Proceedings of the Royal Society of Edinburgh, **115A**, 1–18, 1990

Existence and stability of travelling wave solutions for an evolutionary ecology model

Roger Lui
Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609, U.S.A.*

(MS received 30 May 1989)

Synopsis

Monotone travelling wave solutions are known to exist for Fisher's equation which models the propagation of an advantageous gene in a single locus, two alleles population genetics model. Fisher's equation assumed that the population size is a constant and that the fitnesses of the individuals in the population depend only on their genotypes. In this paper, we relax these assumptions and allow the fitnesses to depend also on the population size. Under certain assumptions, we prove that in the second heterozygote intermediate case, there exists a constant $\theta^* > 0$ such that monotone travelling wave solutions for the reaction–diffusion system exist whenever $\theta > \theta^*$. We also discuss the stability properties of these waves.

1. Introduction

Consider the equation

$$u_t = u_{xx} + h(u) \quad \text{in } \mathbb{R} \times [0, \infty), \tag{1.1}$$

where $h \in C^1[0, 1]$, $h(0) = 0$ and $h(1) = 0$. In 1937, Fisher used this equation with $h(u) = u(1 - u)$ to model the spatial spread of an advantageous gene in a population living in a homogeneous one-dimensional habitat [**4**]. Fisher assumed that the individuals in the population carry a gene that occurs in two forms, A and a. Then there are three genotypes: the homozygotes AA, aa and the heterozygote Aa. Individuals in this population are classified according to the genotype to which they belong. The ability of an individual to survive to adulthood depends entirely on its genotype. The specific form of h that Fisher assumed means that the fitness of the heterozygotes is between those of the homozygotes.

In 1975, Aronson and Weinberger studied (1.1), allowing h to have an intermediate zero, denote by α, between 0 and 1 [**1**]. Under such an assumption, there are two additional cases to consider: when $h > 0$ on $(0, \alpha)$, $h < 0$ on $(\alpha, 1)$ and when $h < 0$ on $(0, \alpha)$, $h > 0$ on $(\alpha, 1)$. Aronson and Weinberger called Fisher's model the heterozygote intermediate case, and the above two cases the heterozygote superior and heterozygote inferior case, respectively. These last two cases correspond to the situation when the heterozygotes are more fit and less fit to survive than the homozygotes, respectively.

Fisher's model is based on assumptions of a highly idealised situation. The author of this paper is interested in extending the results of Fisher's model to the case when the fitnesses of the individuals also depend on the population size.

* Current address: Department of Mathematics, University of Utah, Salt Lake City, Utah 84112, U.S.A.

Equation (1.1) then becomes a system of reaction–diffusion equations,

$$\begin{cases} p_t = p_{xx} + f(p, N)p(1-p), \\ N_t = N_{xx} + g(p, N)N, \end{cases} \tag{1.2}$$

where $0 \leqq p \leqq 1$ is allele A's frequency and $N \geqq 0$ is the population size. The function $f(p, N) = \eta_A(p, N) - \eta_a(p, N)$, where η_A, η_a represent the fitnesses of allele A and allele a, respectively. If we denote the fitnesses of genotypes AA, Aa and aa by $\eta_1(N)$, $\eta_2(N)$ and $\eta_3(N)$, then $\eta_A = p\eta_1 + (1-p)\eta_2$ and $\eta_a = p\eta_2 + (1-p)\eta_3$. The second equation in (1.2) models the growth and spread of the population. The function g is the specific growth-rate and may be thought of as the fitness of the entire population. Thus $g(p, N) = p\eta_A + (1-p)\eta_a = p^2\eta_1 + 2p(1-p)\eta_2 + (1-p)^2\eta_3$. It is important to observe the relation

$$g_p(p, N) = 2f(p, N). \tag{1.3}$$

The mathematical theory for equation (1.1) is very rich and well understood (see **[3]** for details). One of the most intriguing properties of (1.1) is the existence of travelling waves. A travelling wave solution of (1.1) with speed θ is a non-constant function $\tilde{u}(\xi)$ such that $\tilde{u}(x + \theta t)$ satisfies (1.1) for all x and $t > 0$. In the heterozygote intermediate case, Kolmogorov, Petrovski and Piscunoff showed in 1937 (among other things) that there exists $\theta^* > 0$ such that increasing travelling waves $\tilde{u}$ exist with the properties that $\tilde{u}(-\infty) = 0$, $\tilde{u}(\infty) = 1$ if and only if $\theta \geqq \theta^*$ **[6]**. For the heterozygote inferior case, Aronson and Weinberger showed that if $\int_0^1 h > 0$, then there exists $\theta^* > 0$ such that increasing travelling wave solutions connecting 0 and 1 exist if and only if $\theta = \theta^*$. For the heterozygote superior case, since 0 and 1 are unstable solutions of the ordinary differential equation $u' = h(u)$, no travelling wave solution exists connecting 0 and 1.

For (1.2), it turns out that under appropriate assumptions on f and g, the model may be classified into four cases: the first and second heterozygote intermediate cases, the heterozygote superior case and the heterozygote inferior case. The asymptotic behaviour of solutions to (1.2) with bounded initial-data in all four cases have been investigated by the author in **[7]**. Now we are interested in proving the existence of travelling waves that connect two critical points of (1.2). This problem involves analysing the behaviour of trajectories of a four-dimensional ordinary differential equation containing a parameter (the wave speed). In this paper, we employ the shooting argument to prove the existence of travelling wave solutions for the second intermediate case for sufficiently large wave speed. We also prove the stability of these waves (in a certain sense) using the maximum principles. We hope to investigate the existence and stability of travelling wave solutions for the first intermediate and heterozygote inferior cases in the future. For the inferior case, the techniques used will be substantially different from those used in this paper, since travelling waves are expected to exist only for a discrete set of values of the wave speed.

The organisation of the paper is as follows. In Section 2, we state the hypotheses of f and g and classify (1.2) into four cases. In Section 3, we prove the existence of travelling waves for the second intermediate case and in Section 4, we study their stability properties. The main results are Theorems 3.1 and 4.3.

2. Hypotheses

Let f and g be twice continuously differentiable in a neighbourhood of the region $Q = \{(p, N) \mid 0 \leqq p \leqq 1, N \geqq 0\}$. We assume that

$$f_N > 0, \quad g_N < 0 \quad \text{in } Q \tag{2.1}$$

and that there exists a positive constant M such that

$$f(p, 0) < 0 \quad \text{and} \quad f(p, M) > 0 \quad \text{for } 0 \leqq p \leqq 1, \tag{2.2}$$

$$g(p, 0) > 0 \quad \text{and} \quad g(p, M) < 0 \quad \text{for } 0 \leqq p \leqq 1. \tag{2.3}$$

From (2.1) and the Implicit Function Theorem, there exist functions $\hat{N}$ and $\tilde{N}$ such that $f(p, \tilde{N}(p)) = 0$ and $g(p, \hat{N}(p)) = 0$ for $0 \leqq p \leqq 1$. Let Γ_1 and Γ_2 be the graphs of $\hat{N}$ and $\tilde{N}$ in Q, respectively. We assume that Γ_1 and Γ_2 do not intersect more than once and if they do intersect, they do so non-tangentially. Then there are four basic types of behaviour for Γ_1 and Γ_2:

Γ_1 lies above Γ_2; (2.4)

Γ_1 lies below Γ_2; (2.5)

Γ_1 intersects Γ_2 once with $\hat{N}(0) > \tilde{N}(0)$ and $\hat{N}(1) < \tilde{N}(1)$; (2.6)

Γ_1 intersects Γ_2 once with $\hat{N}(0) < \tilde{N}(0)$ and $\hat{N}(1) > \tilde{N}(1)$. (2.7)

Following the reasoning given in **[7]**, we call (2.4) and (2.5) the first and second heterozygote intermediate cases and (2.6), (2.7) the heterozygote superior and inferior cases, respectively.

As mentioned in Section 1, our model implies that $g_p = 2f$. This condition is stronger than necessary and all we need to assume for our theorems is the fact that

$$\operatorname{sign} g_p = \operatorname{sign} f \quad \text{in } Q. \tag{2.8}$$

From the definition of $\hat{N}$, we have $\hat{N}' = -(g_p/g_N)$. Therefore (2.8) implies that Γ_1 is increasing in case (2.4), decreasing in case (2.5) and has a local maximum (minimum) at the point of intersection with Γ_2 in case (2.6) (case (2.7)).

A travelling wave solution to (1.2) with speed θ is a pair of non-constant functions $\tilde{p}(\xi)$, $\tilde{N}(\xi)$ such that $\tilde{p}(x + \theta t)$, $\tilde{N}(x + \theta t)$ satisfy (1.2) for all x and $t > 0$. This is equivalent to saying that $\tilde{p}$, $\tilde{N}$ satisfy the system

$$\begin{cases} p' = q \\ q' = \theta q - f(p, N)p(1-p) \\ N' = M \\ M' = \theta M - g(p, N)N \end{cases} \tag{2.9}$$

for $-\infty < \xi < \infty$ (we dropped the tilda in (2.9)). $(\tilde{p}, \tilde{N})$ should also have a limit as $\xi \to \pm\infty$ which must then be a root of the equations $f(p, N)p(1-p) = 0$ and $g(p, N)N = 0$.

Let us find the roots of the equations $f(p, N)p(1-p) = 0$ and $g(p, N)N = 0$. For cases (2.4) to (2.7), there are four roots which lie on the boundary of Q. They are $(p, N) = (0, 0)$, $(1, 0)$, $(0, K_3)$ and $(1, K_1)$. The last two roots are the endpoints of Γ_1 in Q. There is also a root (p^*, N^*) in the interior of Q in cases (2.6) and (2.7). All these roots are equilibrium solutions of (1.2) with or without diffusion. We can analyse their local stability properties without diffusion by finding the eigenvalues of the linearised matrix at the respective equilibrium points. Doing so, we find that $(0, 0)$ and $(1, 0)$ are always unstable. For (2.4), $(1, K_1)$ is stable and $(0, K_3)$ is unstable while the opposite is true for (2.5). For (2.6), all the boundary equilibrium points are unstable and (p^*, N^*) is stable. For (2.7), $(0, K_3)$ and $(1, K_1)$ are stable while (p^*, N^*) is unstable.

We now return to the problem of finding travelling waves. Let (p_1, N_1), (p_2, N_2) be two equilibrium points mentioned above. If $p_1 = p_2 = 0$ or 1 or if $N_1 = N_2 = 0$, then finding a solution of (2.9) connecting the critical points $(p_1, 0, N_1, 0)$ and $(p_2, 0, N_2, 0)$ concerns only two of the equations of (2.9). The methods and results for such smaller systems are well known (see the discussion in Section 1). Here we are interested in connecting critical points which lie on opposite sides of $p = 0$ and $p = 1$. We are able to do so for the case (2.5) and for that we need the additional assumption that

$$\Gamma_2 \text{ lies above the line } N = K_3. \tag{2.10}$$

The result will then be the existence of monotone travelling wave solutions connecting $(1, 0, 0, 0)$ to $(0, 0, K_3, 0)$ for sufficiently large θ. There may exist non-monotone travelling waves connecting the same critical points so that we do not have uniqueness.

There are a number of papers in the recent past which proved the existence of travelling wave solutions for a reaction–diffusion system using techniques similar to the one used in this paper (see [**2**, **8**, **10**]). In fact, some of the ideas of the author in this paper grew out of reading a very nice paper by Dunbar [**2**]. Dunbar treated a more difficult problem than (2.5) in that, in his paper, the critical point at $-\infty$ has a three-dimensional unstable manifold, whereas in (2.5), the critical point $(1, 0, 0, 0)$ has a four-dimensional unstable manifold.

3. Existence of travelling waves

Consider the equation $p_t = p_{xx} + f(p, 0)p(1-p)$. Let $r = 1 - p$. Then r satisfies $r_t = r_{xx} - f(1-r, 0)r(1-r)$. Let $h(r) = -f(1-r, 0)r(1-r)$. Then $h \in C^1[0, 1]$, $h(0) = 0$, $h(1) = 0$, $h > 0$ on $(0, 1)$ and $h'(0) = -f(1, 0) > 0$. Therefore there exists $\theta_1 \geqq 2\sqrt{-f(1, 0)}$ such that decreasing travelling wave solutions connecting $p = 1$ to $p = 0$ exist for $p_t = p_{xx} + f(p, 0)p(1-p)$ if $\theta \geqq \theta_1$. We define $\theta_2 \geqq 2\sqrt{-f(1, K_3)}$ similarly for the equation $p_t = p_{xx} + f(p, K_3)p(1-p)$. Next consider the equation $N_t = N_{xx} + g(1, N)N$. Let $h(N) = g(1, N)N$. Then $h \in C^1[0, K_1]$, $h(0) = 0$, $h(K_1) = 0$, $h > 0$ on $(0, K_1)$ and $h'(0) = g(1, 0) > 0$. Therefore, there exists $\theta_3 \geqq 2\sqrt{g(1, 0)}$ such that increasing travelling wave solutions joining $N = 0$ to $N = K_1$ exist for $N_t = N_{xx} + g(1, N)N$ if $\theta \geqq \theta_3$. We define $\theta_4 \geqq 2\sqrt{g(0, 0)}$ similarly for the equation $N_t = N_{xx} + g(0, N)N$. We denote these four monotone travelling waves by w_1, w_2, w_3, w_4, respectively.

We assume the reader is familiar with the basic results on stable and unstable manifolds and the LaSalle–Lyapunov Invariant Principle [**5**]. We also need to use Wazewski's Theorem which is a formalisation of the shooting argument. Its statement and proof may be found in [**2**, Section 2]. We are now ready to state and prove our main theorem.

THEOREM 3.1. *Suppose f and g are continuously differentiable in a neighbourhood of Q and* (2.1)–(2.3), (2.5), (2.8) *and* (2.10) *hold. Let $\theta^* = \max(\theta_1, \theta_2, \theta_3, \theta_4)$. Then travelling wave solutions $p(\xi)$, $N(\xi)$ for* (1.1) *exist if $\theta > \theta^*$. Furthermore, $p' < 0$, $N' > 0$, (p, p', N, N') approaches* $(1, 0, 0, 0)$ *as $\xi \to -\infty$ and it approaches* $(0, 0, K_3, 0)$ *as $\xi \to \infty$.*

Proof. We first find the dimensions of the stable and unstable manifolds at the critical points $A = (1, 0, 0, 0)$ and $B = (0, 0, K_3, 0)$. The Jacobian matrix for (2.9) at A, denoted by J_A, has four positive eigenvalues as long as $\theta^2 > \max(-4f(1, 0), 4g(1, 0))$, which is satisfied. At B, J_B has two positive and two negative eigenvalues. Therefore the unstable manifold at A is four-dimensional and the stable manifold at B is two-dimensional.

Since J_A and J_B have real eigenvalues, it is reasonable to look for monotone travelling waves. Since f is negative below K_3, the wave $p(\xi)$ connecting 0 and 1 should either be decreasing and moving left or increasing and moving right. We assume that $\theta > 0$ so that p should be decreasing. Likewise, the wave $N(\xi)$ should be increasing. We then try to look for solutions of (2.9) in the set $W = \{(p, q, N, M) \mid 0 < p < 1,\ q < 0,\ 0 < N < K_3,\ M > 0\}$ in $\mathbb{R}^4$. For the rest of this section, $\mathbf{y}(\xi, \mathbf{y}_0)$ will denote the solution to (2.9) that equals $\mathbf{y}_0$ at $\xi = 0$.

If $\mathbf{y}_0 \in W$, then $\mathbf{y}(\xi, \mathbf{y}_0)$ either remains in W for all $\xi \geqq 0$ or exits W through W^-. W^- contains part of the boundaries of W obtained by setting $p = 0$ or $q = 0$ or $N = K_3$ or $M = 0$ in the definition of W. We denote these four sets by $\{p = 0\}$, $\{q = 0\}$, $\{N = K_3\}$ and $\{M = 0\}$, respectively. The solution $\mathbf{y}$ cannot exit at the boundary $\{p = 0, q = 0\} = \{(p, q, N, M) \mid p = 0,\ q = 0,\ 0 < N < K_3,\ M > 0\}$ since it is contained in the invariant manifold $p = q = 0$ and $\mathbf{y}_0 \in W$. Similarly, $\mathbf{y}$ cannot exit at the boundary $\{N = K_3, M = 0\}$. The union of the above four boundaries of W plus $\{p = 0, M = 0\}$, $\{M = 0, q = 0\}$, $\{q = 0, N = K_3\}$ and $\{N = K_3, p = 0\}$ is the exit set W^-. A good way to look at W^- is shown in Figure 1, where sets on opposite sides do not intersect.

An important observation here is that if $\mathbf{y}_0 \in W$ and $\mathbf{y}(\xi^*, \mathbf{y}_0) \in W^-$, then

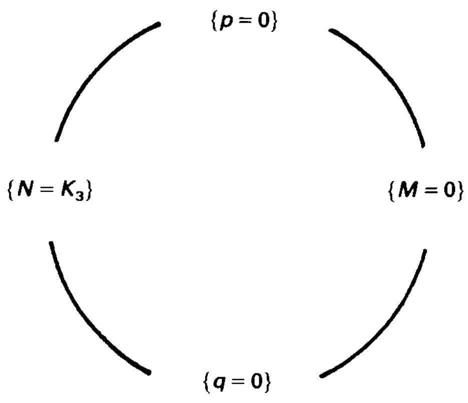

Figure 1. The set W^-.

$\mathbf{y}(\xi, \mathbf{y}_0) \notin \bar{W}$ (closure of W) for ξ immediately after ξ^*. This is obvious if $\mathbf{y}(\xi, \mathbf{y}_0)$ lands on $\{p=0\}$ or $\{N=K_3\}$. If $\mathbf{y}(\xi^*, \mathbf{y}_0) \in \{q=0\}$, then $q' = -f(p, N)p(1-p)$ is positive at ξ^* since $0<p(\xi^*)<1$ and (2.10) holds. Similarly, if $\mathbf{y}(\xi^*, \mathbf{y}_0) \in \{M=0\}$, then $M'(\xi^*)<0$ from the last equation in (2.9). The immediate exit set of W is defined as $W^* = \{\mathbf{y}_0 \in \partial W \mid \mathbf{y}(\xi, \mathbf{y}_0) \notin \bar{W}$ for $\xi>0$, ξ sufficiently small$\}$. (∂W means the boundary of W.) It is clear from above that $W^- \subset W^*$. In fact, W^* is just W^- union the following sets: $\{p=0, N=0\}$, $\{p=1, N=K_3\}$, $\{p=1, M=0\}$, $\{q=0, N=0\}$, $\{p=0$ or 1, $q=0$, $N=K_3\}$, $\{p=0$ or 1, $q=0, M=0\}$, $\{N=0$ or K_3, $M=0, p=0\}$, $\{N=0$ or K_3, $M=0$, $q=0\}$. Note that a trajectory of (2.9) with initial data in W cannot reach any of these sets in forward ξ. W^- will play an important role in our proof.

We claim that

$$\text{for } \theta>\theta^*, \text{ there exists a } \mathbf{y}_0 \in W \text{ such that } \mathbf{y}^*(\xi)=\mathbf{y}(\xi, \mathbf{y}_0) \text{ remains in } W \text{ for all } \xi \text{ and } \mathbf{y}^*(\xi)\to A \text{ as } \xi \to -\infty. \tag{3.1}$$

Suppose for the moment that (3.1) is true. Then $\mathbf{y}^*=(p^*, q^*, N^*, M^*)$ must be bounded in $\mathbb{R}^4$. To see this, let C be chosen such that $|-f(p, N)p(1-p)| \leqq C$ for $0 \leqq p \leqq 1$ and $0 \leqq N \leqq K_3$. If $q^*(\xi^*)<-C/\theta$ at some ξ^*, then $[q^*]'(\xi^*)<0$ so that $q^*(\xi)<-C/\theta$ for $\xi \leqq \xi^*$. This implies that $p^*(\xi)=0$ for some finite value of ξ, a contradiction. Therefore $q^* \geqq -C/\theta = -C_1$. Similarly, we can show that M^* is bounded above by some positive constant C_2.

Consider the function $V(p, q, N, M)=g(1-p, N)-q+\theta p$ which is defined and differentiable on W. Then $\dot{V}=-g_p(1-p, N)q+f(p, N)p(1-p)+g_N(1-p, N)M$ where $\dot{V}$ means the directional derivative of V in the direction given by the vector whose components are the right-hand side of (2.9). From (2.8), (2.1) and (2.10), each term in the above expression is non-positive so that $\dot{V} \leqq 0$ in $\bar{W}$. Thus V is a Lyapunov function on W. Let $E=\{(p, q, N, M) \in \bar{W} \mid \dot{V}=0\}$. Since $\mathbf{y}^*$ lies in W for $\xi \geqq 0$ and is bounded, the Invariant Principle implies that $\mathbf{y}^*$ must converge to the largest invariant set inside E. Such a set contains only two elements, $(0,0,0,0)$ or $(0,0,K_3,0)$. For $\mathbf{y}^*$ to approach $(0,0,0,0)$, it must intersect the stable manifold at $(0,0,0,0)$ which has dimension one and lies inside the invariant manifold $N=0$, $M=0$. Since $\mathbf{y}^*$ never intersects this invariant manifold, $\mathbf{y}^*$ must approach $(0,0,K_3,0)$ as $\xi \to \infty$. Therefore the proof of Theorem 3.1 is complete once we prove (3.1).

To prove (3.1), we shall define a set $\Sigma \in \mathbb{R}^4$ with the following properties (a)–(d):

(a) Σ is a simply connected, compact subset of W.

(b) If a trajectory of (2.9) intersects Σ, then it must intersect W^- before it intersects Σ again in forward ξ.

(c) If $\mathbf{y}_0 \in \Sigma$ and $\mathbf{y}(\xi, \mathbf{y}_0) \in \bar{W}$ but $\mathbf{y}(\xi, \mathbf{y}_0) \notin W^*$, then there exists an open neighbourhood of $\mathbf{y}(\xi, \mathbf{y}_0)$ in W which is disjoint from W^*.

Suppose such a set Σ can be found. Then there exists $\mathbf{y}_0 \in \Sigma$ such that $\mathbf{y}(\xi, \mathbf{y}_0) \in W$ for $\xi \geqq 0$. To see this, assume the contrary. Then Wazewski's Theorem (which uses (b) and (c) above) implies that the solution map of (2.9) is a homeomorphism between Σ and its image on W^-. We shall also show that the image of $\partial\Sigma$ forms a closed loop around $B=(0,0,K_3,0)$ in W^- (see Fig. 1).

These and the fact that Σ is simply connected contradict the fact that

$$W^- \text{ is not simply connected.} \tag{3.2}$$

Therefore there exists $\mathbf{y}_0 \in \Sigma$ such that $\mathbf{y}(\xi, \mathbf{y}_0) \in W$ for all $\xi \geqq 0$. The proof of (3.1) is completed by showing the following:

(d) Let $\mathbf{y}_0 \in \Sigma$. Then $\mathbf{y}(\xi, \mathbf{y}_0) \in W$ for all $-\infty < \xi \leqq 0$ and $\mathbf{y}(\xi, \mathbf{y}_0) \to A$ as $\xi \to -\infty$.

The rest of this paper is devoted to finding Σ with the desired properties and proving (3.2).

Let $\theta > \theta^*$ and let w_1, w_2, w_3, w_4 be the monotone travelling waves defined at the beginning of this section. Consider the phase plane of the system $p' = q$, $q' = \theta q - f(p, 0)p(1-p)$. Let $z_1 = w_1'$. Then $(w_1(\xi), z_1(\xi))$ traces a curve joining $(1, 0)$ to the origin in the phase plane (see Fig. 2). Let $\lambda_1 > \lambda_2 > 0$ be the eigenvalues of the matrix $\begin{bmatrix} 0 & 1 \\ f(1,0) & \theta \end{bmatrix}$. It is known that $\dfrac{dz_1}{dw_1}(-\infty) = \lambda_2$ (here we need $\theta > \theta_1$). The phase plane for the system $p' = q$, $q' = \theta q - f(p, K_3)p(1-p)$ is similar to Figure 2, except that we have to replace λ_1, λ_2 by μ_1, μ_2, where $\mu_1 > \mu_2 > 0$ are the eigenvalues of the matrix $\begin{bmatrix} 0 & 1 \\ f(1, K_3) & \theta \end{bmatrix}$. Note that $\mu_2 < \lambda_2$ since $f_N > 0$ and $\dfrac{dz_2}{dw_2}(-\infty) = \mu_2$ (here we need $\theta > \theta_2$).

Let $\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$ be two positive unit vectors such that

$$0 < \frac{x_{22}}{x_{21}} < \mu_2 < \lambda_2 < \frac{x_{12}}{x_{11}} < \lambda_1. \tag{3.3}$$

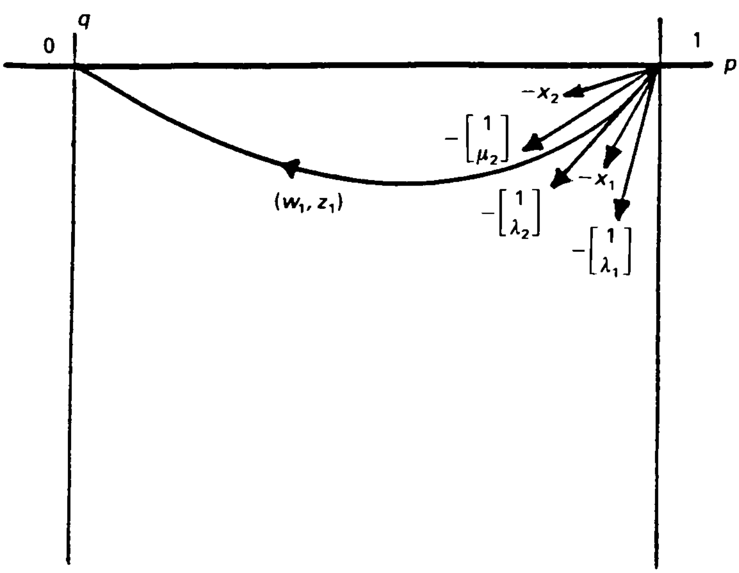

Figure 2. Phase plane diagram for the system $p' = q$, $q' = \theta q - f(p, 0)p(1-p)$. The vector $-\begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$ is tangent to the curve (w_1, z_1) at $(1, 0)$.

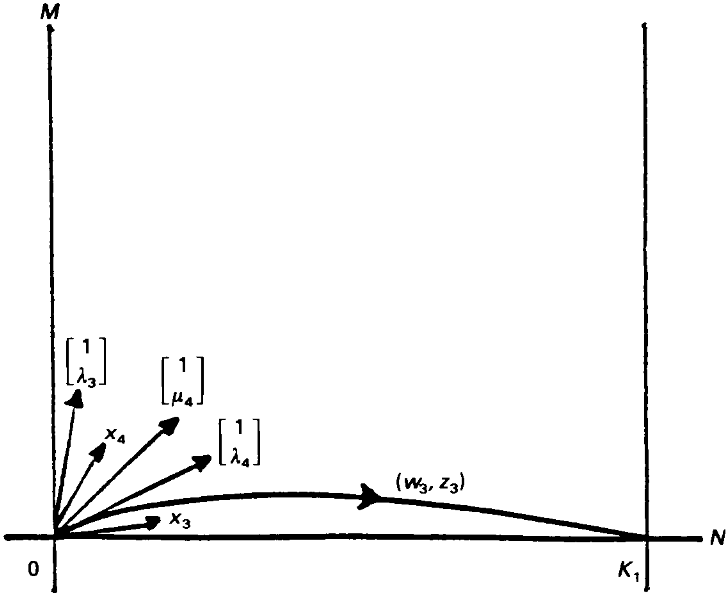

Figure 3. Phase plane diagram for the system $N' = M,\ \ M' = \theta M - g(1, N)N$. The vector $\begin{bmatrix}1\\ \lambda_4\end{bmatrix}$ is tangent to the curve (w_3, z_3) at $(0, 0)$.

This condition is equivalent to choosing $\mathbf{x}_1$, $\mathbf{x}_2$ at certain relative positions to the eigenvectors $\begin{bmatrix}1\\ \lambda_1\end{bmatrix}$, $\begin{bmatrix}1\\ \lambda_2\end{bmatrix}$ and $\begin{bmatrix}1\\ \mu_2\end{bmatrix}$ as indicated in Figure 2.

Next we consider the phase plane of the system $N' = M,\ M' = \theta M - g(1, N)N$. Let $z_3 = w_3'$. Then $(w_3(\xi), z_3(\xi))$ traces out a curve joining the origin to $(K_1, 0)$ in the phase plane (see Fig. 3). Let $\lambda_3 > \lambda_4 > 0$ be the eigenvalues of the matrix $\begin{bmatrix}0 & 1\\ -g(1,0) & \theta\end{bmatrix}$. It is known that $\dfrac{dz_3}{dw_3}(-\infty) = \lambda_4 (\theta > \theta_3)$. The phase plane of the system $N' = M,\ M' = \theta M - g(0, N)N$ is similar to Figure 3, except that we have to replace K_1 by K_3 and λ_3, λ_4 by μ_3, μ_4, where $\mu_3 > \mu_4 > 0$ are the eigenvalues of the matrix $\begin{bmatrix}0 & 1\\ -g(0,0) & \theta\end{bmatrix}$. Also $\dfrac{dz_4}{dw_4}(-\infty) = \mu_4(\theta > \theta_4)$ and $\lambda_4 < \mu_4 < \dfrac{\theta}{2} < \lambda_3$, since $g_p(p, 0) < 0$.

Let $\mathbf{x}_3 = \begin{bmatrix}x_{31}\\ x_{32}\end{bmatrix}$, $\mathbf{x}_4 = \begin{bmatrix}x_{41}\\ x_{42}\end{bmatrix}$ be positive unit vectors such that

$$0 < \frac{x_{32}}{x_{31}} < \lambda_4 < \mu_4 < \frac{x_{42}}{x_{41}} < \lambda_3. \tag{3.4}$$

This condition is equivalent to choosing $\mathbf{x}_3$, $\mathbf{x}_4$ at certain relative positions to the eigenvectors $\begin{bmatrix}1\\ \lambda_4\end{bmatrix}$, $\begin{bmatrix}1\\ \mu_4\end{bmatrix}$ and $\begin{bmatrix}1\\ \lambda_3\end{bmatrix}$. See Figure 3.

Having defined $\mathbf{x}_i$, $i = 1, 2, 3, 4$, we consider the trajectories of (2.9) with initial data

$$\mathbf{y}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \alpha \begin{bmatrix} \mathbf{x}_1 \\ 0 \\ 0 \end{bmatrix} - \beta \begin{bmatrix} \mathbf{x}_2 \\ 0 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ 0 \\ \mathbf{x}_3 \end{bmatrix} + \delta \begin{bmatrix} 0 \\ 0 \\ \mathbf{x}_4 \end{bmatrix}, \tag{3.5}$$

where α, β, γ and δ are nonnegative constants. The following statements can be easily verified:

(3.6)

(i) If $\beta = \gamma = \delta = 0$, then $\mathbf{y}(\xi, \mathbf{y}_0)$ lies on the invariant manifold $M = N = 0$. From Figure 2, if $\alpha > 0$ is sufficiently small, then $\mathbf{y}(\xi, \mathbf{y}_0)$ intersects $p = 0$ first, before $q = 0$.

(ii) If $\alpha = \gamma = \delta = 0$, then $\mathbf{y}(\xi, \mathbf{y}_0)$ lies in the invariant manifold $M = N = 0$ and if $\beta > 0$ is sufficiently small, then $\mathbf{y}(\xi, \mathbf{y}_0)$ intersects $q = 0$ first, before $p = 0$.

(iii) If $\alpha = \beta = \delta = 0$, then $\mathbf{y}(\xi, \mathbf{y}_0)$ lies in the invariant manifold $p = 1$, $q = 0$. From Figure 3, if $\gamma > 0$ is sufficiently small, then $\mathbf{y}(\xi, \mathbf{y}_0)$ intersects $M = 0$ before $N = K_1$.

(iv) If $\alpha = \beta = \gamma = 0$, then $\mathbf{y}(\xi, \mathbf{y}_0)$ lies in the invariant manifold $p = 1$, $q = 0$ and if $\delta > 0$ is sufficiently small, then $\mathbf{y}(\xi, \mathbf{y}_0)$ intersects $N = K_1$ before $M = 0$. Once $\mathbf{y}(\xi, \mathbf{y}_0)$ crosses $N = K_1$, M' and M will continue to be positive (see (2.9)) so that $\mathbf{y}(\xi, \mathbf{y}_0)$ reaches $N = K_3$ before $M = 0$.

One of the problems with choosing initial data $\mathbf{y}_0$ according to (3.6) is that $\mathbf{y}_0 \notin W$. The following lemmas say that we may choose $\mathbf{y}_0 \in W$ and $\mathbf{y}(\xi, \mathbf{y}_0)$ will still exit W in the right way. Recall the set W^-.

LEMMA 3.2. *Let $\beta = 0$ and let $\mathbf{y}_0$ satisfy* (3.5) *with $\alpha > 0$ sufficiently small. Then $\mathbf{y}(\xi, \mathbf{y}_0)$ reaches $\{p = 0\}$ first before $\{q = 0\}$ as long as $N > 0$.*

Proof. Think of $N(\xi)$ as a given function and consider the system

$$\begin{cases} p' = q, \\ q' = \theta q - f(p, N)p(1-p). \end{cases} \tag{3.7}$$

Plot the solution $(p(\xi), q(\xi))$ on the phase plane in Figure 2. Since $\beta = 0$, $(p(0), q(0))$ lies below the curve representing the travelling wave (w_1, z_1). Suppose $(p(\xi), q(\xi))$ crosses this curve at $\xi = \xi^*$. Let $(w_1^*, z_1^*) = (p(\xi^*), q(\xi^*))$. Then at (w_1^*, z_1^*), $\dfrac{dw_1}{dz_1} \geqq \dfrac{dq}{dp}$ which, according to (3.7) and (2.9), implies that $f(p(\xi^*), N(\xi^*)) \leqq f(w_1^*, 0)$. This contradicts the fact that $f_N > 0$ if $N(\xi^*) > 0$. Therefore $(p(\xi), q(\xi))$ remains below (w_1, z_1) for $\xi \geqq 0$ and it reaches $\{p = 0\}$ first before $\{q = 0\}$. This completes the proof of Lemma 3.2. □

The following lemmas may be proved in the same way as Lemma 3.2. In Lemma 3.3, we have to use the travelling wave (w_2, z_2), and the fact that $\dfrac{x_{22}}{x_{21}} < \mu_2$. In Lemma 3.4, we have to use the travelling wave (w_3, z_3) in the phase plane Figure 3, and the fact that $\frac{x_{32}}{x_{31}} < \lambda_4$. Finally, in Lemma 3.5, we have to use the travelling wave (w_4, z_4) and the fact that $\mu_4 < \frac{x_{42}}{x_{41}}$.

LEMMA 3.3. *Let $\alpha = 0$ and $\beta > 0$ be sufficiently small. Then $\mathbf{y}(\xi, \mathbf{y}_0)$ reaches $\{q = 0\}$ first before $\{p = 0\}$ as long as $N < K_3$.*

LEMMA 3.4. *Let $\delta = 0$ and $\gamma > 0$ be sufficiently small. Then $\mathbf{y}(\xi, \mathbf{y}_0)$ reaches $\{M = 0\}$ first before $\{N = K_3\}$ as long as $p < 1$.*

LEMMA 3.5. *Let $\gamma = 0$ and $\delta > 0$ be sufficiently small. Then $\mathbf{y}(\xi, \mathbf{y}_0)$ reaches $\{N = K_3\}$ first before $\{M = 0\}$ as long as $p > 0$.*

We now define a closed loop σ in W which will become the boundary of Σ. Recall that we are still trying to define Σ. Let $\sigma = \{\mathbf{y}_0(s) \mid 0 \leqq s \leqq 2\pi\}$, where

$$\mathbf{y}_0(s) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \alpha(s)\begin{bmatrix} \mathbf{x}_1 \\ 0 \\ 0 \end{bmatrix} - \beta(s)\begin{bmatrix} \mathbf{x}_2 \\ 0 \\ 0 \end{bmatrix} + \gamma(s)\begin{bmatrix} 0 \\ 0 \\ \mathbf{x}_3 \end{bmatrix} + \delta(s)\begin{bmatrix} 0 \\ 0 \\ \mathbf{x}_4 \end{bmatrix}. \tag{3.8}$$

The loop σ has the following properties:

(i) α, β, γ and δ are nonnegative continuous functions defined on the interval $[0, 2\pi]$.
(ii) $\mathbf{y}_0(s_1) = \mathbf{y}_0(s_2)$ if and only if $s_1 = s_2$ or $s_1 = 0$ and $s_2 = 2\pi$.
(iii) $\sigma \subset W$.
(iv) σ lies on the ellipsoid. (3.9)

$$\mathcal{S} = \left\{(x, y, z, w) \;\middle|\; \frac{\theta^2(x-1)^2}{2} + 2y^2 + \frac{\theta^2 z^2}{2} + 2w^2 = \varepsilon^2\right\}. \tag{3.10}$$

The constant $\varepsilon > 0$ in (3.10) is chosen to be sufficiently small so that if $(p, q, N, M) \in \mathcal{S}$, then $\theta^2 + 4f(p, N)p > 0$ and $\theta^2 - 4g(p, N) > 0$. This is possible because $\theta > \theta_1 \geqq 2\sqrt{-f(1, 0)}$ and $\theta > \theta_3 \geqq 2\sqrt{g(1, 0)}$. We also choose ε so small that if $\mathbf{y}_0 \in \mathcal{S}$, then $\mathbf{y}(\xi, \mathbf{y}_0) \to A$ as $\xi \to -\infty$. Recall that J_A has four positive eigenvalues.

The best way to describe σ is draw the graphs of α, β, γ and δ. From Figure 4, (3.9)(i) and (ii) are obvious, (3.9)(iii) follows because $\alpha(s)$, $\beta(s)$ or $\gamma(s)$, $\delta(s)$ do not vanish simultaneously. Finally, (3.9)(iv) is equivalent to choosing α, β, γ and δ to satisfy the equation

$$\frac{\theta^2(\alpha x_{11} + \beta x_{21})^2}{2} + 2(\alpha x_{12} + \beta x_{22})^2 + \frac{\theta^2(\gamma x_{31} + \delta x_{41})^2}{2} + 2(\gamma x_{32} + \delta x_{42})^2 = \varepsilon^2. \tag{3.11}$$

At the same time as we are defining $\mathbf{y}_0(s)$, we also show that assuming that $\mathbf{y}(\xi, \mathbf{y}_0(s))$ exits W for all $0 \leqq s \leqq 2\pi$, its image on W^- forms a closed loop around $B = (0, 0, K_3, 0)$. (If for some $\bar{s}$, $\mathbf{y}(\xi, \mathbf{y}_0(\bar{s}))$ remains in W for all $\xi \geqq 0$, then we can show that $\mathbf{y}(\xi, \mathbf{y}_0(\bar{s}))$ approaches A in W as $\xi \to -\infty$ and (3.1) is valid.)

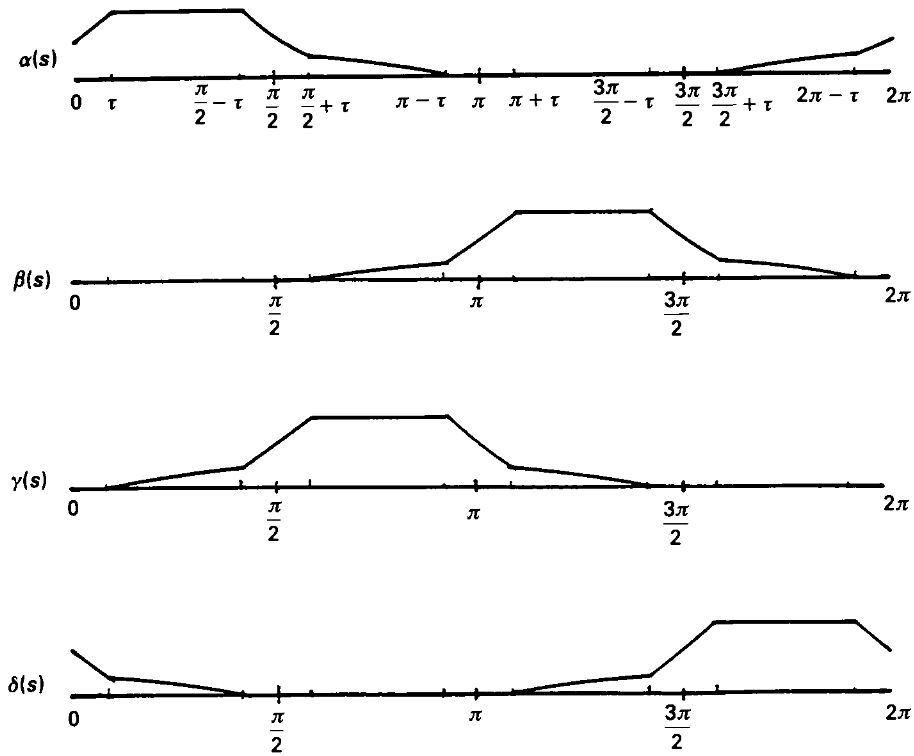

Figure 4. The graphs of $\alpha(s)$, $\beta(s)$, $\gamma(s)$ and $\delta(s)$. $\alpha(0)=\alpha(2\pi)$, $\beta(0)=\beta(2\pi)$, $\gamma(0)=\gamma(2\pi)$, $\delta(0)=\delta(2\pi)$.

We introduce the notation $\mathbf{y}_0(s)=(p_0, q_0, N_0, M_0)$ and $\mathbf{y}(\xi, \mathbf{y}_0(s))=(p(\xi), q(\xi), N(\xi), M(\xi))$ temporarily for the next three pargraphs.

Consider first the interval $\tau \leqq s \leqq \frac{\pi}{2}-\tau$ where τ appears in Figure 4. Let $\alpha(s)>0$ be constant and $\beta(s)=0$ on this interval in the definition of $\mathbf{y}_0$ (independent of s). Suppose γ, δ also equal to zero. Then 3.6(i) implies that $N(\xi)=0$, $M(\xi)=0$ for all ξ and $\mathbf{y}(\xi, \mathbf{y}_0)$ first exits $\bar{W}$ through $p=0$, say when $\xi=T$. Let $\varepsilon_1>0$ be sufficiently small. Then $\mathbf{y}(T+\varepsilon_1, \mathbf{y}_0) \notin \bar{W}$. Let V be a neighbourhood of $\mathbf{y}(T+\varepsilon_1, \mathbf{y}_0)$ which is disjoint from $\bar{W}$. Then there exists a neighbourhood U of $\mathbf{y}_0$ such that if $\mathbf{y}_1 \in U$, then $\mathbf{y}(T+\varepsilon_1, \mathbf{y}_1) \in V$. Therefore given $0<\varepsilon_2<K_3$, we can choose $\gamma(s)$, $\delta(s)$ on $\left[\tau, \frac{\pi}{2}-\tau\right]$ to be positive and sufficiently small so that $\mathbf{y}_0(s) \in U$ and $|N(\xi)| \leqq \varepsilon_2$, $|M(\xi)| \leqq \varepsilon_2$ for $0 \leqq \xi \leqq T+\varepsilon_1$. From the above, $\mathbf{y}(T+\varepsilon_1, \mathbf{y}_0(s)) \in V$. From Lemma 3.1, $\mathbf{y}(\xi, \mathbf{y}_0(s))$ could not have first exited W through $\{q=0\}$. Since $0<N(\xi)<K_3$ for $0 \leqq \xi \leqq T+\varepsilon_1$, $\mathbf{y}(\xi, \mathbf{y}_0(s))$ could not have first exited W through $\{N=K_3\}$ either. Therefore $\mathbf{y}(\xi, \mathbf{y}_0(s))$ must first exit W through $\{p=0\} \cup \{M=0\} \cup \{p=0, M=0\}$ if $\tau \leqq s \leqq \frac{\pi}{2}-\tau$ (see Fig. 1). This fact continues to hold even for the interval $\frac{\pi}{2}-\tau<s \leqq \frac{\pi}{2}+\tau$ because of

Lemmas 3.2, 3.4 and Figure 4. Using the same idea, we can conclude the following about how $\mathbf{y}(\xi, \mathbf{y}_0(s))$ exits W.

$$\left.\begin{array}{ll}\text{(i)} & s \in \left[\tau, \dfrac{\pi}{2}+\tau\right], \\ & \mathbf{y}(\xi, \mathbf{y}_0(s)) \text{ exits through } \{p=0\} \cup \{M=0\} \cup \{p=0, M=0\}; \\ \text{(ii)} & s \in \left[\dfrac{\pi}{2}+\tau, \dfrac{3\pi}{2}-\tau\right], \\ & \mathbf{y}(\xi, \mathbf{y}_0(s)) \textit{ exits through } \{M=0\} \cup \{q=0\} \cup \{M=0, q=0\}; \\ \text{(iii)} & s \in \left[\dfrac{3\pi}{2}-\tau, 2\pi-\tau\right], \\ & \mathbf{y}(\xi, \mathbf{y}_0(s)) \text{ exits through } \{q=0\} \cup \{N=K_3\} \cup \{q=0, N=K_3\}; \\ \text{(iv)} & s \in [2\pi-\tau, \tau], \\ & \mathbf{y}(\xi, \mathbf{y}_0(s)) \text{ exits through } \{N=K_3\} \cup \{p=0\} \cup \{N=K_3, p=0\}. \end{array}\right\} \quad (3.12)$$

It is best to refer to Figure 1 when studying (3.12). Note that if $\alpha(s)=0$, $\beta(s)=0$, then $p(\xi)=1$ and $q(\xi)=0$ for all ξ.

On the intervals $\left[\tau, \frac{\pi}{2}-\tau\right]$, $\left[\frac{\pi}{2}+\tau, \pi-\tau\right]$, $\left[\pi+\tau, \frac{3\pi}{2}-\tau\right]$ and $\left[\frac{3\pi}{2}+\tau, 2\pi-\tau\right]$, one of the four functions α, β, γ and δ is a constant, one is zero and the other two are sufficiently small, monotone and arranged to satisfy (3.11) (decreasing $\varepsilon>0$ if necessary). On the rest of $[0, 2\pi]$, two of the above functions are zero, the other two are arranged to be monotone and satisfy (3.11). We now show that there exists $\tau_1 \in \left(\tau, \frac{\pi}{2}-\tau\right)$, $\tau_2 \in \left(\frac{\pi}{2}+\tau, \pi-\tau\right)$, $\tau_3 \in \left(\pi+\tau, \frac{3\pi}{2}-\tau\right)$ and $\tau_4 \in \left(\frac{3\pi}{2}+\tau, 2\pi-\tau\right)$ such that $\mathbf{y}(\xi, \mathbf{y}_0(s))$ first exits W through $\{p=0\}$ if $\tau<s<\tau_1$, first exits W through $\{M=0\}$ if $\frac{\pi}{2}+\tau<s<\tau_2$, first exits W through $\{q=0\}$ if $\tau_3<s<\frac{3\pi}{2}-\tau$ and first exits W through $\{N=K_3\}$ if $\tau_4<s<2\pi-\tau$. These facts plus (3.12) and Figure 1 imply that the first exit points of $\mathbf{y}(\xi, \mathbf{y}_0(s))$ on W^- form a closed loop around B as s increases from 0 to 2π.

To define τ_1, let $z=\frac{M}{N}$. Then $z'-\theta z+z^2=-g(p, N)>-g(0,0)$ as long as $p>0$ and $0<N<K_3$. Let $\bar{z}$ satisfy the equation $\bar{z}'=\theta\bar{z}-\bar{z}^2-g(0,0)$. It is easy to see that if $\bar{z}(0)>\mu_4$, then $\bar{z}(\xi)>0$ and $\lim_{\xi\to\infty} \bar{z}(\xi)=\mu_3$. The constants $\mu_3>\mu_4>0$ are defined in the paragraph after (3.3) and satisfy $\theta\mu-\mu^2-g(0,0)=0$. Let $u=z-\bar{z}$. Then u satisfies the inequality $u'-\theta u+(z+\bar{z})u>0$ so that $u(\xi)>0$ for $\xi>0$ if $u(0)>0$. If we choose $z(0)>\bar{z}(0)>\mu_4$, then $z(\xi)>0$ (hence $M(\xi)>0$) as long as $p(\xi)>0$ and $0<N(\xi)<K_3$. From (3.5), $z(0)=\frac{M_0}{N_0}\to\frac{x_{42}}{x_{41}}$ as $\gamma\to 0$. From Figure 3, $\frac{x_{42}}{x_{41}}>\mu_4$. Therefore from Figure 4, there exists $\tau_1 \in \left(\tau, \frac{\pi}{2}-\tau\right)$ such that $\frac{M_0}{N_0}>\mu_4$ if $\tau \leqq s<\tau_1$. Putting all the facts together, for s in this interval, $\mathbf{y}(\xi, \mathbf{y}_0(s))$ must exit W through $\{p=0\}$. A similar argument shows that there exists $\tau_3 \in \left(\pi + \tau, \frac{3\pi}{2} - \tau\right)$ such that $\mathbf{y}(\xi, \mathbf{y}_0(s))$ exits W through $\{q=0\}$ if $\tau_3 < s < \frac{3\pi}{2} - \tau$. We define τ_2 and τ_4 similarly using the functions $z = \frac{q}{p-1}$, $\bar{z}$ which is a solution to the equation $\bar{z}' = \theta\bar{z} - \bar{z}^2 + f(1, 0)$, the constants $\lambda_1 > \lambda_2 > 0$ and the vector $\mathbf{x}_1$ which appears in Figure 2.

We now define Σ and show that (3.2) and conditions (a)–(d) hold. Let S^3 be the unit sphere in $\mathbb{R}^4$ and let

$$\varphi(x, y, z, w) = \left(\frac{\theta(1-x)}{\sqrt{2}\varepsilon}, \frac{-\sqrt{2}y}{\varepsilon}, \frac{\theta z}{\sqrt{2}\varepsilon}, \frac{\sqrt{2}w}{\varepsilon}\right).$$

Then $\varphi: \mathcal{S} \to S^3$ is a homeomorphism; $\varphi(\sigma)$ lies in S^3 intersecting the positive octant of $\mathbb{R}^4$. A point $(x', y', z', w') \in S^3$, $w' \geqq 0$, may be identified with a point (x', y', z') inside the unit ball in $\mathbb{R}^3$, with distant $\sqrt{1-(w')^2}$ from the origin. Let this map be ψ and let $\sigma' = \psi \circ \varphi(\sigma)$. Then σ' is a closed non-self-intersecting loop inside P. P is defined to be the unit ball intersecting the positive quadrant in $\mathbb{R}^3$.

Consider the following open connected set

$$U = \left\{(\bar{x}, \bar{y}, \bar{z}) \in P \,\middle|\, \frac{x_{22}}{x_{21}} < \frac{\theta\bar{y}}{2\bar{x}} < \frac{x_{12}}{x_{11}}, \frac{x_{32}}{x_{31}} < \frac{\theta\bar{w}}{2\bar{z}} < \frac{x_{42}}{x_{41}}\right\},$$

where $\bar{w} = \sqrt{1 - \bar{x}^2 - \bar{y}^2 - \bar{z}^2}$. U is non-empty because of (3.3) and (3.4). Let $(\bar{x}, \bar{y}, \bar{z}) \in U$, let $\bar{w}$ be defined as above and let $(x, y, z, w) = (\psi \circ \varphi)^{-1}(\bar{x}, \bar{y}, \bar{z})$. Since $\begin{bmatrix} \mathbf{x}_1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{x}_2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \mathbf{x}_3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \mathbf{x}_4 \end{bmatrix}$ are linearly independent, we can write $\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$ in the form of the right-hand side of (3.5). From (3.3), (3.4) and the inequalities satisfied by $\bar{x}$, $\bar{y}$, $\bar{z}$, $\bar{w}$ in the definition of U, we find that α, β, γ and δ in (3.5) are positive.

Let Σ' be a compact, simply connected surface in U with σ' as its boundary. Let $\Sigma = (\psi \circ \varphi)^{-1}(\Sigma')$. Then Σ is compact, simply connected, lies in $\mathcal{S}$ and has σ as its boundary. Condition (a) is therefore satisfied. In addition, if $(x, y, z, w) \in \Sigma$, then $0 < x < 1$, $y < 0$, $z > 0$, $w > 0$ (assuming that $\sqrt{2}\varepsilon < \theta$). If we express an interior point of Σ in the form of (3.5), then α, β, γ and δ are all positive. This last property of Σ will be useful later.

To prove condition (b), let (p, q, N, M) be a trajectory of (2.9) and define $d(\xi) = \frac{\theta^2}{2}(p-1)^2 + 2q^2 + \frac{\theta^2}{2}N^2 + 2M^2$. Then $d'(\xi) = 4\theta q^2 + 4\theta M^2 - (1-p)q[\theta^2 + 4f(p, N)p] + MN[\theta^2 - 4g(p, N)]$. Suppose $(p, q, N, M)(\xi_i) \in \Sigma$ for $i = 1, 2$ and $\xi_1 < \xi_2$. Then from (3.10) and the remarks after it, $d(\xi_i) = \varepsilon^2$ and

$d'(\xi_i) > 0$. Therefore there exists $\xi_1 < \xi^* < \xi_2$ such that $d(\xi^*) = \varepsilon^2$ and $d'(\xi^*) \leqq 0$. $d(\xi^*) = \varepsilon^2$ implies that the terms inside the square brackets of $d'(\xi^*)$ are positive. Therefore, either $(1 - p(\xi^*))q(\xi^*) > 0$ or $M(\xi^*)N(\xi^*) < 0$. This forces $(p, q, N, M)(\xi^*)$ to be outside $\bar{W}$ and proves condition (b).

To prove condition (c), suppose $\mathbf{y}(\xi, \mathbf{y}_0) \notin \partial W$. Then there exists a neighbourhood containing $\mathbf{y}(\xi, \mathbf{y}_0)$ which is disjoint from $W^* \subset \partial W$. On the other hand, the discussion following the proof of Theorem 3.1 implies that if $\mathbf{y}_0 \in \Sigma$ and $\mathbf{y}(\xi, \mathbf{y}_0) \in \partial W$, then $\mathbf{y}(\xi, \mathbf{y}_0) \in W^*$. Therefore condition (c) is satisfied.

We now prove (3.2) by showing that a closed loop around $B = (0, 0, K_3, 0)$ in W^- (like the σ we have just described) cannot be continuously deformed to a point while remaining in W^-. Assume the contrary and suppose $z(s, t)$: $[0, 1] \times [0, 1] \to W^-$ is a continuous deformation such that $z(s, 0)$ is the closed loop around B in W^- and $z(s, 1)$ is a point which lies in, say, $\{p = 1\}$. Then for some t, the loop $z(s, t)$ has to leave $\{q = 0\}$ (see Fig. 1) and this forces the set $\{M = 0\} \cap \{N = K_3\}$ to be non-empty, a contradiction. We leave the technical details to the reader.

Finally we have to prove condition (d). We already know that if $\mathbf{y}_0 \in \Sigma \subset \mathscr{S}$, then $\mathbf{y}(\xi, \mathbf{y}_0) \to A$ as $\xi \to -\infty$. Thus we only need to show that $\mathbf{y}(\xi, \mathbf{y}_0) \in W$ for all $\xi \leqq 0$. From the second and last equation of (1.10), we see that if $q(\xi^*) = 0$ or $M(\xi^*) = 0$, then since $\mathbf{y}(\xi^*, \mathbf{y}_0)$ is near $A = (1, 0, 0, 0)$, $q'(\xi^*) > 0$ and $M'(\xi^*) < 0$. Therefore $\mathbf{y}(\xi, \mathbf{y}_0)$ leaves W for $\xi < 0$ only through $\{p = 1\}$ or $\{N = 0\}$.

To show that $\mathbf{y}(\xi, \mathbf{y}_0)$ cannot reach $\{p = 1\}$, recall that $\lambda_1 > \lambda_2 > 0$ are the eigenvalues of the matrix $\begin{bmatrix} 1 & 1 \\ f(1, 0) & \theta \end{bmatrix}$. It is known that there is exactly one trajectory $(\bar{p}, \bar{q})$ to the system $\bar{p}' = \bar{q}$, $\bar{q}' = \theta\bar{q} - f(\bar{p}, 0)\bar{p}(1 - \bar{p})$ which approaches $(1, 0)$ as $\xi \to -\infty$ tangent to the eigenvector $\begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix}$ corresponding to the eigenvalue λ_1. We may assume that $(\bar{p}, \bar{q})$ lies in the sector between $-\mathbf{x}_1$ and $-\begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix}$ shown in Figure 2. Recall that Σ was chosen so that if $\mathbf{y}_0 \in \Sigma$, then $\mathbf{y}_0 = (p_0, q_0, N_0, M_0)$ is of the form (3.5) with α, β, γ and δ nonnegative. This implies that (p_0, q_0) lies in the sector between $-\mathbf{x}_1$ and $-\mathbf{x}_2$ in Figure 2. From the same figure, in order for $\mathbf{y}(\xi, \mathbf{y}_0)$ to reach $\{p = 1\}$, (p, q) of (2.9) must intersect $-\begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix}$ and hence $(\bar{p}, \bar{q})$. At the point of intersection, $\frac{dq}{dp} \leqq \frac{d\bar{q}}{d\bar{p}}$ which implies that $f(p, N) \leqq f(p, 0)$, a contradiction. Thus $\mathbf{y}(\xi, \mathbf{y}_0)$ never reaches $\{p = 1\}$ if $N > 0$ for $\xi \leqq 0$. A similar argument may be used to show that $\mathbf{y}(\xi, \mathbf{y}_0)$ never reaches $\{N = 0\}$ if $p < 1$ and $\xi \leqq 0$. Let $(\bar{N}, \bar{M})$ be the solution to $\bar{N}' = \bar{M}$, $\bar{M}' = \theta\bar{M} - g(1, \bar{N})\bar{N}$, approaching $(0, 0)$ as $\xi \to -\infty$ tangent to the eigenvector $\begin{bmatrix} 1 \\ \lambda_3 \end{bmatrix}$ corresponding to the largest positive eigenvalue λ_3 of the matrix $\begin{bmatrix} 0 & 1 \\ -g(1, 0) & \theta \end{bmatrix}$. Then $(\bar{N}, \bar{M})$ prevents $\mathbf{y}(\xi, \mathbf{y}_0)$ from reaching $\{N = 0\}$ (see Fig. 3). The proof of condition (d) is complete and so is Theorem 3.1. □

4. Stability of travelling waves

We first state two results which are needed to prove our stability theorem:

THEOREM 4.1. *Let u,v be bounded solutions of the reaction-diffusion system*

$$\begin{cases} u_t = d_1 u_{xx} + \tilde{f}(x, t, u, v) \\ v_t = d_2 v_{xx} + \tilde{g}(x, t, u, v) \end{cases} \quad \text{in } \mathbb{R} \times [0, T), \tag{4.1}$$

where $\tilde{f}$ is a decreasing function of v and g is an increasing function of u. Suppose α, β, γ and δ are bounded and smooth functions defined on $\mathbb{R} \times [0, T)$ such that $(\alpha(x, 0), \beta(x, 0)) \leqq (u(x, 0), v(x, 0)) \leqq (\gamma(x, 0), \delta(x, 0))$ and $(\alpha(x, t), \beta(x, t)) \leqq (\gamma(x, t), \delta(x, t))$. Suppose also that α, β, γ and δ satisfy the inequalities:

$$\begin{aligned} &\gamma_t - d_1\gamma_{xx} - \tilde{f}(x, t, \gamma, \beta) \geqq 0, \\ &\delta_t - d_2\delta_{xx} - \tilde{g}(x, t, \gamma, \delta) \geqq 0, \\ &\alpha_t - d_1\alpha_{xx} - \tilde{f}(x, t, \alpha, \delta) \leqq 0, \\ &\beta_t - d_2\beta_{xx} - \tilde{g}(x, t, \alpha, \beta) \leqq 0, \end{aligned} \tag{4.2}$$

on $\mathbb{R} \times [0, T)$. Then $(\alpha, \beta) \leqq (u, v) \leqq (\gamma, \delta)$ on $\mathbb{R} \times [0, T)$.

Proof. See the proof in [**9**].

LEMMA 4.2. *Let $w(x, t)$ satisfy the equation $w_t = w_{xx} + \lambda w$ and the initial conditions $w(x, 0) = C_1$ if $x > 0$, $w(x, 0) = C_2 e^{bx}$ if $x \leqq 0$, where λ is real and b, C_i are positive constants. Let $0 < \theta < 2b$. Then for each $L > 0$, $w(x_0 - \theta t, t) = 0\left[\exp\left(\lambda - \frac{\theta^2}{4}\right)t\right]$ as $t \to \infty$ uniformly for $x_0 \in [-L\sqrt{t}, L]$.*

Proof. We have

$$e^{-\lambda t} w(x, t) = \int_{-\infty}^{0} \frac{1}{\sqrt{4\pi t}} e^{-(x-y)^2/4t} C_2 e^{by}\, dy + \int_0^\infty \frac{C_1}{\sqrt{4\pi t}} e^{-(x-y)^2/4t}\, dy.$$

Let $z = -\dfrac{(x-y)}{\sqrt{2t}}$; then the second integral equals

$$\frac{C_1}{\sqrt{2\pi}} \int_{-x/\sqrt{2t}}^{\infty} e^{-z^2/2}\, dz \sim -C_1 \frac{\sqrt{2t}}{x} e^{-x^2/4t} \quad \text{as } \frac{x}{\sqrt{2t}} \to -\infty.$$

($\sim$ means ratio goes to one). Similarly, the first integral equals

$$\frac{C_2 e^{bx}}{\sqrt{2\pi}} \int_{-\infty}^{-x/\sqrt{2t}} e^{-z^2/2 + \sqrt{2t}\, bz}\, dz = \frac{C_2 e^{bx + b^2 t}}{\sqrt{2\pi}} \int_{-\infty}^{-(x+2bt)/\sqrt{2t}} e^{-v^2/2}\, dv.$$

If $\dfrac{x + 2bt}{\sqrt{2t}} \to \infty$, then the above integral is asymptotic to $\dfrac{C_2\sqrt{2t}}{x + 2bt} e^{-x^2/4t}$. If we write $x = x_0 - \theta t$, where $0 < \theta < 2b$, then as long as $\dfrac{x_0}{\sqrt{2t}}$ is bounded

$$e^{-\lambda t} w(x, t) \sim \left[\frac{C_2}{x + 2bt} - \frac{C_1}{x}\right] \sqrt{2t}\, e^{-x^2/4t} \quad \text{as } t \to \infty.$$

The coefficient in front of $e^{-x^2/4t}$ goes to zero as $t\to\infty$. Therefore $w(x_0-\theta t, t)=0\left[\exp\left(\lambda-\frac{\theta^2}{4}\right)t\right]$ as $t\to\infty$ if x_0 is bounded above. The proof of Lemma 4.2 is complete. □

We now turn to the proof of stability of travelling waves for the case (2.5). From condition (2.10), we have $\frac{\partial}{\partial N}[f(p, N)p(1-p)]\geqq 0$ and $\frac{\partial}{\partial p}[g(p, N)N]\leqq 0$ for $0\leqq p\leqq 1$, $N\geqq 0$. It is well known that comparison theorems are harder to apply in this case than in the case when both partial derivatives are nonnegative.

The travelling waves $\tilde{p}(\xi)$, $\tilde{N}(\xi)$ with speed θ constructed in Section 3 are facing in opposite directions. It is more convenient to let $r=1-p$ so that (1.2) becomes

$$\begin{aligned} r_t &= r_{xx}+F(r, N),\\ N_t &= N_{xx}+G(r, N), \end{aligned}\tag{4.3}$$

where $F(r, N)=-f(1-r, N)r(1-r)$ and $G(r, N)=g(1-r, N)N$. For each $\theta>\theta^*$, let $\tilde{r}(\xi)\equiv 1-\tilde{p}(\xi)$ and $\tilde{N}(\xi)$ denote a travelling wave solution of (4.3) that connects $(0, 0)$ to $(1, K_3)$. We do not require $\tilde{r}$ or $\tilde{N}$ to be monotone.

THEOREM 4.3. *Let r, N be solutions to* (4.3) *with initial data satisfying $0\leqq r(x, 0)\leqq 1$, $0\leqq N(x, 0)\leqq M$ for all x where $M>K_3$. Suppose that $|r(x, 0)-\tilde{r}(x)|\leqq Ce^{bx}$ and $|N(x, 0)-\tilde{N}(x)|\leqq Ce^{bx}$ for x sufficiently small. Let $k_1^*=\max|F_p(p, N)|$, $k_2^*=\max|F_N(p, N)|$, $h_1^*=\max|G_p(p, N)|$ and $h_2^*=\max|G_N(p, N)|$, where the maximum is taken over $0\leqq p\leqq 1$ and $0\leqq N\leqq M$. Let λ_1 be the largest positive eigenvalue of the matrix $A=\begin{bmatrix} k_1^* & k_2^*\\ h_1^* & h_2^*\end{bmatrix}$. Suppose θ, b satisfy the inequalities $2\sqrt{\lambda_1}<\theta<2b$. Then there exists a positive constant C^* such that*

$$|r(x_0-\theta t, t)-\tilde{r}(x_0)|\leqq C^*e^{(\lambda_1-\theta^2/4)t},\quad |N(x_0-\theta t, t)-\tilde{N}(x_0)|\leqq C^*e^{(\lambda_1-\theta^2/4)t}$$

uniformly for $x_0\in[-L\sqrt{t}, L]$ and $L>0$, as $t\to\infty$.

Proof. Let $r_1(x, t)=r(x, t)-\tilde{r}(x+\theta t)$ and $N_1(x, t)=N(x, t)-\tilde{N}(x+\theta t)$. Then r_1, N_1 satisfy the equations

$$\begin{cases} r_{1t}=r_{1xx}+k_1(x, t)r_1+k_2(x, t)N_1,\\ N_{1t}=N_{1xx}+h_1(x, t)r_1+h_2(x, t)N_1, \end{cases}\tag{4.4}$$

where $k_1(x, t)=[F(r(x, t), N(x, t))-F(\tilde{r}(x+\theta t), N(x, t))]/r_1(x, t)$, $k_2(x, t)=[F(\tilde{r}(x+\theta t), N(x, t))-F(\tilde{r}(x+\theta t), \tilde{N}(x+\theta t))]/N_1(x, t)$, and h_1, h_2 are similarly defined with F replaced by G. Since $g(p, M)$ is negative, the theory of invariant rectangles for parabolic systems implies that $0\leqq r(x, t)\leqq 1$, $0\leqq N(x, t)\leqq M$ for all x and $t>0$. Therefore $|k_i(x, t)|\leqq k_i^*$, $|h_i(x, t)|\leqq h_i^*$ for $i=1, 2$. Note that $k_2\leqq 0$ and $h_1\geqq 0$ because of conditions (2.1) and (2.8).

Let γ, δ satisfy the equations

$$\begin{cases} \gamma_t=\gamma_{xx}+k_1(x, t)\gamma-k_2(x, t)\delta,\\ \delta_t=\delta_{xx}+h_1(x, t)\gamma+h_2(x, t)\delta, \end{cases}\tag{4.5}$$

with initial data

$$\gamma(x,0)=\begin{cases}1 & \text{if } x>0\\ Ce^{bx} & \text{if } x\leqq 0\end{cases},\quad \delta(x,0)=\begin{cases}M & \text{if } x>0,\\ Ce^{bx} & \text{if } x\leqq 0.\end{cases}$$

Since $-k_2\geqq 0$ and $h_1\geqq 0$, the maximum principle implies that $\gamma(x,t)\geqq 0$ and $\delta(x,t)\geqq 0$ for all $t>0$. Define $\alpha=-\gamma$, $\beta=-\delta$. From our hypotheses, we may assume that $(\alpha(x,0),\beta(x,0))\leqq(r_1(x,0),N_1(x,0))\leqq(\gamma(x,0),\beta(x,0))$ for all x.

Let $\tilde{f}(x,t,u,v)=k_1(x,t)u+k_2(x,t)v$ and $\tilde{g}(x,t,u,v)=h_1(x,t)u+h_2(x,t)v$. Then $\tilde{f}$ is decreasing in v and $\tilde{g}$ is increasing in u. Also, from (4.5), α, β, γ and δ satisfy the inequalities listed in (4.2) with $d_1=d_2=1$. Therefore Theorem 4.1 implies that $(\alpha,\beta)\leqq(r_1,N_1)\leqq(\gamma,\delta)$ for all x and $t>0$. We now try to estimate γ and δ.

Let $\bar{\gamma}$, $\bar{\delta}$ be solutions to the $\bar{\gamma}_t=\bar{\gamma}_{xx}+k_1^*\bar{\gamma}+k_2^*\bar{\delta}$, $\bar{\delta}_t=\bar{\delta}_{xx}+h_1^*\bar{\gamma}+h_2^*\bar{\delta}$, $(\bar{\gamma}(x,0),\bar{\delta}(x,0))=(\gamma(x,0),\delta(x,0))$. Then the maximum principle implies that $(\gamma,\delta)\leqq(\bar{\gamma},\bar{\delta})$ for all x and $t>0$. To solve the above system for $\bar{\gamma}$ and $\bar{\delta}$, let A be defined as in Theorem 4.3. By increasing A, we may assume that A is diagonalisable so that there exists a matrix Γ such that $A\Gamma=\Gamma\Lambda$, where $\Lambda=\operatorname{diag}(\lambda_1,\lambda_2)$ and $\lambda_1>\lambda_2$ are the eigenvalues of A. Since A is a positive matrix, we have $\lambda_1>0$.

Let $\begin{bmatrix}\bar{\gamma}\\ \bar{\delta}\end{bmatrix}=\Gamma\begin{bmatrix}\mu\\ \nu\end{bmatrix}$. Then μ, ν satisfy $\mu_t=\mu_{xx}+\lambda_1\mu$, $\nu_t=\nu_{xx}+\lambda_2\nu$, and

$$\mu(x,0)=\begin{cases}C_1 & \text{if } x>0,\\ C_2e^{bx} & \text{if } x\leqq 0;\end{cases}\quad \nu(x,0)=\begin{cases}C_3 & \text{if } x>0,\\ C_4e^{bx} & \text{if } x\leqq 0.\end{cases}$$

From Lemma 4.2, there exist positive constants C_5, C_6 such that $\mu(x_0-\theta t,t)\leqq C_5\exp\left[\left(\lambda_1-\frac{\theta^2}{4}\right)t\right]$ and $\nu(x_0-\theta t,t)\leqq C_6\exp\left[\left(\lambda_2-\frac{\theta^2}{4}\right)t\right]$ uniformly for $x_0\in[-L\sqrt{t},L]$ as $t\to\infty$. Theorem 4.3 then follows by multiplying $\begin{bmatrix}\mu\\ \nu\end{bmatrix}$ by Γ. □

Remark 4.4. Since all the travelling wave solutions have the same limit as $|\xi|\to\pm\infty$, one cannot expect stability in the supremum norm. Theorem 4.3 simply says that the wave with speed θ is stable under perturbation by functions that decay faster than e^{bx} as $x\to-\infty$ if $2\sqrt{\lambda_1}<\theta<2\sqrt{b}$.

Remark 4.5. With some effort, one can show that $2\sqrt{\lambda_1}>\theta^*$ so that there is a gap, $\theta\in(\theta^*,2\sqrt{\lambda_1}]$, where stability of waves with speed θ has not been established. Note that we have used the fact that p and N diffuse at the same rate to uncouple $\bar{\gamma}$ and $\bar{\delta}$ in the proof of Theorem 4.3.

Remark 4.6. The inequalities satisfied by the initial data in Theorem 4.3 imply that $\frac{r(x,0)}{\tilde{r}(x)}\to 1$ and $\frac{N(x,0)}{\tilde{N}(x)}\to 1$ as $x\to-\infty$, since $\tilde{r}(x)$, $\tilde{N}(x)=0(e^{\sigma x})$ near $-\infty$; $\sigma\in\left[0,\frac{\theta}{2}\right]$ is the smallest of the four eigenvalues of the matrix obtained by linearising (2.9) at $(1,0,0,0)$. These two conditions on the initial data apparently are not enough to imply the results of Theorem 4.3. This is because Theorem 4.1

requires four sets of inequalities involving both the upper and lower comparison functions to hold simultaneously.

Acknowledgment

This research was partially supported by NSF grant DMS-8601585 and NSF grant DMS-8801968.

References

1 D. G. Aronson and H. F. Weinberger. Nonlinear diffusion in population genetics, combustion and nerve propatation. In *Partial Differential Equations and related topics*, ed. J. Goldstein, pp. 5–49. Lecture Notes in Mathematics 446 (New York: Springer, 1975).

2 S. R. Dunbar. Travelling wave solutions of diffusive Lotka-Volterra equations: a heteroclinic connection in $\mathbb{R}^4$. *Trans. Amer. Math. Soc.* **286** (1984), 557–594.

3 P. Fife. Mathematical aspects of reacting and diffusing systems. Lecture Notes in Biomathematics 28 (New York: Springer, 1979).

4 R. A. Fisher. The advance of advantageous genes. *Ann. Eugenics* **7** (1937), 355–369.

5 P. Hartman. *Ordinary differential equations* (New York: Wiley, 1964).

6 A. Kolmogoroff, I. Petrovsky and N. Piscounoff. Étude de l'équations de la diffusion avec croissance de la quantité de matieré et son application a un probleme biologique. *Bull. Univ. Moscow Ser. Internat. Sect. A* **1** (1937), 1–25.

7 R. Lui. Asymptotic behavior of solutions to an evolutionary ecology model with diffusion. *SIAM J. Applied Math.* **49** (1989), 1447–1461.

8 M. M. Tang and P. Fife. Propagating fronts for competing species equations with diffusion. *Arch. Rational Mech. Anal.* **73** (1980), 69–77.

9 D. Terman. *Comparison theorems for reaction-diffusion systems defined in an unbounded domain* (Madison: University of Wisconsin MRC Technical Report #2374, April, 1982).

10 W. C. Troy. The existence of travelling wave front solutions of a model of the Belousov-Zhabotinsky chemical reaction. *J. Differential Equations* **36** (1980), 89–98.

(*Issued* 28 *June* 1990)