Electronic Journal of Probability

Vol. 5 (2000) Paper no. 14, pages 1–40.

Journal URL
http://www.math.washington.edu/~ejpecp/
Paper URL
http://www.math.washington.edu/~ejpecp/EjpVol5/paper14.abs.html

TRAVELLING WAVES FOR A CERTAIN FIRST-ORDER COUPLED PDE SYSTEM

Owen D. Lyne
School of Mathematical Sciences, University of Nottingham
University Park, Nottingham, NG7 2RD, United Kingdom
owen.lyne@nottingham.ac.uk

Abstract. This paper concentrates on a particular first-order coupled PDE system. It provides both a detailed treatment of the *existence* and *uniqueness* of monotone travelling waves to various equilibria, by differential-equation theory and by probability theory and a treatment of the corresponding hyperbolic initial-value problem, by analytic methods.

The initial-value problem is studied using characteristics to show existence and uniqueness of a bounded solution for bounded initial data (subject to certain smoothness conditions). The concept of *weak* solutions to partial differential equations is used to rigorously examine bounded initial data with jump discontinuities.

For the travelling wave problem the differential-equation treatment makes use of a shooting argument and explicit calculations of the eigenvectors of stability matrices.

The probabilistic treatment is careful in its proofs of *martingale* (as opposed to merely local-martingale) properties. A modern *change-of-measure* technique is used to obtain the best lower bound on the speed of the monotone travelling wave — with Heaviside initial conditions the solution converges to an approximate travelling wave of that speed (the solution tends to one ahead of the wave-front and to zero behind it). Waves to different equilibria are shown to be related by Doob h-transforms. *Large-deviation theory* provides heuristic links between alternative descriptions of minimum wave speeds, rigorous algebraic proofs of which are provided.

Key words. Travelling waves. Martingales. Branching processses

AMS Classification. 35L35; 60J27; 60G44

Submitted to EJP on March 17, 1999. Final version accepted on August 17, 2000.

1 Introduction

Connections between parabolic and elliptic equations and diffusion processes are well-known. In this paper we explore the connections between hyperbolic partial differential equations and stochastic processes by treating the following system.

Let r_1, r_2, q_1, q_2 be positive (since they correspond to rates of probabilistic processes) constants and b_1, b_2 unrestricted real constants, fixed throughout. Let θ be a *positive* rate parameter. We consider an equation system related to the generalized FKPP system discussed in Champneys, Harris, Toland, Warren and Williams [2]. Here the system of interest is

$$\frac{\partial u}{\partial t} = B\frac{\partial u}{\partial x} + R(u^2 - u) + \theta Q u, \tag{1}$$

where u is a vector-valued function from $[0,\infty) \times \mathbb{R}$ to $\mathbb{R}^2$, $u = (u_1, u_2) \in \mathbb{R}^2$ and $u^2 = (u_1^2, u_2^2)$, and where

$$B := \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}, \qquad R := \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix}, \qquad Q := \begin{pmatrix} -q_1 & q_1 \\ q_2 & -q_2 \end{pmatrix}.$$

(We use ':=' to mean 'is defined to equal'.)

Dunbar [8] considered a similar system but with $q_1 = q_2$, $r_1 = r_2$ and the nonlinearity was $\frac{1}{4}(u_1 + u_2)^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ rather than u^2. Holmes [10] compared reaction-diffusion systems to reaction-telegraph models (again with symmetry in the nonlinearity in both terms) for animal movement; if equation (1) is rewritten as a single second-order PDE it can be seen that our system is a reaction-telegraph model, though we find it more convenient to work with the pair of first-order equations. Hadeler [9] discussed more general non-linearities but still retained the condition $q_1 = q_2$. We contrast the probabilistic interpretation of these models and ours in section 1.4.

A travelling-wave solution of equation (1) is a solution of the form $u(t, x) := w(x - ct)$ (where $w : \mathbb{R} \to \mathbb{R}^2$). w describes a travelling wave if and only if

$$(B + cI)w' + R(w^2 - w) + \theta Q w = 0. \tag{2}$$

The '*source point*' $S = (0, 0)$ and the '*target point*' $T = (1, 1)$ are clearly equilibria of equation (2). If $r_1 r_2 \geq 4\theta^2 q_1 q_2$, there will also be equilibria at the two points

$$E_\pm = \left(\frac{1}{2} + \theta\rho_1 \pm \sqrt{\Delta}, \frac{1}{2} + \theta\rho_2 \mp \sqrt{\Delta}\right),$$

where $\rho_i := q_i/r_i$ and $\Delta := \frac{1}{4} - \theta^2\rho_1\rho_2$.

For $\theta \in (0, \infty)$ we study the existence of monotone travelling waves from S to T. Waves from S to the other equilibria can be obtained through a transformation detailed in section 1.3. Monotone travelling waves from S to T have a direct probabilistic interpretation which we will explore in this paper, but whether such an intepretation of other travelling waves is possible is unclear.

1.1 Stability of equilibria

Suppose that $(B + cI)$ is invertible and write equation (2) in the form

$$\frac{\mathrm{d}w}{\mathrm{d}x} = F(w). \tag{3}$$

Then F is a quadratic polynomial. Let E be an equilibrium point of (2), thus $F(E) = 0$. Then write $w(x) - E = v(x)$ and expand equation (3) to first order in v. This yields

$$\frac{\mathrm{d}v}{\mathrm{d}x} = K(E)v, \qquad K_{ij}(E) := \frac{\partial F_i}{\partial w_j} \text{ evaluated at } w = E.$$

Thus $K_{c,\theta}(T)$ (we write $K_{c,\theta}$ to emphasize the dependence on c and θ) satisfies

$$(B + cI)K_{c,\theta}(T) + R + \theta Q = 0 \tag{4}$$

and $K_{c,\theta}(S)$ satisfies

$$(B + cI)K_{c,\theta}(S) - R + \theta Q = 0.$$

The stability properties of these matrices are investigated in section 3.

1.2 The main ODE theorem

DEFINITION. An eigenvalue λ of a real 2×2 matrix M will be called *stable* [respectively, *unstable*] *monotone* if

(i) λ is real and negative [positive], and

(ii) M has an eigenvector (v_1, v_2) corresponding to λ with $v_1 v_2 \geq 0$.

This definition links nicely with the Perron-Frobenius Theorem (see Seneta [15]). The theorem implies that a square matrix M with all off-diagonal entries strictly positive has a special eigenvalue $\Lambda_{PF}(M)$ with an associated eigenvector with all entries positive and such that every other eigenvalue of M has real part less than $\Lambda_{PF}(M)$. Moreover, any eigenvector with all entries positive must be a multiple of the Perron-Frobenius eigenvector.

As will be proven in section 3 (Lemma 3.1), for any fixed $\theta > 0$, there exists a critical value $c(\theta)$ in the interval $\min(-b_1, -b_2) \leq c(\theta) \leq \max(-b_1, -b_2)$ — with the property that if $c > c(\theta)$, $c \neq \max(-b_1, -b_2)$, then $K_{c,\theta}(T)$ has at least one stable monotone eigenvalue, and if $c < c(\theta)$, $c \neq \min(-b_1, -b_2)$, it has no such eigenvalues. Since a necessary condition for the existence of a monotone travelling wave of (2) which converges to T as $x \to \infty$ is the existence of a stable monotone eigenvalue of $K_{c,\theta}(T)$ this provides a lower bound on possible values of c for which a monotone travelling wave can exist. It is shown later that this condition on c is also sufficient for the existence of a monotone connection from S to T, see section 4.5. Large-deviation theory gives probabilistic heuristics for this critical value, see section 5.2. Thus we can state our main result as follows.

Theorem 1.1 *Suppose that $c > c(\theta)$, then there exists one and, modulo translation, only one monotone solution of equation* (2) *with $w(x) \to S$ as $x \to -\infty$ and $w(x) \to T$ as $x \to +\infty$. For $c < c(\theta)$ there is no such solution.*

This theorem is proven in section 4 using a shooting argument. Behaviour at $c(\theta)$ itself depends on the relative values of parameters as follows (these special cases are detailed in section 4.6):

- If $c(\theta)$ is in the interval $\min(-b_1, -b_2) < c(\theta) < \max(-b_1, -b_2)$ then there is a unique, monotone solution for $c = c(\theta)$;
- For $i = 1, 2$, if $c(\theta) = -b_i$, then there is a unique monotone solution for $c = c(\theta)$ if and only if $\theta q_i = r_i$.

1.3 A Doob h-transform

As in Champneys et al. [2] there is a *Doob h-transform* that maps monotone waves from S to one of the equilibria to monotone waves from S to another of the equilibria. This fact is summarised by the following Lemma.

Lemma 1.2 *Suppose that θ is fixed at a value where $\Delta \geq 0$, so that E_+ and E_- exist. If $E = (\alpha_1, \alpha_2)$ denotes either E_+ or E_-, then the substitution*

$$\tilde{q}_i := q_i \alpha_j / \alpha_i \quad (j \neq i), \quad \tilde{r}_i := r_i \alpha_i, \quad \tilde{u}_i := u_i / \alpha_i, \quad \tilde{w}_i := w_i / \alpha_i \tag{5}$$

transforms (1) and (2) into their $\sim$ versions, monotone waves from S to E for the original problem corresponding exactly to monotone waves from S to T for the $\sim$ problem. The possibility that $E_+ = E_-$ and $\Delta = 0$ is not excluded.

Much of our work therefore automatically transfers to the case when T is replaced by E_+ or E_-, though the critical values $c^{\pm}(\theta)$ of θ corresponding to waves from S to $E^{\pm}$ will be different.

1.4 Key probability theorems

Our aim in section 5 will be to prove the following theorems which connect a probability model, defined below, with the system (1). Note that we often switch between two equivalent notations in the probability part of this paper to reduce the use of subscripts. Thus we will sometimes write, for example, $b(y)$ for b_y, $w(x,y)$ for $w_y(x)$ and $u(t,x,y)$ for $u_y(t,x)$ (for $y = 1, 2$).

Let $I := \{1, 2\}$ and consider the following two-type branching system of particles. At time $t \geq 0$, there are $N(t)$ particles, the k-th particle — *in order of birth* — having *position* $X_k(t)$ in $\mathbb{R}$ and *type* $Y_k(t)$ in I. The *state* of the system at time t is therefore

$$\Big(N(t);\, X_1(t), \ldots, X_{N(t)}(t);\, Y_1(t), \ldots, Y_{N(t)}(t) \Big). \tag{6}$$

Particles, once born, behave independently of one another. Each particle lives forever. The type of a particle (once born) is an autonomous Markov chain on I with Q-matrix θQ. While a particle is of type $y \in I$, it moves with constant velocity $b(y)$, and it gives birth — *to one child each time, at its own current position and of its own current type* — in a Poisson process of rate $r(y)$. So, $r(y)$ is the breeding rate of type y.

The branching system in Dunbar [8] had $q_1 = q_2$, $r(1) = r(2)$ and particles, rather than giving birth to one particle of the same type and living on themselves, die and give birth to two particles of independent *random* types, with each having equal probability of being of either type. The equations studied by Hadeler [9] correspond to the new pair of particles having type correlated to that of their parents, for any correlation except ± 1. Our model permits $q_1 \neq q_2$, so that the particle-type Markov Chain can have any equilbrium distribution on I

For our model it makes no difference whether you consider that one new particle has been born and the old one lives on too, or that two new particles of the same type are born. However, when the type of the new particles is random, this distinction is crucial. Our model is thus distinct from those previously considered.

Write $\mathbb{P}_{x,y}$ (with associated expectation $\mathbb{E}_{x,y}$) for the law of this process when it starts from one particle of type $Y_1(0) = y$ at position $X_1(0) = x$. By martingale [respectively, local martingale, supermartingale, ...] we mean a process which is for every $\mathbb{P}_{x,y}$ a martingale [respectively, ...] relative to the natural filtration $\mathcal{F}_t$ ($\mathbb{P}_{x,y}$-augmented, to be precise) of the process at (6).

The state-space for this process is

$$\mathcal{S} := \bigcup_{n \geq 1} \Big(\{n\} \times \mathbb{R}^n \times I^n \Big). \tag{7}$$

Define $L(t) := \inf_{k \leq N(t)} X_k(t)$. This is the position of the left-most particle. The asymptotic speed of the left-most particle is $\lim_{t \to \infty} t^{-1} L(t)$.

Theorem 1.3 *As $t \to \infty$, the following holds almost surely* (a.s.)

$$t^{-1}L(t) \to -c(\theta). \tag{8}$$

If u satisfies the coupled system (1)*, for $t \geq 0$ and $x \in \mathbb{R}$ and if*

$$u(0,x,y) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

then for $t > 0$, $0 \leq u(t,x,y) \leq 1$, $u(t,x,y) = \mathbb{P}_{x,y}[L(t) > 0]$, and u is an approximate travelling wave of speed $c(\theta)$ in the sense that

$$u(t, x + \gamma t, y) \to \begin{cases} 0 & \text{if } \gamma < c(\theta), \\ 1 & \text{if } \gamma > c(\theta). \end{cases}$$

This theorem is proved, along with the following one, in sections 5.4 and 5.5. The theorem allows us to relate the speed of the spread of the particles in the probabilistic model with the wave speed of travelling waves. Specifically, we are claiming that the left-most particle travels, in the limit, at the speed $-c(\theta)$, where $c(\theta)$ is the critical speed above which a unique monotone travelling wave exists and below which no such wave exists.

An analytic proof that the weak solution of (1) is between 0 and 1 for the Heaviside initial data is included in section 2.4, as well as a proof that for continuous initial data between 0 and 1 the solution remains between 0 and 1 (see Lemma 2.2 and the subsequent remarks). The work of section 2 is not necessary for the probabilistic approach to the problem but adds insight from the viewpoint of classical analysis, and vice-versa. In section 5 we show that any (smooth) solution to the coupled system (1) that is between 0 and 1 has a McKean representation. We use this representation to motivate the (probabilistic) construction of a solution for the initial-value problem with Heaviside initial data. This constructed solution is then directly verified to satisfy the appropriate equations and does remain between 0 and 1.

Consider the case when $X_1(0) = 0$ and $Y_1(0) = 1$, that is, work with the $\mathbb{P}_{0,1}$ law: $\mathbb{P} := \mathbb{P}_{0,1}$. The terminology — *probabilistic* eigenvalue of $K_{c,\theta}(T)$ — is a shorthand explained fully after Theorem 5.2.

Theorem 1.4 (i) *Let $c > c(\theta)$. Let λ be the* probabilistic *eigenvalue of $K_{c,\theta}(T)$. Define*

$$Z_\lambda(t) := \sum_{k=1}^{N(t)} v_\lambda\big(Y_k(t)\big) \exp\Big\{ \lambda\big[X_k(t) + ct\big] \Big\},$$

with v_λ being the eigenvector (with $v_\lambda(1) = 1$) corresponding to $\Lambda_{PF}(\lambda)$, the Perron-Frobenius eigenvalue of $\lambda B + \theta Q + R$. The fact that Z_λ is a martingale (see the discussion after Theorem 5.2*) implies that*

$$\liminf_{t\to\infty} t^{-1}L(t) \geq \lambda^{-1}\Lambda_{PF}(\lambda) \quad \text{(a.s.)}.$$

(ii) *Since $Z_\lambda(\infty)$ exists in $\mathcal{L}^1$ (by Theorem* 5.4*) and $Z_\lambda(0) = 1$, we can define a measure Q_λ equivalent to $\mathbb{P}$ on $\mathcal{F}_\infty$ by*

$$dQ_\lambda/d\mathbb{P} = Z_\lambda(\infty) \text{ on } \mathcal{F}_\infty, \text{ whence } dQ_\lambda/d\mathbb{P} = Z_\lambda(t) \text{ on } \mathcal{F}_t.$$

Then

$$M_\lambda(t) := Z_\lambda(t)^{-1}\frac{\partial}{\partial\lambda}Z_\lambda(t)$$

defines a Q_λ-martingale, and

$$t^{-1}M_\lambda(t) \to 0 \quad \text{(a.s.)}.$$

This implies that

$$\limsup_{t\to\infty} t^{-1}L(t) \leq \frac{\partial}{\partial\lambda}\Lambda_{PF}(\lambda) \quad \text{(a.s.)}.$$

(iii) *As $c \downarrow c(\theta)$, we have*

$$\lambda^{-1}\Lambda_{PF}(\lambda) \to -c(\theta) \quad \text{and} \quad \frac{\partial}{\partial\lambda}\Lambda_{PF}(\lambda) \to -c(\theta),$$

so that (8) *follows.*

This result gives us the weaponry to prove Theorem 1.3 using martingale techniques. Note that Theorem 1.3 is proved using Theorem 1.4, and that in section 5.4 we prove Theorem 1.4 via the Theorems and Lemmas in the preceding sections of section 5 — hence various references to Theorems of section 5 in the statement of Theorem 1.4 are not circular.

1.5 Summary chart of results

This table is intended to help keep track of the various cases determined by the values of parameters. There is no monotone travelling wave from S to T if No appears in both right-hand columns. The inside and outside regions mentioned here are defined in section 4 where it is shown that if a monotone travelling wave exists then it lies either in an inside, or an outside region.

	Number of unstable *monotone* eigenvalues at S	Number of stable *monotone* eigenvalues at T	Monotone Connection through	
			inside region	outside region
$c > \max(-b_1, -b_2)$	1	1	Yes	No
$\max(-b_1, -b_2) > c > c(\theta)$	1	2	No	Yes
$c(\theta) > c > \min(-b_1, -b_2)$	1	0	No	No
$c < \min(-b_1, -b_2)$	0	0	No	No

When $c \in \{-b_1, -b_2\}$ the coupled system (1) becomes an ODE and an algebraic equation. The only candidate for a travelling wave from S to T is the corresponding segment of the solution of the algebraic equation (which is a parabola) — full details of this special case are in section 4.6.

2 Existence and uniqueness results for the PDE initial-value problem

To study the question of global existence and uniqueness of solutions to the initial-value problem for the system (1), it is convenient to change to moving coordinates (moving at a speed of $\frac{1}{2}(b_1 + b_2)$) and then re-scale space so that the coefficients of u_x are 1 and -1. This is possible unless $b_1 = b_2$ — we deal with this case in section 2.6. We use subscript notation to represent derivatives and relabel so that the system becomes:

$$u_t - u_x = \quad r_1(u^2 - u) + \theta q_1(v - u) =: \quad f(u, v); \tag{9}$$

$$v_t + v_x = \quad r_2(v^2 - v) + \theta q_2(u - v) =: \quad g(u, v). \tag{10}$$

The functions f and g are introduced to simplify notation.

We are particularly interested in the Cauchy problem for Heaviside initial data:

$$u(0, x) = v(0, x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

which we study in section 2.4, but first we study the Cauchy problem with C^1-initial data.

The characteristics are $x+t=$ constant, on which $\frac{\partial u}{\partial t}=f(u,v)$ and $x-t=$ constant, on which $\frac{\partial v}{\partial t}=g(u,v)$. Integrating along the characteristics from $t=0$ to a point (T,X) gives:

$$\begin{aligned}u(T,X) &= u(0,X+T)+\int_0^T f\big((u,v)(t,X+T-t)\big)dt;\\ v(T,X) &= v(0,X-T)+\int_0^T g\big((u,v)(t,X-T+t)\big)dt.\end{aligned}$$

We work in the Banach space C^1 of 2-vector-valued functions w for which w and w_x are continuous and bounded for all x, in which we choose as norm $|||w|||=\max(||w||,||w_x||)$, where $||w||:=\sup_{x\in\mathbb{R}}|w(x)|$, the usual L^∞-norm. These equations can be used with the Contraction Mapping Principle to prove existence and uniqueness of classical solutions for a short time for C^1-initial data. See Courant and Hilbert [5, pages 461–471] for details.

2.1 Proofs for smooth initial data

To go from short-time existence and uniqueness of solutions to global existence and uniqueness we prove the following lemmas which give bounds on the solutions for all time (for a certain class of initial data). These bounds then allow iteration of the Contraction Mapping argument — hence local existence and uniqueness become global. This iteration is done by using the solution obtained from local existence and uniqueness, up until a small, fixed time, τ, then taking the value of this solution at time τ as initial data, and repeating the argument. The bounds obtained in the following two lemmas allow repeated use of the same τ at each step, rather than having to take a sequence of τ_n (whose sum may converge to a finite *blow-up* time), thus we obtain existence and uniqueness for all time. To set our notation, note that we will write $u(0,x)=u_0(x), v(0,x)=v_0(x)$ for $-\infty<x<\infty$.

Lemma 2.1 *If $0<u_0(x),v_0(x)<K\le 1$ for all x, and (u_0,v_0) is in C^1, then C^1 solutions (u,v) of equations* (9), (10) *satisfy $0<u,v<K$ for all time.*

Proof. Suppose, for a contradiction, that there exists a point (T,X) where (u,v) is outside the square, $(0,K)^2$. The value of (u,v) at this point only depends on the initial data in the interval $[X-T,X+T]$ — this is the domain of dependence (see Courant and Hilbert [5, pages 438–440]). This data determines the solution (u,v) throughout the closed triangle, which we shall denote by Ω, of (t,x)-space whose corners are $(T,X),(0,X-T)$ and $(0,X+T)$.

Since Ω is compact and u and v are continuous, for each of the possible violations (that is, violations of the four inequalities $u<K$, $u>0$, $v<K$ and $v>0$) we can find a *first time* it occurs in Ω. For example, if $u\ge K$ at some point in Ω then there exists a point $(t_0,x_0)\in\Omega$ such that $u(t_0,x_0)=K$ and, for all $(t,x)\in\Omega$ with $t<t_0$, $u(t,x)<K$. Similarly we can find a time (and corresponding spatial position) where the first violations of $u>0$, $v<K$ and $v>0$ occur in Ω (if such violations do occur). We can then study the *first violation* that happens, by taking the one corresponding to the minimum of these 4 times (taking the time to be $T+1$ if it does not occur in Ω). This is well-defined because we know there is at least 1 violation and at most 4.

So, consider each possible *first violation* in turn. Firstly, that there exists a point $(t_0,x_0)\in\Omega$ such that $u(t_0,x_0)=K$ and, for all $(t,x)\in\Omega$ with $t<t_0$, $\big(u(t,x),v(t,x)\big)\in(0,K)^2$. Then consider u restricted to the characteristic $x=x_0+t_0-t$ through (t_0,x_0) (which lies entirely in Ω). From (9), with $u=u(t,x_0+t_0-t)$,

$$\begin{aligned}\frac{\partial u}{\partial t}=r_1(u^2-u)+\theta q_1(v-u) &< r_1(u^2-u)+\theta q_1(K-u)\\ &\le \theta q_1(K-u).\end{aligned}$$

Therefore, $u \leq K + (u_0 - K)\exp(-\theta q_1 t) < K$ where $u_0 = u(0, x_0 + t_0)$, for $0 \leq t \leq t_0$, which contradicts $u(t_0, x_0) = K$.

Similarly, we can show that $v(t,x) < K$ for all $(t,x) \in \Omega$, and we use a similar argument to show $u > 0$ and $v > 0$:

$$\frac{\partial u}{\partial t} \geq -(r_1 + \theta q_1)u,$$

so $u \geq u_0 \exp\big(-(r_1 + \theta q_1)t\big) > 0$ if $u_0 > 0$.

Hence no violation occurs in Ω which is a contradiction. This completes the proof. $\square$

Lemma 2.1 allows us to prove the following:

Lemma 2.2 *If $0 \leq u_0(x), v_0(x) \leq K$ for all x, for some constant, $0 < K \leq 1$, and u_0, v_0 are in C^1, then there exists a unique (global) C^1 solution (u,v) of equations* (9), (10) *satisfying $0 \leq u, v \leq K$ for all time.*

Proof. Consider a sequence $(u^n, v^n)(x)$ of initial data satisfying the conditions of Lemma 2.1, which converges in C^1 to $(u_0, v_0)(x)$. That is, each $(u^n, v^n) \in C^1$, $0 < u^n, v^n < K$,

$$\sup_x \left(|u_0 - u^n|, |v_0 - v^n|\right) \to 0 \text{ as } n \to \infty$$

and

$$\sup_x \left(\left| \frac{\partial u_0}{\partial x} - \frac{\partial u^n}{\partial x} \right|, \left| \frac{\partial v_0}{\partial x} - \frac{\partial v^n}{\partial x} \right| \right) \to 0 \text{ as } n \to \infty.$$

Up to any fixed time τ, solutions $u^{(n)}(t,x), v^{(n)}(t,x)$ satisfy $0 < u^{(n)}, v^{(n)} < K$ by Lemma 2.1, so that, $\lim_{n\to\infty}(u^{(n)}, v^{(n)}) = (u,v)$ lies in $[0,K]^2$. The fact that this limit exists and is a unique solution is standard (see Courant and Hilbert [5, pages 467–468]). Since τ is arbitrary the result is true for all time. $\square$

This lemma completes the proof of global existence and uniqueness of bounded solutions for smooth initial conditions between 0 and any constant $0 < K \leq 1$. This upper limit on K is the best possible, since $u^2 - u > 0$ for $u > 1$, so the solution tends to grow. Indeed, it is clear that for $u_0(x) = v_0(x)$ identically equal to $1 + \epsilon$, for any $\epsilon > 0$, the solution blows up in finite time.

However it is possible to extend the above lemmas to deal with initial data below 0, since the $u^2 - u$ nonlinearity will tend to push the solution up towards 0.

Lemma 2.3 *If $-\infty < K < u_0(x), v_0(x) < 0$ for all x, and (u_0, v_0) is in C^1, then C^1 solutions (u,v) of equations* (9), (10) *satisfy $K < u, v < 0$ for all time.*

Proof. We can again look for the *first violation*, this time of the four restrictions $u > K, u < 0, v > K, v < 0$, and look at a characteristic going through a point at which the first violation occurs.

For the case $u = K$, note that

$$\begin{aligned} \frac{\partial u}{\partial t} = r_1(u^2 - u) + \theta q_1(v - u) \quad &> \quad \theta q_1(v - u) \\ &> \quad \theta q_1(K - u), \end{aligned}$$

so that u does not hit K.

For $u = 0$, note that, looking at the characteristic sufficiently close to the violation point (so that $u > -1$)

$$\begin{aligned} \frac{\partial u}{\partial t} = r_1(u^2 - u) + \theta q_1(v - u) \quad &< \quad r_1(-2u) + \theta q_1(-u) \\ &< \quad -(2r_1 + \theta q_1)u, \end{aligned}$$

so that u does not hit 0 either.

Similarly for v. Thus all the contraction mapping arguments extend to all time in the same way that the bounds on data between 0 and K (for $0 < K \leq 1$) extended existence and uniqueness in that case. $\square$

In the same way that we passed from Lemma 2.1 to Lemma 2.2 we can relax the strict inequalities in the hypotheses of Lemma 2.3 to obtain the following result.

Lemma 2.4 *If $-\infty < K \leq u_0(x), v_0(x) \leq 0$ for all x, for some constant $K < 0$, and (u_0, v_0) is in C^1, then C^1 solutions (u, v) of equations* (9), (10) *satisfy $K \leq u, v \leq 0$ for all time.*

Combining Lemma 2.2 and Lemma 2.4 prepares the ground for the following result.

Lemma 2.5 *If $K_1 < u_0(x), v_0(x) < K_2$ for all x, for some constants $-\infty < K_1 \leq 0$ and $0 \leq K_2 \leq 1$, and (u_0, v_0) is in C^1, then C^1 solutions (u, v) of equations* (9), (10) *satisfy $K_1 < u, v < K_2$ for all time.*

Proof. If $K_1 = 0$ then this result is simply a restatement of Lemma 2.2, and Lemma 2.4 deals with the case $K_2 = 0$. So, we may assume that $K_1 < 0 < K_2$.

For a violation such as, say, $u = K_1$, note that, considering a section of the characteristic sufficiently close to the violation point for $u < 0$

$$\begin{aligned} \frac{\partial u}{\partial t} = r_1(u^2 - u) + \theta q_1(v - u) \quad &> \quad \theta q_1(v - u) \\ &> \quad \theta q_1(K_1 - u), \end{aligned}$$

so that u does not hit K_1.

For $u = K_2$, note that, considering the characteristic sufficiently close to the violation point (so that $u > 0$)

$$\begin{aligned} \frac{\partial u}{\partial t} = r_1(u^2 - u) + \theta q_1(v - u) \quad &< \quad \theta q_1(v - u) \\ &< \quad \theta q_1(K_2 - u), \end{aligned}$$

so that u does not hit K_2.

Similarly for v, hence we are done. $\square$

Finally, again using the inequality relaxation, and noting that if the initial data is identically zero then the solution is identically zero for all time, we have:

Lemma 2.6 *If $K_1 \leq u_0(x), v_0(x) \leq K_2$ for all x, for some constants $-\infty < K_1 \leq 0$ and $0 \leq K_2 \leq 1$, and (u_0, v_0) is in C^1, then C^1 solutions (u, v) of equations* (9), (10) *satisfy $K_1 \leq u, v \leq K_2$ for all time.*

No such result will be true for a pair of constants K_1 and K_2 both strictly on the same side of zero — the solution to the initial value problem with $u_0 = v_0$ identically equal to some constant K, such that $-\infty < K < 1$, tends monotonically to zero (see section 2.6). This fact is put together with comparison arguments in section 2.2 to give a much stronger result than Lemma 2.6.

Given bounded initial data with an upper bound no greater than 1, we can read off the appropriate values of K_1 and K_2 by defining:

$$\begin{aligned} K_1 &= \min\Big(0, \inf_x\big(u_0(x)\big), \inf_x\big(v_0(x)\big)\Big), \\ K_2 &= \max\Big(0, \sup_x\big(u_0(x)\big), \sup_x\big(v_0(x)\big)\Big). \end{aligned}$$

Local existence of solutions with piecewise smooth initial data (i.e. continuous but with a finite number of jump discontinuities in the x-derivative) follows by approximating (in the supremum norm) piecewise smooth data by C^1 data — the limiting solutions are classical except on characteristics $x \pm t = x_0$ propagating from points x_0 of discontinuity of $\frac{\partial u_0}{\partial x}, \frac{\partial v_0}{\partial x}$.

2.2 Comparison results

Consider two solutions (u, v) and $(\tilde{u}, \tilde{v})$ of the equations (9),(10), with $u(0, x) = u_0(x)$, $v(0, x) = v_0(x)$, $\tilde{u}(0, x) = \tilde{u}_0(x)$, $\tilde{v}(0, x) = \tilde{v}_0(x)$. On the characteristics of the form $x + t =$ constant, we know that

$$\begin{aligned} \frac{\partial u}{\partial t} &= r_1(u^2 - u) + \theta q_1(v - u), \\ \frac{\partial \tilde{u}}{\partial t} &= r_1(\tilde{u}^2 - \tilde{u}) + \theta q_1(\tilde{v} - \tilde{u}), \end{aligned}$$

and similar equations for v and $\tilde{v}$ along the other characteristics.

Studying the difference between the solutions along a characteristic $x+t =$ constant, we see that it obeys the equation

$$\begin{aligned} \frac{\partial(u - \tilde{u})}{\partial t} &= r_1\big((u^2 - \tilde{u}^2) - (u - \tilde{u})\big) + \theta q_1\big((v - \tilde{v}) - (u - \tilde{u})\big), \\ &= (u - \tilde{u})\big(r_1(u + \tilde{u} - 1) - \theta q_1\big) + \theta q_1(v - \tilde{v}). \end{aligned}$$

Provided that the initial data (u_0, v_0) and $(\tilde{u}_0, \tilde{v}_0)$ satisfies the conditions of Lemma 2.6, then $(u + \tilde{u})$ will be bounded below (for all time), enabling us to write, along the characteristic up until a putative equality of u and $\tilde{u}$,

$$\frac{\partial(u - \tilde{u})}{\partial t} \geq C(u - \tilde{u}) + \theta q_1(v - \tilde{v}),$$

for some constant C. There is a similar equation for the difference of v and $\tilde{v}$.

Thus, if $u_0 > \tilde{u}_0$ and $v_0 > \tilde{v}_0$, then, for all time, $u > \tilde{u}$ and $v > \tilde{v}$. By taking sequences of initial data we obtain the following result.

Lemma 2.7 *If (u, v) and $(\tilde{u}, \tilde{v})$ are C^1 solutions of equations (9),(10) with $u_0 \geq \tilde{u}_0$ and $v_0 \geq \tilde{v}_0$, then, for all time, $u \geq \tilde{u}$ and $v \geq \tilde{v}$.*

Since the solution to the initial value problem with $u_0 = v_0$ identically equal to some constant K, such that $-\infty < K < 1$, tends monotonically to zero (see section 2.6), then any solution bounded between $-\infty < K_1 \leq 0$ and $0 \leq K_2 < 1$ will tend to zero. Thus Lemma 2.6 can be modified to the following result.

Lemma 2.8 *If $K_1 \leq u_0(x), v_0(x) \leq K_2$ for all x, for some constants $-\infty < K_1 \leq 0$ and $0 \leq K_2 < 1$, and (u_0, v_0) is in C^1, then C^1 solutions (u, v) of equations (9), (10) satisfy $K_1 \leq u, v \leq K_2$ for all time and $u(t, x)$ and $v(t, x) \to 0$ (uniformly in x) as $t \to 0$.*

2.3 Weak solutions and the Rankine-Hugoniot jump conditions

To deal with discontinuous initial data it is necessary to utilize the concept of *weak* solution.

DEFINITION. A *test function* is a 2-vector-valued function $\Phi := \big(\phi_1(t, x), \phi_2(t, x)\big)$ such that each of ϕ_1 and ϕ_2 is infinitely differentiable with compact support (i.e. $\phi_1, \phi_2 \in C_0^\infty$).

DEFINITION. For u, v bounded and measurable, we say that $U := (u, v)$ is a *weak solution* of equation (1) if, for all *test functions* Φ,

$$\int_{t=0}^{\infty}\int_{x=-\infty}^{\infty}\Big(U\cdot\Phi_t - BU\cdot\Phi_x + \big(R(U^2-U)+\theta QU\big)\cdot\Phi\Big)dtdx + \int_{x=-\infty}^{\infty}U(0,x)\cdot\Phi(0,x)dx = 0. \quad (11)$$

For a piecewise classical solution to be a weak solution its curves of discontinuity in the xt-plane must satisfy certain conditions, known as the *jump* or *Rankine-Hugoniot conditions*. For a semi-linear system such as (1) it is well known that these conditions imply that a component of the weak solution can only have discontinuities across characteristics corresponding to that component (shock paths can only be characteristic curves).

Moving to the weak setting requires care since jump conditions are not necessarily sufficient to guarantee uniqueness of solutions. Lemma 2.9 proves existence and uniqueness of *weak* solutions for a certain class of bounded, measurable initial data.

2.4 Heaviside initial data

We now construct explicitly a piecewise classical solution for Heaviside initial data that satisfies the Rankine-Hugoniot jump conditions (hence it is a weak solution and will be shown to be unique by Lemma 2.9).

For the Heaviside initial data the jump conditions reduce to two requirements — that the discontinuity in u propagates along the characteristic for u that goes through zero, i.e. u jumps across $x = -t$ and is continuous elsewhere, and that the discontinuity in v propagates along the characteristic for v that goes through zero, i.e. v jumps across $x = t$ and is continuous elsewhere. We have defined the Heaviside function so as to be left-continuous and construct so that our solution inherits this property — this matches the continuity which the probabilistic approach implies. The probabilistic interpretation of this solution is discussed in section 5.5.

Clearly the solution for the Heaviside data is identically zero for $x \le -t$ and identically one for $x > +t$.

We are constructing a left-continuous solution — therefore on $x = +t$, $u = 1$ while v remains to be calculated.

We integrate along the $x = +t$ characteristic to find v there. Since $u = 1$,

$$\frac{\partial v(t,+t)}{\partial t} = r_2(v^2-v) + \theta q_2(1-v)$$

and $v(0,0) = 0$. Hence, when $r_2 \neq \theta q_2$,

$$v(t,+t) = \frac{\theta q_2\Big(\exp\big((r_2-\theta q_2)t\big)-1\Big)}{r_2\exp\big((r_2-\theta q_2)t\big)-\theta q_2},$$

and v increases from 0 to $\min\left(\frac{\theta q_2}{r_2}, 1\right)$ as t goes from 0 to ∞. When $r_2 = \theta q_2$,

$$v(t,+t) = \frac{r_2 t}{r_2 t + 1},$$

and v increases from 0 to 1 as t goes from 0 to ∞.

We can also investigate the discontinuity in u along $x = -t$ similarly. v is continuous across this characteristic and zero on it, so is zero on $x = -t+$. Thus, integrating along the inside edge of the characteristic,

$$\frac{\partial u(t,-t+)}{\partial t} = r_1(u^2-u) - \theta q_1 u$$

and $u(0,0+)=1$. Hence,

$$u(t,-t+)=\frac{r_1+\theta q_1}{r_1+\theta q_1\exp\bigl((r_1+\theta q_1)t\bigr)},$$

and u decreases from 1 to 0 as t goes from 0 to ∞.

We now know the values of both u and v on the inside edge of the wedge $|x|<t$, between the discontinuities — and this data is continuous and between 0 and 1. Thus there is a unique solution to the problem with these as initial/boundary values and this solution is between 0 and 1. It can be found by following characteristics $x+t=$ constant from the $x=+t$ characteristic. This solution is piecewise classical with the only discontinuities being across the characteristics as required and so is indeed a weak solution.

Results of computational work on this initial-value problem will be reported elsewhere.

2.5 Discontinuous initial data

We follow the method of Beale [1] in his work on the Broadwell model. Let ϕ be a test function with

$$\phi\in C_0^\infty,\qquad \phi\geq 0,\qquad \int_{-\infty}^{\infty}\phi(x)dx=1,$$

and let $\phi_k(x)=k\phi(kx)$, $k\in\mathbb{N}$. For non-smooth initial data (u_0,v_0) we consider initial data formed by mollifying (u_0,v_0). We use the convolutions $u_0^{(k)}=u_0*\phi_k$ and $v_0^{(k)}=v_0*\phi_k$ and check that the corresponding solutions of the PDE converge as $k\to\infty$. Note that if u_0 and v_0 are between constants K_1 and K_2, then so are the mollified versions, $u_0^{(k)}$ and $v_0^{(k)}$.

In the following argument we consider differences between solutions for different mollifications of the same measurable initial data. We work with initial data that is bounded between constants K_1 and K_2, such that $-\infty<K_1\leq 0$ and $0\leq K_2\leq 1$, i.e. $\bigl(u_0(x),v_0(x)\bigr)\in[K_1,K_2]^2$ for all x. For the argument to work these differences between solutions should be in L^1, which we show follows if the difference between two different mollifications of the initial data is in L^1. This is clearly true for the Heaviside data, and for initial data that is itself in L^1. Then, for example, it is true for bounded initial data that only differs from the Heaviside function by an L^1-function.

With this in mind define a *step function* s to be a function of the form

$$s(x)=\begin{cases}k_1 & \text{if } x>0,\\ k_2 & \text{if } x\leq 0,\end{cases}$$

for some constants k_1 and k_2 such that $-\infty<k_1,k_2\leq 1$. Then if both components of the initial data differ from step functions (with possibly different constants) by L^1-functions, the difference between mollifications will be L^1.

Lemma 2.9 *Let $\bigl(u_0(x),v_0(x)\bigr)\in[K_1,K_2]^2$ for all x, such that each of $u_0(x)$ and $v_0(x)$ differs from a step function by an L^1-function. Define $u_0^{(k)}=u_0*\phi_k$ and $v_0^{(k)}=v_0*\phi_k$ as above and let $(u^{(k)},v^{(k)})$ be the solution of (1) with $\bigl(u^{(k)}(0),v^{(k)}(0)\bigr)=(u_0^{(k)},v_0^{(k)})$. Then the $u^{(k)}$ and $v^{(k)}$ are between K_1 and K_2 for all t,x and, for $1\leq p<\infty$, for $\tau>0$ arbitrary, $(u^{(k)}-u^{(1)},v^{(k)}-v^{(1)})$ converges in $C(0,\tau;L^p)$ to $(u-u^{(1)},v-v^{(1)})$ where (u,v) is a weak solution of (1) with $\bigl(u(0),v(0)\bigr)=(u_0,v_0)$. This weak solution is also bounded by K_1 and K_2 and is unique.*

Proof. By Lemma 2.6, $K_1\leq u^{(k)}(t,x)\leq K_2$ and $K_1\leq v^{(k)}(t,x)\leq K_2$. Now, if $U=(u^{(k)}-u^{(k')},v^{(k)}-v^{(k')})$, then, for $i=1,2$ and integers k,k',

$$U_{i,t}=(-1)^{i-1}U_{i,x}+h_i(t,x),\qquad |h_i(t,x)|\leq C|U(t,x)|, \tag{12}$$

where C depends on the parameters θ, q_1, q_2, r_1 and r_2, but not on k, k'. Define f_n by $f_n(y) = |y| - 1/(2n)$ for $|y| \geq 1/n$, $f_n(y) = ny^2/2$ for $|y| \leq 1/n$. Note that, for each n, f_n is continuous with continuous first derivative. Also $0 \leq f_n(y) \leq |y|$, and $f_n(y) \to |y|$ as $n \to \infty$.
Multiplying equation (12) by $f_n'(U_i)$, yields:

$$\big(f_n(U_i)\big)_t = (-1)^{i-1}\big(f_n(U_i)\big)_x + f_n'(U_i)h_i(t,x).$$

Integrating in x over a finite interval $[x_1, x_2]$ gives, for $i \in \{1, 2\}$,

$$\begin{aligned}
\frac{\partial}{\partial t}\left(\int_{x_1}^{x_2} f_n(U_i)dx\right) &\leq (-1)^{i-1}\left(f_n\big(U_i(t,x_2)\big) - f_n\big(U_i(t,x_1)\big)\right) \\
&+ C\int_{x_1}^{x_2} |U_i(t,x)|dx.
\end{aligned}$$

We now integrate on $[0, t]$ to get

$$\begin{aligned}
\int_{x_1}^{x_2} f_n\big(U_i(t,x)\big)dx &\leq \int_{x_1}^{x_2} f_n\big(U_i(0,x)\big)dx + C\int_0^t\int_{x_1}^{x_2} |U_i(s,x)|dxds \\
&+ (-1)^{i-1}\int_0^t \Big(f_n\big(U_i(s,x_2)\big) - f_n\big(U_i(s,x_1)\big)\Big)ds.
\end{aligned}$$

Letting $n \to \infty$ and taking limits using the bounded convergence theorem implies that

$$\begin{aligned}
\int_{x_1}^{x_2} |U_i(t,x)|dx &\leq \int_{x_1}^{x_2} |U_i(0,x)|dx + C\int_0^t\int_{x_1}^{x_2} |U_i(s,x)|dxds \\
&+ (-1)^{i-1}\int_0^t \big(|U_i(s,x_2)| - |U_i(s,x_1)|\big)ds \qquad (13)\\
&\leq \int_{x=-\infty}^{\infty} |u_{i,0}^{(k)} - u_{i,0}^{(k')}|dx + C\int_0^t\int_{x_1}^{x_2} |U_i(s,x)|dxds \\
&+ 2t||U_i||_{L^\infty[\mathbb{R}\times[0,t]]},
\end{aligned}$$

where in the latter step we are using the fact that the initial data is only an L^1-function from a step function to say that $\int_{x=-\infty}^{\infty} |u_0^{(k)} - u_0^{(k')}|dx$ is finite.
Hence, for any T,

$$\begin{aligned}
\sup_{t\in[0,T]}\int_{x_1}^{x_2} |U_i(t,x)|dx &\leq \int_{x=-\infty}^{\infty} |u_0^{(k)} - u_0^{(k')}|dx + CT\left(\sup_{t\in[0,T]}\int_{x_1}^{x_2} |U_i(t,x)|dx\right) \\
&+ 2T||U_i||_{L^\infty[\mathbb{R}\times[0,T]]}.
\end{aligned}$$

Choosing T sufficiently small so that $CT < 1$ and rearranging gives,

$$(1 - CT)\left(\sup_{t\in[0,T]}\int_{x_1}^{x_2} |U_i(t,x)|dx\right) \leq \int_{x=-\infty}^{\infty} |u_0^{(k)} - u_0^{(k')}|dx + 2T||U_i||_{L^\infty[\mathbb{R}\times[0,T]]},$$

where the constant on the right-hand side (which we will now denote by γ) does not depend on x_1 and x_2 (γ is finite due to boundedness of solutions for smooth, bounded initial data). Thus, for $0 \leq t < T$,

$$\int_{-\infty}^{\infty} |U_i(t,x)|dx \leq \frac{\gamma}{1 - CT} < \infty,$$

and hence each $U_i(t,\cdot) \in L^1((0,T)\times\mathbb{R})$. Now it is possible to return to equation (13) and improve our estimate. Firstly, rewrite it as:

$$\begin{aligned}\int_{x_1}^{x_2} |U_i(t,x)|dx &\leq \int_{x_1}^{x_2} |U_i(0,x)|dx + C\int_0^t \int_{x_1}^{x_2} |U_i(s,x)|dxds \\ &+ \int_0^t |U_i(s,x_2)|ds + \int_0^t |U_i(s,x_1)|ds. \end{aligned} \tag{14}$$

Since each $U_i(t,\cdot) \in L^1((0,T)\times\mathbb{R})$ there exist sequences such that, letting $x_1 \to -\infty$ and $x_2 \to \infty$ along these sequences, keeping t fixed, the latter two terms of equation (14) tend to zero. The fact that each $U_i(t,\cdot) \in L^1((0,T)\times\mathbb{R})$ keeps the other terms finite as $x_1 \to -\infty$ and $x_2 \to \infty$.

Hence:

$$\int_{x=-\infty}^{\infty} |U(t,x)|dx \leq \int_{x=-\infty}^{\infty} |u_0^{(k)} - u_0^{(k')}|dx + C\int_0^t \int_{x=-\infty}^{\infty} |U(s,x)|dxds.$$

Thus, from Gronwall's inequality, $|U(t,\cdot)|_{L^1} \leq e^{Ct}|u_0^{(k)} - u_0^{(k')}|_{L^1}$. Since $(u_0^{(k)} - u_0^{(1)}, v_0^{(k)} - v_0^{(1)}) \to (u_0 - u_0^{(1)}, v_0 - v_0^{(1)})$ in $L^1(\mathbb{R})$, it follows that $(u^{(k)} - u^{(1)}, v^{(k)} - v^{(1)})$ converges in $C\big(0,\tau;L^1(\mathbb{R})\big)$ to a function $w(t,x) = \big(w_1(t,x), w_2(t,x)\big)$. In fact, since the $u^{(k)}$ and $v^{(k)}$ are all bounded between K_1 and K_2, the convergence takes place in $C\big(0,\tau;L^p(\mathbb{R})\big), 1 \leq p < \infty$. Then it is easily seen that $\big(u(t,x), v(t,x)\big) := w(t,x) + \big(u^{(1)}(t,x), v^{(1)}(t,x)\big)$ is a weak solution of the equation and that (u,v) is again bounded by K_1 and K_2 for all time.

We now verify the uniqueness property. For these solutions we can show, for u that, for arbitrary $\phi \in C_0^\infty(\mathbb{R})$ and $t > 0$,

$$\int_{-\infty}^{\infty} u(t,x+t)\phi(x)dx - \int_{-\infty}^{\infty} u_0(x)\phi(x)dx =$$

$$\int_{-\infty}^{\infty}\int_0^t \Big(r_1\big(u(s,x+s)^2 - u(s,x+s)\big) + \theta q_1\big(v(s,x+s) - u(s,x+s)\big)\Big)\phi(x)dsdx$$

and similarly for v.

Therefore

$$u(t,x+t) - u_0(x) =$$

$$\int_0^t \Big(r_1\big(u(s,x+s)^2 - u(s,x+s)\big) + \theta q_1\big(v(s,x+s) - u(s,x+s)\big)\Big)ds$$

for almost all (t,x). Hence if there are two solutions with the same initial data and $y(t)$ is the L^∞-norm of the difference at time t, we obtain (using the boundedness between K_1 and K_2) an estimate

$$y(t) \leq c\int_0^t y(s)ds, \qquad y(0) = 0,$$

and this implies y is identically zero. □

The restriction to initial data that is an L^1 function different from a step function can be removed using domains of dependence and a truncation argument as follows.

The value of u on the line (T,X) to $(T,X+1)$ depends only on the initial data in the interval $[X-T, X+T+1]$. Thus it should agree with the solution whose initial data is identically zero outside this interval, and matches on the interval, which we shall describe as the *truncated* initial data. However the truncated initial data satisfies the conditions of Lemma 2.9 and so the truncated initial data has a unique corresponding solution. Thus we can *define* the solution for more general initial data to be that constructed by piecing together solutions for the truncated initial data, Lemma 2.9 guarantees that this will be well-defined.

2.6 The case when $b_1 = b_2$

When $b_1 = b_2$ we change to moving coordinates at speed b_1. Relabelling as before we obtain the ODE:

$$\begin{aligned} u_t &= r_1(u^2-u)+\theta q_1(v-u) = f(u,v); \\ v_t &= r_2(v^2-v)+\theta q_2(u-v) = g(u,v). \end{aligned} \tag{15}$$

For this pair of equations the square $[0,1]^2$ is positively invariant — since $f(0,v)>0$ and $f(1,v)<0$ for $0<v<1$ and similarly $g(u,0)>0$ and $g(u,1)<0$ for $0<u<1$. S and T are clearly equilibria. At all other boundary points the flow is into the unit square in forwards time. In fact, any square of the form $[K_1,K_2]^2$ for $-\infty<K_1\le 0$ and $0\le K_2\le 1$ is positively invariant.

For the Heaviside initial data the solution is clear — the Heaviside function simply propagates at speed $-b_1$. This is also clear from the probabilistic method (see section 5.5).

The pair of equations (15) is also relevant because, if the initial data in the PDE system (equation (1)) is constant, that is $u(0,x)=K_1, v(0,x)=K_2$ (independently of x), then the solution satisfies (15), with initial conditions $u(0)=K_1, v(0)=K_2$. The constant initial data case is particularly relevant in light of the comparison arguments in section 2.2. Solutions with initial data bounded below and above by constants K_1 and K_2 can be bounded below and above by the solutions for constant initial data K_1 and K_2, but these solutions clearly go to zero for $-\infty<K_1\le 0$ and $0\le K_2<1$ (since the square $[K_1,K_2]^2$ is positively invariant, contains only one equilibrium, which is S, and a periodic solution within the square is not possible due to the direction of the vector field).

3 Useful algebraic results

Throughout this section we shall assume, without loss of generality, that $b_1\ge b_2$.

Lemma 3.1 *For $\theta>0$, there exists a (finite) number $c(\theta)$ that satisfies $-b_1\le c(\theta)\le -b_2$ and such that*

(i) *for $c>c(\theta)$, $c\ne -b_2$, the matrix $K_{c,\theta}(T)$ has at least one stable monotone eigenvalue, and*

(ii) *for $c<c(\theta)$, $c\ne -b_1$, the matrix $K_{c,\theta}(T)$ has no stable monotone eigenvalues.*

Proof. In fact, we can say rather more than this about $K_{c,\theta}(T)$. Explicitly, if $(B+cI)$ is invertible,

$$K_{c,\theta}(T) = \begin{pmatrix} \frac{\theta q_1 - r_1}{b_1+c} & -\frac{\theta q_1}{b_1+c} \\ -\frac{\theta q_2}{b_2+c} & \frac{\theta q_2 - r_2}{b_2+c} \end{pmatrix}.$$

Thus, its two eigenvalues, λ^+ and λ^-, are given by:

$$\lambda^{\pm} = \frac{1}{2}\left\{ \frac{\theta q_1 - r_1}{b_1+c} + \frac{\theta q_2 - r_2}{b_2+c} \pm \sqrt{\left(\frac{\theta q_1 - r_1}{b_1+c} - \frac{\theta q_2 - r_2}{b_2+c}\right)^2 + \frac{4\theta^2 q_1 q_2}{(b_1+c)(b_2+c)}}\right\}. \tag{16}$$

These eigenvalues correspond to eigenvectors of the form $\begin{pmatrix} 1 \\ v^{\pm} \end{pmatrix}$, where

$$v^{\pm} = \frac{(\theta q_1 - r_1) - \lambda^{\pm}(b_1+c)}{\theta q_1}. \tag{17}$$

Bifurcation diagrams of the eigenvalues as functions of c are given in Figures 1 and 2.

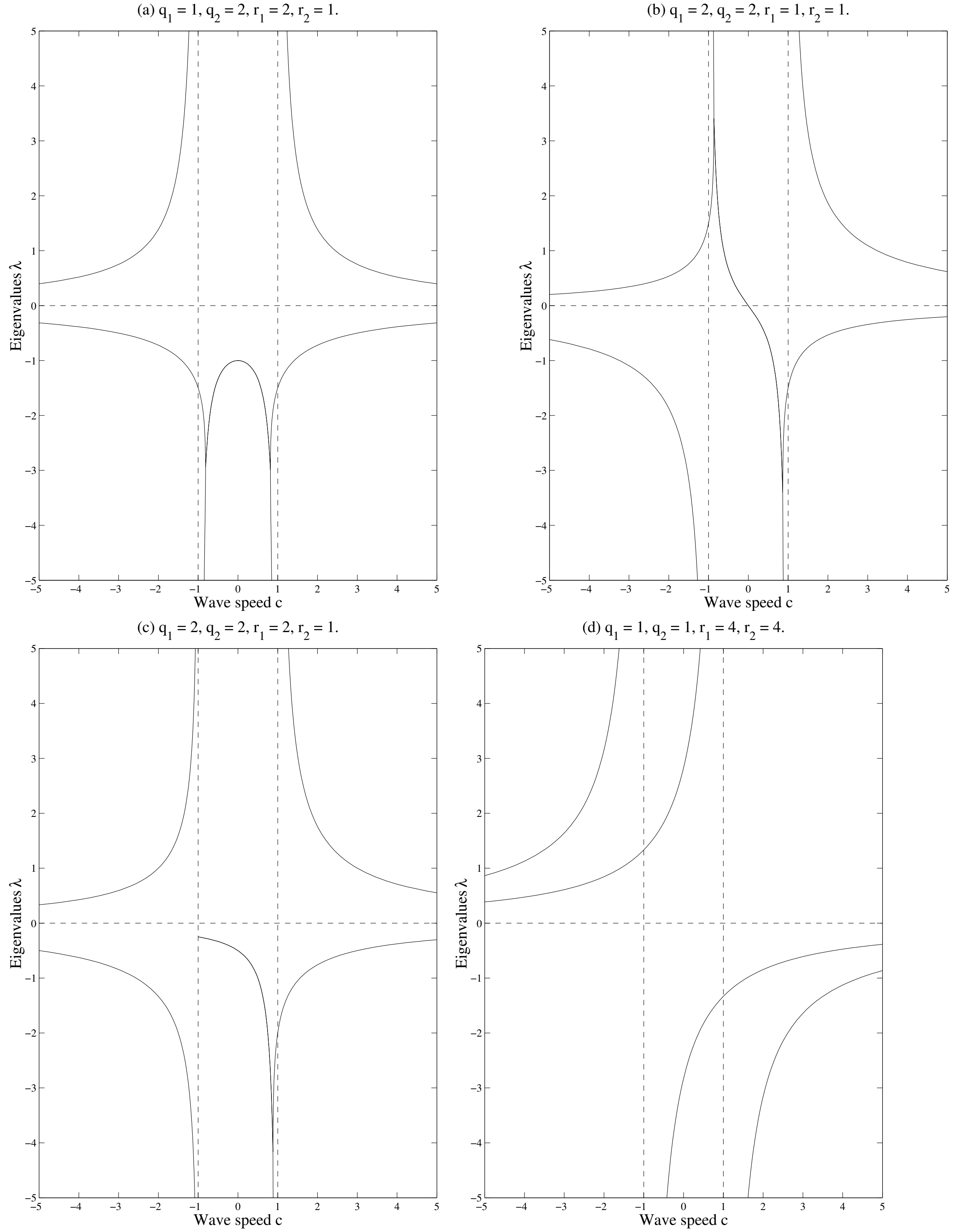

Figure 1: Bifurcation diagrams — plots of the eigenvalues λ^+ and λ^-, of $K_{c,\theta}(T)$, against c (where the eigenvalues are a complex conjugate pair the real part is plotted). All diagrams use $\theta = 1$, $b_1 = 1$ and $b_2 = -1$. The asymptotes $c = -1$ and $c = 1$ are marked. $c(\theta) = 1$ in (d), and is slightly smaller than that in each of (a)-(c) — it is the right-hand intersection of λ^+ and λ^-.

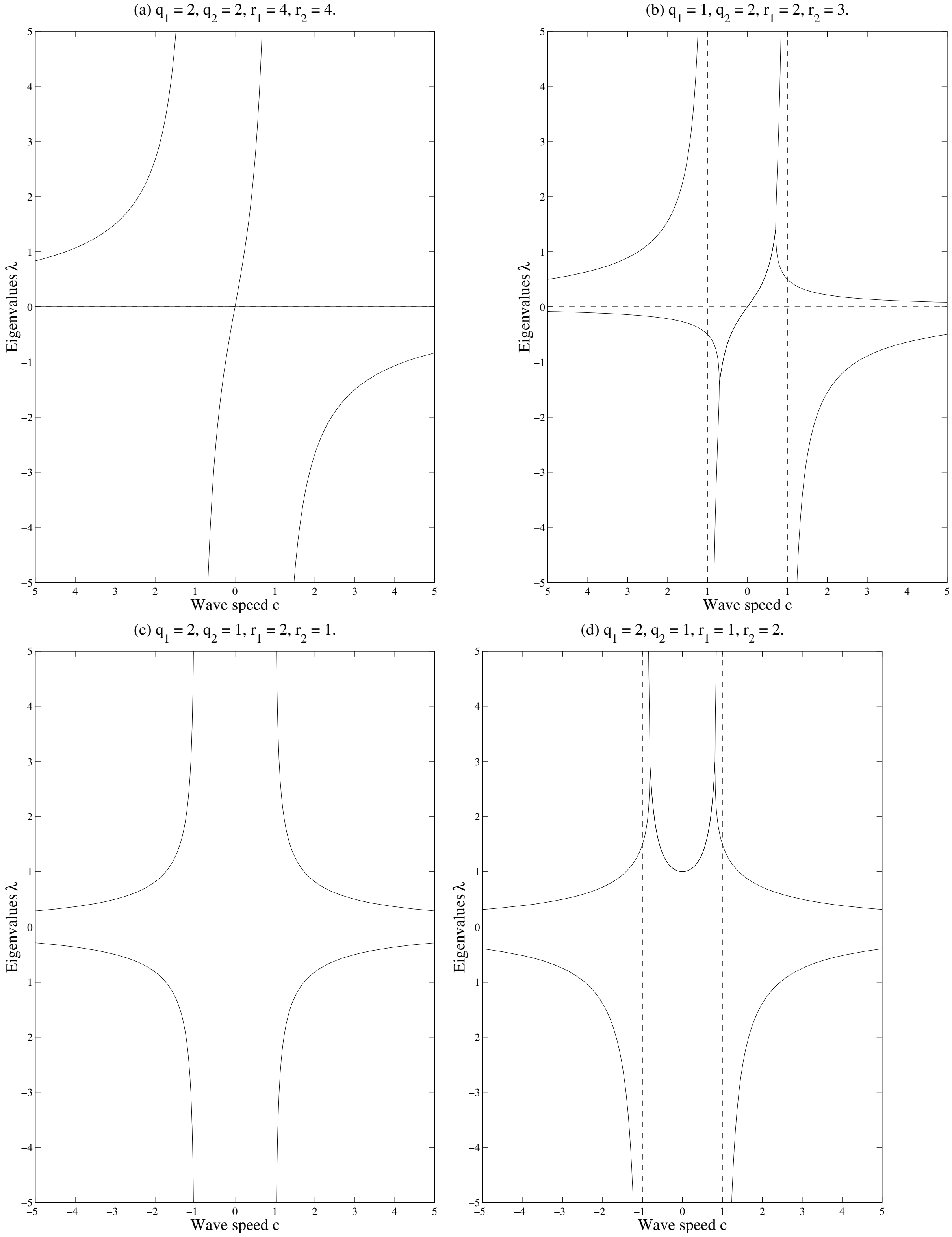

Figure 2: Bifurcation diagrams — plots of the eigenvalues λ^+ and λ^-, of $K_{c,\theta}(T)$, against c (where the eigenvalues are a complex conjugate pair the real part is plotted). All diagrams use $\theta = 1$, $b_1 = 1$ and $b_2 = -1$. The asymptotes $c = -1$ and $c = 1$ are marked. $c(\theta) = 1$ in each of these plots.

For $c > -b_2$, note that λ^- is a stable monotone eigenvalue and λ^+ is not, since $v^+ < 0$ (as $\lambda^+ > \frac{\theta q_1 - r_1}{b_1 + c}$ in this case). For $c < -b_1$, note that neither eigenvalue is stable monotone as $v^- < 0$ and $\lambda^+ > 0$. Thus, if there is a critical value $c(\theta)$ it lies in the interval claimed. For $b_1 = b_2 = b$, say, this is enough to prove the lemma, $c(\theta) = -b$, independent of θ.

So, to complete the proof of the lemma it is sufficient to consider the case $b_1 > b_2$, where it is necessary to examine the eigenvalues when $-b_1 < c < -b_2$, i.e. $(b_1 + c) > 0 > (b_2 + c)$.

Examining the term under the square root in equation (16) (which we will denote by $h(c)$, for fixed θ), for $-b_1 < c < -b_2$, note that it can have one, two or no zeroes. If there are two zeroes let us denote them by c_1 and c_2, with $-b_1 < c_1 < c_2 < -b_2$. When $\theta q_2 > r_2$ and $\theta q_1 \neq r_1$, there are two zeroes of h and for $c_1 < c < c_2$, $h(c) < 0$. So in this case the eigenvalues of $K_{c,\theta}(T)$ are complex. When $-b_1 < c < c_1$ and when $c_2 < c < -b_2$, $h(c) > 0$, so the eigenvalues are real. When $c_2 < c < -b_2$ both eigenvalues are stable monotone. When $-b_1 < c < c_1$ the eigenvalues are not stable monotone — $\theta q_1 < r_1$ implies that the eigenvectors have components of opposite signs (as shown in Figure 1 (a)) while $\theta q_1 > r_1$ implies that the eigenvalues are positive (as shown in Figure 1 (b)). Thus, in this case, $c(\theta) = c_2$.

When $\theta q_2 > r_2$ and $\theta q_1 = r_1$ there is a single zero of h, at c_1 say (as shown in Figure 1 (c)). When $c_1 < c < -b_2$, $h(c) > 0$ and both eigenvalues are stable monotone. When $-b_1 < c < c_1$, $h(c) < 0$ and so the eigenvalues are complex. Hence $c(\theta) = c_1$ in this case.

When $\theta q_2 \leq r_2$ matters are simpler because $\theta q_1 < r_1$ implies that the eigenvectors have components of opposite signs or the eigenvalues are complex, within the interval $-b_1 < c < -b_2$ (as shown in Figures 1 (d) and 2 (a) and (b)), and $\theta q_1 \geq r_1$ implies that the eigenvalues have non-negative real part throughout the interval $-b_1 < c < -b_2$ (as shown in Figures 2 (c) and (d)). Thus, for $\theta q_2 \leq r_2$, $c(\theta) = -b_2$.

Hence the lemma is proven. □

When $c = c(\theta)$ there is a repeated stable monotone eigenvalue, unless $c(\theta) = -b_i$, in which case the point is moot since the stability matrix does not exist, but the analysis can be completed by direct methods (see section 4.6).

For the probabilistic method we need to look at one of these special cases in more detail — the case where $-b_1 < c(\theta) < -b_2$ and $c = -b_2$. Note that $-b_1 < c(\theta) < -b_2$ if and only if $r_2 < \theta q_2$.

Recall that $K_{c,\theta}(T)$ is defined by equation (4), thus an eigenvector v (with corresponding eigenvalue λ) of $K_{c,\theta}(T)$ will satisfy the equation

$$\lambda(B + cI)v + (R + \theta Q)v = 0. \tag{18}$$

For $(B + cI)$ invertible, this relation is also true in the opposite direction — a non-trivial vector v that satisfies equation (18) is an eigenvector of $K_{c,\theta}(T)$ with eigenvalue λ. However, for $(B + cI)$ singular, equation (18) can still have non-trivial solutions, in the case of interest it has one solution,

$$\lambda = -\frac{(r_1 - \theta q_1)(r_2 - \theta q_2) - \theta^2 q_1 q_2}{(b_1 - b_2)(r_2 - \theta q_2)} < 0,$$

with $v = \begin{pmatrix} 1 \\ v_2 \end{pmatrix}$, where

$$v_2 = \frac{\theta q_2}{\theta q_2 - r_2} > 0.$$

Thus this solution (which is what we are really interested in, analysis of $K_{c,\theta}(T)$ is a short-cut, and simplifies discussion by giving us a way of referring to eigenvalues and eigenvectors as those of $K_{c,\theta}(T)$) can be thought of as a generalized stable monotone eigenvalue of $K_{c,\theta}(T)$ and we have proved that, for $c > c(\theta)$ the matrix $K_{c,\theta}(T)$ has at least one (possibly generalized) stable monotone eigenvalue.

The value of $c(\theta)$ can also be derived from the probabilistic model of the system: the theory of large deviations gives a formula for $c(\theta)$ which effectively summarises all these cases; see section 5.2. That formula, which is shown in section 5.2 to be entirely equivalent to the preceding characterization, makes it easy to observe that $c(\theta)$ is decreasing as θ increases and that

$$c^* := \lim_{\theta\to\infty} c(\theta) = -\left(\frac{b_1 q_2 + b_2 q_1}{q_1 + q_2}\right).$$

This limit will be discussed in section 5.2, for now note that the lower bound on $c(\theta)$ given by Lemma 3.1 can therefore be tightened to c^*.

It is also possible to arrive at this limit by further manipulating $h(c)$. Rearranging the equation $h(c) = 0$ to express it as a quadratic in c (where the coefficients are quadratic in the other parameters, including θ) enables us to consider the limit as $\theta \to \infty$ by picking out the terms of highest order (i.e. order two in θ) only. Simplifying this expression yields

$$\theta^2\left\{c(\theta)(q_1 + q_2) + (b_1 q_2 + b_2 q_1)\right\}^2 + \text{terms of lower order in } \theta = 0.$$

Thus the θ^2 term is zero if, and only if, $c = c^*$.

Further explicit calculation on $K_{c,\theta}(S)$ and $K_{c,\theta}(T)$ allows us to summarize the locations (i.e. left- or right-half plane) of the eigenvalues as follows:

- For $c > -b_2$: both eigenvalues of $K_{c,\theta}(S)$ are real and positive, the smaller is unstable monotone, the larger not; $K_{c,\theta}(T)$'s eigenvalues are both real, for $\theta < \frac{1}{\rho_1+\rho_2}$ both are negative, for $\theta = \frac{1}{\rho_1+\rho_2}$ one is zero and the other is negative and for $\theta > \frac{1}{\rho_1+\rho_2}$ one is positive and one is negative.
- For $c < -b_1$: both eigenvalues of $K_{c,\theta}(S)$ are real and negative, the eigenvalue closer to zero is stable monotone, the other not; $K_{c,\theta}(T)$'s eigenvalues are real and positive for $\theta < \frac{1}{\rho_1+\rho_2}$, for $\theta = \frac{1}{\rho_1+\rho_2}$ one is positive and the other is zero and for $\theta > \frac{1}{\rho_1+\rho_2}$ one is positive and one is negative.
- For $-b_1 < c < -b_2$: the eigenvalues of $K_{c,\theta}(S)$ are real and have opposite signs, one is stable monotone, the other unstable monotone; for $\theta < \frac{1}{\rho_1+\rho_2}$, $K_{c,\theta}(T)$'s eigenvalues are real and have opposite signs, details for larger θ depend on the relative sizes of c and $c(\theta)$.

The following result is required in our proof of $\mathcal{L}^1$ convergence of Z_λ in Theorem 5.4.

Lemma 3.2 *(i) Suppose that $c > c(\theta)$. Let $\lambda_s(c)$ be the stable monotone eigenvalue of $K_{c,\theta}(T)$ (the one nearer to 0 if there are two). From the definition of $K_{c,\theta}$ and $\lambda_s(c)$ it is easily seen that $-\lambda_s(c)c$ is the Perron-Frobenius eigenvalue of $(\lambda_s(c)B + \theta Q + R)$. For $\mu < \lambda_s(c)$, with μ sufficiently close to $\lambda_s(c)$,*

$$\Lambda_{PF}(\mu B + \theta Q + R) = -\mu c_1(\mu) \text{ for some } c_1(\mu) < c.$$

(ii) As $c \downarrow c(\theta)$, we have

$$\lambda_s(c)^{-1}\Lambda_{PF}(\lambda_s(c)B + \theta Q + R) \to -c(\theta) \text{ and } \left[\frac{\partial}{\partial\mu}\Lambda_{PF}(\mu B + \theta Q + R)\right]_{\mu=\lambda_s(c)} \to -c(\theta).$$

(iii) When $c(\theta) < -b_2$, $K_{c(\theta),\theta}(T)$ has a double eigenvalue, which we will denote by λ_0. This eigenvalue is geometrically simple, i.e. it has only one normalised eigenvector even though it has algebraic multiplicity two, and is stable monotone.

Proof. We follow the proof of Lemma 4.4 of Crooks [6].

(i) We can explicitly write $\lambda_s(\cdot)$ as a function of c — for $c > -b_2$ it is λ^- from equation (16), for $c < -b_2$ it is λ^+ from equation (16). Considered as a function of c it is easy to see that it is continuously differentiable away from $c = -b_i$ for $i = 1, 2$ and to check that $\lambda_s(c) \downarrow \lambda_0$ as $c \downarrow c(\theta)$. So for $c > c(\theta)$

and $\lambda_0 < \mu < \lambda_s(c)$, there exists some $c_1(\mu)$ such that $c(\theta) < c_1(\mu) < c$ and $\lambda_s\big(c_1(\mu)\big) = \mu$ and thus $\Lambda_{PF}(\mu B + \theta Q + R) = -\mu c_1(\mu)$.

(ii) The first part follows from the fact that $\lambda_s(c)^{-1}\Lambda_{PF}(\lambda_s(c)B + \theta Q + R) = c$. For the second part consider again the explicit form of the Perron-Frobenius eigenvalue as a function of μ:

$$\begin{aligned}\Lambda_{PF}(\mu B + \theta Q + R) &= \frac{1}{2}\big(\mu(b_1+b_2) - (\theta q_1 - r_1) - (\theta q_2 - r_2)\big) \\ &+ \frac{1}{2}\sqrt{\Big(\mu(b_1-b_2) - (\theta q_1 - r_1) + (\theta q_2 - r_2)\Big)^2 + 4\theta^2 q_1 q_2}.\end{aligned}$$

It is clear, by continuity (considering Λ_{PF} as a function of μ and noting that the term under the square root is strictly positive), that

$$\left[\frac{\partial}{\partial\mu}\Lambda_{PF}(\mu B + \theta Q + R)\right]_{\mu=\lambda_s(c)} \to \left[\frac{\partial}{\partial\mu}\Lambda_{PF}(\mu B + \theta Q + R)\right]_{\mu=\lambda_0}$$

as $c \downarrow c(\theta)$ because $\lambda_s(c) \to \lambda_0$ as $c \downarrow c(\theta)$. Now

$$\frac{\partial}{\partial\mu}\Lambda_{PF}(\mu B + \theta Q + R) = \frac{\partial}{\partial\mu}\Lambda_{PF}(\mu B + \mu c(\theta)I + \theta Q + R) - c(\theta). \tag{19}$$

We claim that $\Lambda_{PF}(\mu B + \mu c(\theta)I + \theta Q + R)$ must attain a local minimum at $\mu = \lambda_0$. Note that, for fixed $c > c(\theta)$, sufficiently close to $c(\theta)$, there are two values of μ such that $\Lambda_{PF}(\mu B + \mu c I + \theta Q + R) = 0$, these two values of μ are the two stable monotone eigenvalues of $K_{c,\theta}(T)$. We now use a convexity result due to Cohen [4].

To be precise, Cohen's result states that the Perron-Frobenius eigenvalue of a matrix $M_1 + M_2$ is a convex function of M_2, where M_1 has positive elements off the main diagonal and M_2 is a diagonal matrix (possibly 0). Thus, for μ between the two eigenvalues $\Lambda_{PF}(\mu B + \mu c I + \theta Q + R) < 0$ (strict inequality since there can be at most two zeroes — each is an eigenvalue of $K_{c,\theta}(T)$, a 2×2 matrix). Thus, for $c = c(\theta)$, $\Lambda_{PF}(\mu B + \mu c(\theta)I + \theta Q + R) > 0$ except at $\mu = \lambda_0$ where it is zero.

Thus the derivative on the right hand side of equation (19) is zero at $\mu = \lambda_0$ and hence the proof of (ii) is complete.

(iii) As c decreases through $c(\theta)$ the two stable monotone eigenvalues of $K_{c,\theta}$ coalesce and, at least for c sufficiently close to $c(\theta)$, become a complex conjugate pair with negative real part. From equation (16), $\lambda_0 = \frac{1}{2}\left\{\frac{\theta q_1 - r_1}{b_1 + c(\theta)} + \frac{\theta q_2 - r_2}{b_2 + c(\theta)}\right\}$, and from equation (17), this corresponds to an eigenvector of the form $\begin{pmatrix} 1 \\ v_0 \end{pmatrix}$, where

$$v_0 = \frac{(\theta q_1 - r_1) - \lambda_0\big(b_1 + c(\theta)\big)}{\theta q_1}.$$

We can verify directly that λ_0 and v_0 as given above have the correct signs (negative and positive respectively) by using inequalities arising from the definition of $c(\theta)$. □

For the probabilistic proof of uniqueness, modulo translation, of monotone travelling waves from S to T, the next lemma is important.

Lemma 3.3 *Suppose that $c > c(\theta)$ and that $K_{c,\theta}(T)$ has two stable monotone eigenvalues. Let β be the stable monotone eigenvalue further from 0. Then, for $\alpha > \beta$ with α sufficiently close to β, the only non-negative 2-vector g such that*

$$0 \le \big(\alpha(B + cI) + \theta Q + R\big)g$$

is the zero vector: $g = 0$.

Proof. Convexity of $\Lambda_{PF}(\mu(B+cI)+\theta Q+R)$ as a function of μ given by Cohen's Theorem [4], for $\beta<\mu<\lambda$, yields $\Lambda_{PF}(\mu(B+cI)+\theta Q+R)\leq 0$. However, there can only be two values (μ_1 and μ_2, say) of μ at which this function is 0 since each μ_i is thus an eigenvalue of a 2×2 matrix. These two values are therefore β and λ and the inequality is strict.

Suppose that for some μ satisfying $\beta<\mu<\lambda$, there exists g, non-negative, with $g\neq 0$, and $(\mu(B+cI)+\theta Q+R)g>0$. Then $\Lambda_{PF}(\mu(B+cI)+\theta Q+R)\geq 0$ which is a contradiction. □

Finally, we need the following result in showing that

$$\mathbb{P}_{x,y}\big[Z_\lambda(\infty)=0\big]=0 \text{ or } 1.$$

Lemma 3.4 *If w is a 2-vector such that $0\leq w\leq 1$ and*

$$R(w^2)=(R-\theta Q)w,$$

then either $w=(1,1)$ or $w=(0,0)$.

Proof. Rearranging the above equation notice that the problem amounts to searching for intersections of two parabolae inside the unit square — $\theta q_1 w_2=(r_1+\theta q_1)w_1-r_1w_1^2$ and $\theta q_2 w_1=(r_2+\theta q_2)w_2-r_2w_2^2$. Since each parabola goes through $w=(0,0)$ and $w=(1,1)$ there can be no other intersections in the square — since from $(0,0)$ to $(1,1)$ the curve $\theta q_1 w_2=(r_1+\theta q_1)w_1-r_1w_1^2$ is above the line $w_2=w_1$ while the other curve is below it. □

4 Analytic proofs of existence and uniqueness of travelling waves

4.1 Plan of attack

To prove Theorem 1.1 we use shooting arguments. First note that the path of any solution w of (2) which satisfies

$$w(x)\to T \quad \text{as} \quad x\to\infty,\quad w(x)\to S \quad \text{as} \quad x\to-\infty \quad \text{and} \quad w'(x)>0, x\in\mathbb{R},$$

must lie entirely inside the open unit square. Whether it must also lie between the nullclines $w_1'=0$ and $w_2'=0$ depends on the values of various parameters, as explained below. Depending on the geometric configuration of the nullclines, we introduce shooting boxes. We observe that these regions have four key properties upon which our proof will be based:

- Non-constant solution curves do not intersect the boundary of the region tangentially and hence any non-constant solution curve which intersects the boundary crosses it transversally;
- Exit-times of solutions in the regions are continuous functions of initial conditions;
- No non-constant solution curve passes through S or T;
- Other than S and T there are no equilibria in the closure of these regions.

There are two types of region that we use — *inside regions* when $c>\max(-b_1,-b_2)$ and *outside regions* when $\min(-b_1,-b_2)<c<\max(-b_1,-b_2)$. The terminology is based on the geometry of the phase plane as will be seen shortly.

4.2 Labelling the phase plane

In the 2-dimensional $w = (w_1, w_2)$-plane consider the parabolae

$$P_1 : \theta q_1 w_2 - (r_1 + \theta q_1) w_1 + r_1 w_1^2 = 0$$

and

$$P_2 : \theta q_2 w_1 - (r_2 + \theta q_2) w_2 + r_2 w_2^2 = 0$$

and let Ω_i denote the open region in the first quadrant between P_i and the w_i-axis, $i = 1, 2$. The parabolae are of course the nullclines for the system, and their intersections are its equilibria. The point $S = (0, 0) \in P_1 \cap P_2$ for all $\theta > 0$, the point $T = (1, 1) \in P_1 \cap P_2$ for all $\theta > 0$ and $E_\pm \in P_1 \cap P_2$ if $0 < \theta \leq (4\rho_1\rho_2)^{-\frac{1}{2}}$.

Note that the relative positions of E_+, E_- and T depend on the value of θ and that no two of them are commensurate with respect to the partial ordering on $\mathbb{R}^2$ induced by the positive quadrant. Let us denote by E_1 the element of $\{E_+, E_-, T\}$ with the largest w_1-component and by E_3 that with the largest w_2-component. Then, for $\theta(\rho_1 + \rho_2) < 1$, $T = E_2$. When $\theta(\rho_1 + \rho_2) > 1$ and $\theta \leq (4\rho_1\rho_2)^{-\frac{1}{2}}$, T could be E_1 or E_3. In this case $T = E_1$ for $\rho_2 > \rho_1$, $T = E_3$ for $\rho_1 > \rho_2$. The various equality cases missed out of this enumeration are those where equilibria coincide.

Also note that the relative position of T and the maximum of the parabola P_i is determined by the sign of $\theta q_i - r_i$ — if $\theta q_i \geq r_i$ then the segment of P_i from S to T is monotone; if $\theta q_i < r_i$ then this segment has a turning point before reaching T.

Regions of interest. Now let $\tilde{\Sigma}$ denote the rectangle in the w-plane with two sides on the axes intersecting at 0, a side through E_1 parallel to $\{w_1 = 0\}$ and one through E_3 parallel to $\{w_2 = 0\}$. Let Σ denote the open convex subset of $\tilde{\Sigma}$ whose boundary comprises four straight-line segments from $\partial\tilde{\Sigma}$, a parabolic segment from P_1 joining E_1 to E_2 and a parabolic segment from P_2 joining E_2 to E_3. Thus the boundary of Σ always has four straight-line segments: in addition it has two parabolic components when E_+, E_- and T are distinct, one parabolic component when two of E_+, E_- and T coincide and no parabolic component when $\theta > (4\rho_1\rho_2)^{-\frac{1}{2}}$. Also Σ consists of the union of three sets: $\omega_2 = \tilde{\Sigma} \cap \Omega_1 \cap \Omega_2$, a relatively closed component ω_3 whose boundary intersects $\{w_1 = 0\}$ away from the origin and a relatively closed component ω_1 whose boundary intersects $\{w_2 = 0\}$ away from the origin. (See Figure 3.) Note that

$$\begin{aligned}
(r_1 + \theta q_1) w_1 - r_1 w_1^2 - \theta q_1 w_2 &> 0, \quad (w_1, w_2) \in \omega_2 \cup \omega_1, \\
(r_2 + \theta q_2) w_2 - r_2 w_2^2 - \theta q_2 w_1 &> 0, \quad (w_1, w_2) \in \omega_2 \cup \omega_3;
\end{aligned}$$

and hence, if $w = (w_1, w_2)$ satisfies (2) then,

$$\begin{aligned}
&(b_1 + c) w_1' \text{ is strictly positive at } x \text{ if } w(x) \in \omega_2 \cup \omega_1, \\
&(b_2 + c) w_2' \text{ is strictly positive at } x \text{ if } w(x) \in \omega_2 \cup \omega_3, \\
&(b_1 + c) w_1' \text{ is strictly negative at } x \text{ if } w(x) \in \omega_3 \setminus \overline{\omega}_2, \\
&(b_2 + c) w_2' \text{ is strictly negative at } x \text{ if } w(x) \in \omega_1 \setminus \overline{\omega}_2.
\end{aligned} \tag{20}$$

A monotone curve connecting S to E_1 must approach E_1 from $\omega_1 \cup \omega_2$. Thus, by (20), if $(b_1 + c) < 0$ there cannot be a monotone connection from S to E_1. Similarly, a monotone curve connecting S to E_3 must approach E_3 from $\omega_2 \cup \omega_3$. Thus, by (20), if $(b_2 + c) < 0$ there cannot be a monotone connection from S to E_3.

A monotone connection to E_2 must eventually approach through ω_2, so a necessary condition for existence is $(b_1 + c) > 0$ *and* $(b_2 + c) > 0$.

When considering monotone connections from S to T it is sufficient to restrict attention to the intersections of each ω_i with the unit square — a monotone connection to T must lie entirely within the unit

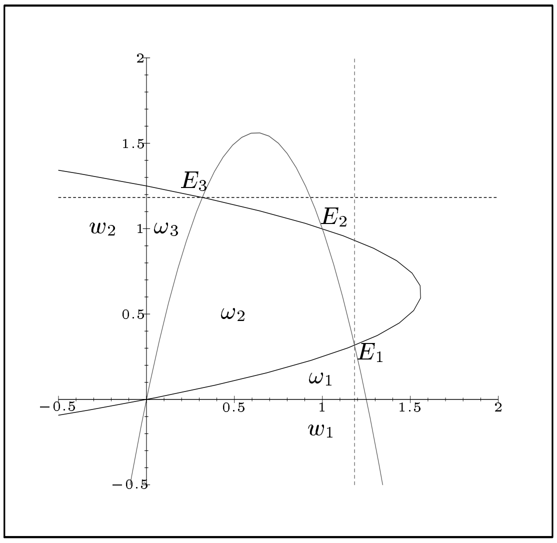

Figure 3: The regions of interest

square and must eventually approach T through some ω_i. It is possible to work with just ω_2 and ω_3 — since ω_1 is mapped to ω_3 by swapping b_1 and b_2, r_1 and r_2 and q_1 and q_2. (It is not possible for a monotone connection from S to T to lie partly within ω_1 and partly within ω_3 since both w_1' and w_2' have signs opposite in ω_1 to their signs in ω_3.) The *outside* region is ω_3 intersected with the unit square; the *inside* region is a subset of ω_2 inside the unit square. Examination of these two regions allows us to cover all possible monotone connections to T — connections must eventually approach from one of the two.

Also note that the nullclines represent points at which solution curves are vertical or horizontal (where the nullclines cross, there are fixed points). Thus, if a solution curve hits a nullcline somewhere other than an equilibria it immediately crosses to the other side of the nullcline, tangency is not possible since the nullclines are nowhere vertical (for the one representing vertical solution curves) and nowhere horizontal (for the other).

4.3 The inside region

Consider the region formed by the segment of the parabola P_1 from S to the first intersection with the line $w_2 = 1$ (this is at T if $\frac{\theta q_1}{r_1} > 1$); if the intersection is not at T then use the line $w_2 = 1$ to connect the endpoint of the segment to T; the segment of P_2 from S to the first intersection with the line $w_1 = 1$ (this is at T if $\frac{\theta q_2}{r_2} > 1$) and if the intersection is not at T then use the line $w_1 = 1$ to connect the endpoint to T. Thus, the *inside region* is defined to be the region *inside* both parabolae and the unit square, in the notation of the previous section it is a subset of ω_2. See Figures 4 and 5.

Denote the edges of the region as follows:

- The open segment of P_1 (not including S or the other endpoint) by A.
- The open segment of P_2 (not including S or the other endpoint) by B.
- If A does not connect S to T then denote the open segment of $w_2 = 1$ (not including endpoints) by C.

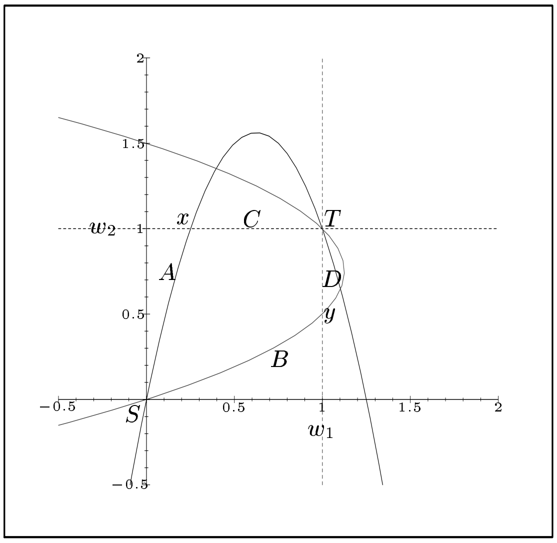

Figure 4: An example of an inside region with 4 distinct edges

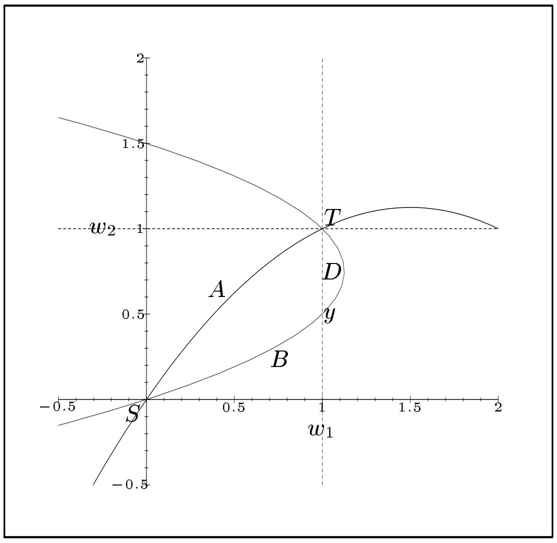

Figure 5: An example of an inside region with only 3 distinct edges

- If B does not connect S to T then denote the open segment of $w_1 = 1$ (not including endpoints) by D.

Notice that a monotone connection from S to T must approach T from inside this region if both C and D exist. If either C or D does not exist then it is possible to reach T monotonically from outside the *inside region* — then we must use the *outside* region as discussed below.

If C exists, then denote its left endpoint by x and if D exists then denote its lower endpoint by y.

By (20) and the following discussion, a necessary condition for a monotone connection that eventually approaches T from the inside region is that $c > \max(-b_1, -b_2)$, i.e.

$$b_1 + c > 0 \qquad \textit{and} \qquad b_2 + c > 0, \tag{C1}$$

since $w_1' > 0$ and $w_2' > 0$ (throughout the interior of the region) if and only if (C1) holds.

Lemma 4.1 *Assume (C1) holds. Then a solution curve that hits the boundary of the inside region at any point other than S or T immediately crosses the boundary.*

Proof. Consider the various segments of the boundary. On A and at the point x, if it exists, $w_1' = 0$ and $w_2' > 0$. Thus (using continuity and the fact that the slope of this boundary is bounded and positive, so the flow is not tangent to the boundary) if $w(t) \in A$ or $w(t) = x$ then there exists an $\epsilon > 0$ such that for $s \in (t - \epsilon, t), w(s)$ is inside the region and for $s \in (t, t + \epsilon), w(s)$ is outside the region.

Similarly for B and y — $w_2' = 0$ and $w_1' > 0$ so the same argument works.

For C, if it exists, note that $w_1' > 0$ and $w_2' > 0$ on C and again the solution curve crosses from inside to outside automatically. Similarly for D, and the lemma is proved. ☐

4.4 The outside region

We have observed that there are monotone curves from S to T that do not intersect the inside region, except at T itself, when the inside region does not have four edges. Without loss of generality assume that the edge C of the inside region does not exist, i.e. $\frac{\theta q_1}{r_1} > 1$, so that there is a possibility of a monotone connection through ω_3. When neither C nor D exists there are two outside regions, by equation (20) the flow is in opposite directions in the two regions. Hence there can be a monotone connection through at most one of them.

Consider the region formed by the segment of the parabola P_1 from S to T; the w_2 axis and the line $w_2 = m(w_1 - 1) + 1$ for some $m \in (0, 1 - \frac{r_1}{\theta q_1})$. This restriction on m ensures that the line has positive slope and lies above the parabola P_1 (the upper bound just given is the slope of the parabola at T and the assumption that $\frac{\theta q_1}{r_1} > 1$ ensures that the slope is strictly positive). We will choose a particular value of m later. In the notation of section 4.2 this region is therefore a subset of ω_3. See Figure 6 for a diagram of a typical example of this region.

Denote the edges of the region as follows:

- The open segment of P_1 (not including S or T) by A.
- The half-open segment of the w_2 axis (not including S but including the other endpoint) by B.
- The open segment of $w_2 = m(w_1 - 1) + 1$ strictly between the w_2 axis and T by C.

For a monotone connection from S to T that does not go through the inside region it is necessary that $w_1' > 0$ and $w_2' > 0$ throughout the interior of the outside region (note that though there are curves from

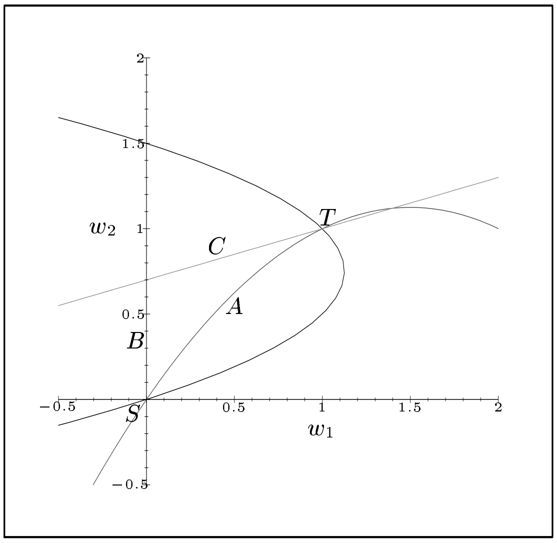

Figure 6: An example of an outside region

S to T that exit the outside region through C and still converge monotonically to T this condition will still be necessary).

Using equation (20) to consider the direction of the flow within the region notice that is is therefore necessary that $b_2 > b_1$ and that $-b_2 < c < -b_1$, i.e. $(b_2 + c) > 0 > (b_1 + c)$. So, for there to be a possibility of a monotone connection that is not through the inside region, the following conditions are necessary:

$$(b_2 + c) > 0 > (b_1 + c) \qquad \textit{and} \qquad \frac{\theta q_1}{r_1} > 1. \tag{C2}$$

For a monotone connection through the other outside region it is necessary that $(b_1 + c) > 0 > (b_2 + c)$ *and* $\frac{\theta q_2}{r_2} > 1$ — the subsequent analysis is entirely equivalent since we can simply interchange the subscripts 1 and 2. The two cases are clearly mutually exclusive.

The condition (C2) ensures that, for each $k > 0$, a curve that satisfies $w_2'/w_1' = k$ is an ellipse passing through S and T. This observation is used below to rule out internal tangencies to the region.

A solution curve cannot exit from the outside region through A or B in forwards time. On A observe that $w_1' = 0$ and $w_2' > 0$ and on B that $w_1' > 0$ and $w_2' > 0$.

We need to check that the flow does not have an internal tangency to C in order to construct a shooting argument for the outside region. The shooting argument used runs backwards in time, but note that a tangency backwards in time is also a tangency forwards in time, and vice versa.

Lemma 4.2 *Assume condition (C2) holds. Then a solution curve that hits the boundary of the outside region (from the inside of the region — an external tangency is immaterial) at any point other than S or T immediately crosses the boundary.*

Proof. If a solution curve hits A or B the observations above show that the curve (in backwards time) will cross from the inside to the outside of the region. So assume, for a contradiction, that at a point x_1 on C the solution curve is internally tangent to C. The solution curve in forwards time will re-enter the interior of the region — where $w_1' > 0$ and $w_2' > 0$ — and must eventually hit the line C again or else approach T. This is since it cannot leave the region through A because $w_1' = 0, w_2' > 0$ there. Call the first point at which this occurs x_2.

Thus, the solution curve is a smooth curve from x_1 to x_2, both of which are on the line $w_2 = m(w_1 - 1) + 1$, therefore there is a point x_3 on this curve at which the curve has slope m. However, the curve on which the slope of the flow is m is an ellipse; an ellipse that passes through S, T and x_1 — so x_3 cannot be on this ellipse. This contradiction proves the result. □

4.5 The shooting argument

To establish the main result (Theorem 1.1) we shall use the preceding observations in a shooting argument through the outside region (backwards in time), as well as a more standard forward time argument from S to T for the inside region. Note that a monotone travelling wave from S to T is possible only if T has stable monotone eigenvalues. A corresponding unstable direction at S is also necessary, but this always exists when a stable monotone eigenvalue at T exists. (See the discussion before Lemma 3.2 in section 3.) It is known so far that if $c > \max(-b_1, -b_2)$ then a monotone connection can only exist through the inside region — we show below that in this case it does indeed exist and is unique. For $c < \min(-b_1, -b_2)$ there cannot be a monotone connection from S to T and for c in between then there may or may not be a monotone connection (which necessarily lies entirely in an outside region) — the determining factor is whether c is greater than $c(\theta)$. When such a connection exists it is unique.

4.5.1 Through the inside region when (C1) holds

Consider the intersection of the inside region with the circle of radius ϵ, centred at the origin, S. This is a segment of a circle with 2 endpoints on the boundary of the inside region. $w_1' > 0$ *and* $w_2' > 0$ are necessary for a monotone connection through the region (i.e. $(b_1 + c) > 0$ *and* $(b_2 + c) > 0$ are necessary in this region) — when this is true a connection looks plausible. Considering the family of solution curves which pass through points of this segment at time zero completes our shooting argument, since, running backwards in time all these curves must go to S (since they cannot exit the region and cannot tend to any point other than an equilibrium), and running forwards in time, the curves from the 2 endpoints leave the region immediately and curves from points in between will exit transversally from the region (by Lemma 4.1), thus exiting in between the exit/end-points. Hence classical continuous dependence theory for initial value problems allows us the conclude that a monotone connection from S to T exists.

4.5.2 Through the outside region when (C2) holds

Assume condition (C2), otherwise there is no monotone connection through the region. This time we shoot backwards in time.

Firstly we deal with $c > c(\theta)$. Choose m so that the edge C has slope equal to half that of the dominant eigenvector of $K_{c,\theta}(T)$ (the dominant eigenvector is that corresponding to the negative real eigenvalue of smallest modulus — when (C2) and $c > c(\theta)$ there are two simple stable monotone eigenvalues so m is well-defined). Thus the upper edge of the region bisects the angle between the line $w_2 = 1$ and the dominant eigenvector at T. This enables us, backwards in time, to obtain paths leaving the outside region on both sides of T, and hence shoot to S. More precisely, for $c > c(\theta)$ we already know that $K_{c,\theta}(T)$ has two stable monotone eigenvalues. Consideration of equation (17) shows that both have slope less than that of the P_1 at T. The dominant eigenvector is $\begin{pmatrix} 1 \\ v^+ \end{pmatrix}$. Thus we choose m to correspond to a direction $\begin{pmatrix} 1 \\ v^+/2 \end{pmatrix}$. Hence, by shooting backwards from points suitably close to T and using the fact that the flow is determined by the dominant eigenvector we observe the flow sweeps through the boundaries of the region (by Lemma 4.2) as we follow a segment of a circle around T that intersects the region. Classical continuous dependence theory for initial value problems tells us that a connection from S to T exists.

For $c < c(\theta)$ there is no stable monotone eigenvalue at T and there cannot be a monotone connection. Thus we have completed showing that a monotone eigenvalue at T is sufficient (except for the cases $c = c(\theta)$ and $c = \max(-b_1, -b_2) > c(\theta)$ which are discussed below in section 4.6) for a monotone connection — we already knew it was necessary.

4.6 Special cases

We have been wary of the cases where $c = c(\theta)$ and $c = -b_i$. Assume, without loss of generality, that $b_1 > b_2$.

If $\theta q_2 \leq r_2$, then $c(\theta) = -b_2$. For $c = c(\theta)$ the only candidate for a connection from S to T is the segment of the w_2-nullcline, P_2, connecting S to T. If $\theta q_2 < r_2$ then this is not monotone from S to T and so there is no monotone travelling wave at this speed. When $\theta q_2 = r_2$ this connection is monotone. The travelling wave equations consists of one algebraic — defining the nullcline — and one differential equation. We can substitute the algebraic into the differential equation to obtain a one-dimensional problem where we are looking for a connection from 0 to 1. Since the derivative is positive between these points this indeed exists and the segment of nullcline is a monotone travelling wave.

When $\theta q_2 > r_2$, $-b_1 < c(\theta) < -b_2$. For the case $c = c(\theta)$ we can use the argument we used in the outside region — by part (iii) of Lemma 3.2 there is a double stable monotone eigenvalue so the corresponding eigenvector determines the nature of the flow near T. We simply repeat the bisection procedure and show that the flow leaves either side of the region — so there is a monotone connection for $c = c(\theta)$ in this case. We also must check the case $c = -b_2$, but here the solution is simply the segment of nullcline exactly as for $\theta q_2 = r_2$.

Thus we can summarize as follows:

- For $\theta q_2 < r_2$, there is a monotone travelling wave if and only if $c > c(\theta)$.
- For $\theta q_2 \geq r_2$. There is a monotone travelling wave if and only if $c \geq c(\theta)$.

4.7 Uniqueness modulo translation

To establish uniqueness of monotone travelling waves from S to E_i, $i = 1, 2, 3$, it suffices, because of the change of variables at (1.26), to prove that the wave from S to T is unique. However, due to the dimensionality of the unstable and stable manifolds at S and T respectively, we have very little left to show. As we summarised in section 3; just above Lemma 3.2; we have that:

- For $c > \max(-b_1, -b_2)$; $K_{c,\theta}(T)$ has a one-dimensional stable manifold for $\theta \geq \frac{1}{\rho_1 + \rho_2}$.
- For $c < \min(-b_1, -b_2)$; $K_{c,\theta}(S)$ has a two-dimensional *stable* manifold. (Of course, here $c < c(\theta)$ anyway.)
- For $\min(-b_1, -b_2) < c < \max(-b_1, -b_2)$; $K_{c,\theta}(S)$ has a one-dimensional unstable manifold.

Thus, since we can only possibly have non-uniqueness when:

- $K_{c,\theta}(S)$ has a two-dimensional unstable manifold *and* $K_{c,\theta}(T)$ has a two-dimensional stable manifold, we have only one case left to check, i.e. $c > \max(-b_1, -b_2)$ and $\theta < \frac{1}{\rho_1+\rho_2}$.

In this case the two eigenvalues of T are negative and distinct, denote them by $\alpha < \beta < 0$, and the corresponding eigenvectors by v_α and v_β, respectively. α is a stable monotone eigenvalue (i.e. v_α has both components of the same sign) and β is not. Note that the dominant eigenvalue β is the non-monotone eigenvalue. This suggests that most of the flow entering T will do so non-monotonically and so there may indeed be a unique monotone connection. We now use this idea to complete the proof of uniqueness.

Firstly, by standard theory (Coddington and Levinson [3, Chapter 13, Theorem 4.4]), all solutions that converge to T do so exponentially. For notational convenience define u by $w_i(x) = u_i(x)+1$ for $i = \{1, 2\}$, so that u converges to the origin when w converges to T. Then there exists a one-dimensional manifold of solutions u such that

$$\frac{\log \parallel u(x) \parallel}{x} \to \alpha \text{ as } x \to \infty.$$

and all other solutions that converge to the origin do so with rate β.

By Coddington and Levinson [3, Chapter 13, Theorem 4.5] it follows that these other solutions are of the form $u(x) = k_\beta e^{\beta x} v_\beta + o(e^{\beta x})$ as $x \to \infty$ for some $k_\beta \neq 0$. However, since v_β has components of opposite sign these are not monotone solutions, leaving us with only the one-dimensional manifold of solutions as candidates for monotone connections.

Hence uniqueness is proven and we have completed the proof of Theorem 1.1.

5 Probabilistic results

5.1 Martingales for a branching two-type process

Consider the two-type branching system of particles which was defined in section 1.4 and recall our notation that $I = \{1, 2\}$ and that we often switch between two equivalent notations in this section to reduce the use of subscripts. Thus we will sometimes write, for example, $b(y)$ for b_y, $w(x, y)$ for $w_y(x)$ and $u(t, x, y)$ for $u_y(t, x)$ (for $y = 1, 2$).

Recalling the definitions of B and R in equation (1), let $\lambda < 0$, let $\Lambda_{PF}(\lambda)$ be the Perron-Frobenius eigenvalue of $\lambda B + \theta Q + R$, and let v_λ be the associated eigenvector with $v_\lambda(1) = 1$.

The formal generator $\mathcal{G}$ of the two-type branching process is

$$\mathcal{G} = \mathcal{G}_B + \mathcal{G}_Q + \mathcal{G}_R,$$

where, for $n \geq 1$, $x \in \mathbb{R}^n$, and $y \in I^n$, we have

$$\begin{aligned}
(\mathcal{G}_B F)(n; x; y) &= \sum_{k=1}^{n} b(y_k) \frac{\partial F}{\partial x_k}, \\
(\mathcal{G}_Q F)(n; x; y) &= \theta \sum_{k=1}^{n} \sum_{j \neq y_k; j \in I} Q(y_k, j) \Big\{ F\big(n; x; s_{k,j}(y)\big) - F\big(n; x; y\big) \Big\}, \\
(\mathcal{G}_R F)(n; x; y) &= \sum_{k=1}^{n} r(y_k) \Big\{ F\big(n+1; (x, x_k); (y, y_k)\big) - F\big(n; x; y\big) \Big\},
\end{aligned}$$

where $s_{k,j}(y) := (y_1, \ldots, y_{k-1}, j, y_{k+1}, \ldots, y_n)$ and $(x, x_k) := (x_1, \ldots, x_n, x_k)$, etc. If $F : [0, \infty) \times \mathcal{S}$ (where $\mathcal{S}$ is the state-space of the branching process as defined by equation (7)) and

$$\left\{ \left(\frac{\partial}{\partial t} + \mathcal{G} \right) F \right\} (t; n; x; y) = 0 \qquad (n \geq 1,\, x \in \mathbb{R}^n,\, y \in I^n), \tag{21}$$

then $F\big(t; N(t); X(t); Y(t)\big)$ is a local martingale.
In particular, if u is a $C^{1,1}$ solution of the coupled system (1), define, for $t > 0$,

$$M(s) := \prod_{k=1}^{N(s)} u\big(t - s; X_k(s); Y_k(s)\big). \tag{22}$$

This satisfies the conditions of equation (21) (the partial differential equation enables us to do the time differentiation of M) and hence defines a 'multiplicative' local martingale M on the time-parameter set $[0, t]$. If in addition $0 \leq u(\cdot, x, y) \leq 1$, then $0 \leq M \leq 1$, so that M is a true martingale and

$$\mathbb{E}_{x,y} M(0) = \mathbb{E}_{x,y} M(t),$$

that is,

$$u(t, x, y) = \mathbb{E}_{x,y} \prod_{k=1}^{N(t)} f\big(X_k(t), Y_k(t)\big) = \mathbb{E}_{0,y} \prod_{k=1}^{N(t)} f\big(x + X_k(t), Y_k(t)\big),$$

where $f(x, y) = u(0, x, y)$ is C^1.
Thus we have proved the following:

Theorem 5.1 *If $u \in C^{1,1}$ satisfies the system* (1) *with $0 \leq u(t, x, y) \leq 1$ on $[0, \infty) \times \mathbb{R} \times I$ and with initial condition*

$$u(0, x, y) = f(x, y) \in C^1,$$

then u has a McKean representation

$$u(t, x, y) = \mathbb{E}_{x,y} \prod_{k=1}^{N(t)} f\big(X_k(t), Y_k(t)\big).$$

The above theorem clearly tells us that there is at most 1 bounded (between 0 and 1) solution to the system (1) for smooth, bounded (between 0 and 1) initial data.
Now we look at 'additive' martingales. If $h : [0, \infty) \times \mathbb{R} \times I \to \mathbb{R}$ satisfies the *linear* equation

$$\frac{\partial h}{\partial t} + B\frac{\partial h}{\partial x} + \theta Q h + R h = 0,$$

then, again by using (21),

$$\sum_{k=1}^{N(t)} h\big(t, X_k(t), Y_k(t)\big) \text{is a local martingale.}$$

Now we have:

Theorem 5.2 *If w is a C^1 function on $\mathbb{R} \times I$, then*

$$\prod_{k=1}^{N(s)} w\big(X_k(s) + cs, Y_k(s)\big)$$

is a local martingale if, and only if, w solves the travelling-wave system (2).
If g is a C^1 function on $\mathbb{R} \times I$, then

$$\sum_{k=1}^{N(t)} g\big(X_k(t) + ct, Y_k(t)\big)$$

is a local martingale if, and only if, g solves the linearization of (2) *at the point* $(1, 1)$:

$$(B + c)g' + \theta Q g + R g = 0.$$

This theorem follows from the fact that in (22) M is a local martingale and from the local-martingale property established immediately before Theorem 5.2.

Let $c > c(\theta)$, where $c(\theta)$ is defined in Theorem 1.1. If $c < \max(-b_1, -b_2)$, take λ to be the stable monotone eigenvalue of $K_{c,\theta}(T)$ nearer to 0; if $c = \max(-b_1, -b_2)$, take λ to be the unique generalized stable monotone eigenvalue of $K_{c,\theta}(T)$ and if $c > \max(-b_1, -b_2)$, take λ to be the unique stable monotone eigenvalue of $K_{c,\theta}(T)$. This definition of λ is possible and well-defined by the the work at the start of section 3. Call, for future reference, this definition of λ the *probabilistic* eigenvalue of $K_{c,\theta}(T)$. Hence, if λ is the *probabilistic* eigenvalue of $K_{c,\theta}(T)$ then $-\lambda c$ is the Perron-Frobenius eigenvalue of $\lambda B + \theta Q + R$. Recall the definition of Z_λ in Theorem 1.4,

$$Z_\lambda(t) = \sum_{k=1}^{N(t)} v_\lambda\big(Y_k(t)\big) \exp\Big\{\lambda\big[X_k(t) + ct\big]\Big\}.$$

Then, by Theorem 5.2, Z_λ is an 'additive' local martingale; and since it is non-negative, it is also a supermartingale. We wish to show that Z_λ is a true martingale, this can be done just as for ζ_λ by noting that $\mathbb{E}N(t) \leq e^{r_0 t}$ and each individual term of the sum is again bounded up to any fixed t.

5.2 A large-deviations approach to $c(\theta)$

As we have already observed in Theorem 1.3, the asymptotic behaviour of the position of the left-most particle of the system should give us the wave speed $c(\theta)$. This is proved later in this section; for now we want to give probabilistic heuristics to obtain this value.

If we consider a particle's type intuitively we note that it flicks between the two types with an equilibrium distribution independent of θ. On average the particle will spend a proportion $\frac{q_2}{q_1+q_2}$ of its time with type 1 and a proportion $\frac{q_1}{q_1+q_2}$ of its time with type 2. A larger θ will tend to force each particle's type distribution closer to the equilibrium distribution.

So, we are interested in considering how far from the equilibrium distribution (for time spent in the two states) a particle can deviate. The largest deviations will give the positions of the most extreme particles, including the left-most particle. Clearly, the higher the breeding rate (the parameters r_1, r_2) for each state, the easier it will be to find a particle in that state. As θ is increased the deviations from the equilibrium distribution will be smaller.

The theory of large deviations is the appropriate technique for tackling such a problem. We use the Dirichlet form $\epsilon(u,v) := -u^T \Pi Q v$, where Π is the diagonal matrix whose elements are the components of the equilibrium distribution for the Markov chain. This is a positive-definite, symmetric bilinear form. Then the large-deviation rate functional $J_\epsilon(\mu)$ is $\epsilon(f^{1/2}, f^{1/2})$ where f is the ratio of μ to the invariant distribution and μ is any distribution for time spent in the two states (see Deuschel and Stroock [7, page 129]). So, if μ corresponds to spending a proportion of time m_1 in state 1, then $J_\epsilon(\mu) = (\sqrt{q_1 m_1} - \sqrt{q_2 m_2})^2$, and hence enables us to state that the probability that a particle, at time t, has spent a proportion m_1 of its time in state 1 and a proportion m_2 of its time in state 2 (for $m_1, m_2 \geq 0, m_1 + m_2 = 1$) is approximately $\exp\big(-t\theta(\sqrt{q_1 m_1} - \sqrt{q_2 m_2})^2\big)$. Thus, for all proportions other than the equilibrium proportions this decays exponentially and an increase in θ increases the rate of decay. Now, this decay is balanced by the breeding — the number of particles grows exponentially at a rate $m_1 r_1 + m_2 r_2$. So, when

$$m_1 r_1 + m_2 r_2 - \theta(\sqrt{q_1 m_1} - \sqrt{q_2 m_2})^2 > 0 \tag{23}$$

there is exponential growth in the number of particles who have spent a proportion m_1 of their time in state 1 and a proportion m_2 of their time in state 2; when this is strictly less than zero there is decay in the number of such particles. This leads us to define $\tilde{c}(\theta)$ as:

$$\tilde{c}(\theta) = \quad -\inf\Big\{ b_1 m_1 + b_2 m_2 : \quad m_1 r_1 + m_2 r_2 - \theta(\sqrt{q_1 m_1} - \sqrt{q_2 m_2})^2 > 0;$$

$$m_1, m_2 \geq 0, m_1 + m_2 = 1\Big\},$$

where m_1 and m_2 represent the proportions of times in the two states. This is well-defined since for $m_1 = \frac{q_2}{q_1+q_2}$ (the equilibrium distribution) the inequality is certainly satisfied.

To demonstrate the correspondence of $c(\theta)$ and $\tilde{c}(\theta)$ we examine some specific cases.

Firstly if $b_1 = b_2 = b$ then the above formula gives $\tilde{c}(\theta) = -b$, which tallies with the result for $c(\theta)$ we obtained in section 3. As before we can now assume, without loss of generality, that $b_1 > b_2$. For the equilibrium proportions the inequality (23) is satisfied, so we immediately have that $c^* = -(\frac{b_1 q_2 + b_2 q_1}{q_1+q_2}) \leq \tilde{c}(\theta) \leq -b_2$.

So we ask, when is $\tilde{c}(\theta) = -b_2$? Putting $m_2 = 1$ the inequality (23) becomes $r_2 - \theta q_2 > 0$, so $r_2 > \theta q_2$ implies that $\tilde{c}(\theta) = -b_2$. For $r_2 = \theta q_2$ we use the concavity of $m_1 r_1 + m_2 r_2 - \theta(\sqrt{q_1 m_1} - \sqrt{q_2 m_2})^2$ to observe that the inequality (23) is satisfied for $m_2 = 1 - \epsilon$ for all ϵ in the interval $0 < \epsilon < \frac{q_2}{q_1+q_2}$, thus the infimum is still $m_2 = 1$ — just it is not attained (this corresponds to the fact that in this case there is a travelling wave corresponding to $\tilde{c}(\theta)\big(= c(\theta)\big)$, for $r_2 > \theta q_2$ there is not).

For $r_2 < \theta q_2$, $\tilde{c}(\theta) < -b_2$ and so all that remains to prove is that $\tilde{c}(\theta) = c(\theta)$ in this case too. We can rearrange our formula for $\tilde{c}(\theta)$ by writing $c = -b_1 m_1 + b_2(m_1 - 1)$ and hence $m_1 = \frac{b_2+c}{b_2-b_1}$. Substituting this into the defining inequality we note that

$$m_1 r_1 + m_2 r_2 - \theta(\sqrt{q_1 m_1} - \sqrt{q_2 m_2})^2 = 0$$

implies that the term under the square root in equation (16) is zero. Now we note that this corresponds to the zero with the larger value of c since $\tilde{c}(\theta)$ is at least c^*.

This formula for $c(\theta)$ makes it easier to observe that we can have $m_i = 1$ only when $r_i > \theta q_i$, which makes intuitive sense — the breeding in state i needs to be larger than the mutation out of the state for there to be a persistent family of that type of particle over time. It also makes it clear that $c(\theta)$ decreases from $\max(-b_1, -b_2)$ to c^* as $\theta \to \infty$.

5.3 Convergence properties of Z_λ martingales

The full assumptions which we have made about Z_λ (defined in Theorem 1.4) — that $c > c(\theta)$, that λ is the *probabilistic* eigenvalue of $K_{c,\theta}$, etc. — will now be needed in proving that Z_λ converges in $\mathcal{L}^p$ for some $p > 1$ (and hence in $\mathcal{L}^1$). According to Doob's $\mathcal{L}^p$ inequality (see Rogers and Williams [14], Theorem II.70.2), we need only show that Z_λ is bounded in $\mathcal{L}^p$ for some $p > 1$. The key result is from Neveu [13] and the method of using it follows Champneys et al. [2].

Lemma 5.3 (Neveu) *Let $p \in (1, 2]$. For any finite sequence $W_1, \ldots, W_n$ of non-negative independent variables in $\mathcal{L}^p$ and any sequence $c_1, \ldots, c_n$ of non-negative real numbers,*

$$\psi\left(\sum_{k=1}^{n} c_k W_k\right) \leq \sum_{k=1}^{n} c_k^p \psi(W_k),$$

where $\psi(W) := \mathbb{E}(W^p) - \big(\mathbb{E}(W)\big)^p$ for $W \in \mathcal{L}^p$.

Proof of $\mathcal{L}^1$ convergence of Z_λ. Fix $t > 0$. Because of the branching character of the (N, X, Y) process, we have for each $s > 0$,

$$Z_\lambda(s+t) = \sum_{k=1}^{N(s)} \exp\big\{\lambda\,[X_k(s) + cs]\big\} W_k(t, s)$$

where, conditionally on $\mathcal{F}_s$, the $W_k(t,s)$ are independent, each with the $\mathbb{P}_{0,y(k)}$ law of $Z_\lambda(t)$ where $y(k) = Y_k(s) \in I$. Since t is fixed, and I is finite, Neveu's lemma, applied conditionally on $\mathcal{F}_s$, gives

$$\mathbb{E}_{x,y}\{Z_\lambda(s+t)^p \mid \mathcal{F}_s\} - Z_\lambda(s)^p \le K_1(t,x,\lambda) \sum_{k=1}^{N(s)} \exp\{\lambda p\,[X_k(s)+cs]\},$$

so that, on taking expectations,

$$\mathbb{E}_{x,y}\{Z_\lambda(s+t)^p\} - \mathbb{E}_{x,y}\{Z_\lambda(s)^p\} \le K_1(t,x,\lambda)\mathbb{E}_{x,y}\left\{\sum_{k=1}^{N(s)} \exp\{\mu\,[X_k(s)+cs]\}\right\},$$

where $\mu := \lambda p < \lambda < 0$. We choose p in $(1,2]$ sufficiently close to 1 that (see Lemma 3.2) $c > c_1$, where $-\mu c_1 = \Lambda_{PF}(\mu B + \theta Q + R)$ and v_μ is the corresponding eigenvector with $v_\mu(1) = 1$. But then

$$\begin{aligned}
&\mathbb{E}_{x,y} \sum_{k=1}^{N(s)} \exp\{\mu\,[X_k(s)+cs]\} \\
\le\ & K_2(x,\mu,y)\exp\{\mu(c-c_1)s\}\mathbb{E}_{x,y}\sum_{k=1}^{N(s)} v_\mu\big(Y_k(s)\big)\exp\{\mu\,[X_k(s)+c_1 s]\} \\
\le\ & K_3(x,\mu,y)\exp\{\mu(c-c_1)s\},
\end{aligned}$$

since Z_μ is a martingale. Hence,

$$\begin{aligned}
&\sum_m \mathbb{E}_{x,y}\{Z_\lambda(ms+s)^p - Z_\lambda(ms)^p\} \\
\le\ & K_1(s,x,\lambda)K_3(x,\mu,y)\sum_m \exp\{\mu m(c-c_1)s\} < \infty,
\end{aligned}$$

and Z_λ is bounded in $\mathcal{L}^p$. □

Now we prove, with the same notation and assumptions, that

$$w(y) := \mathbb{P}_{x,y}\big(Z_\lambda(\infty) = 0\big) = 0 \text{ for all } (x,y). \tag{24}$$

(The fact that $w(y)$ does not depend on x is obvious.)

Proof of (24). Let J be the first jump time of Y_1 and let T be the first branch time of (N,X,Y). On decomposing $w(y)$ according as $T < J$ or $T > J$, we obtain

$$w(y) = \frac{r(y)w(y)^2 + \theta\sum_{z\ne y} Q(y,z)w(z)}{r(y)+\theta q(y)},$$

so that $Rw = R(w^2) + \theta Q w$. By Lemma 3.4, $w \equiv 0$ on I or $w \equiv 1$ on I. When Z_λ converges in $\mathcal{L}^1$, then, obviously, $w \equiv 0$ on I. □

The following theorem is now proven.

Theorem 5.4 *Let $c > c(\theta)$. Let λ be the* probabilistic *eigenvalue of $K_{c,\theta}(T)$. Thus $-\lambda c$ is the Perron-Frobenius eigenvalue of $\lambda B + \theta Q + R$ and, as usual, we denote by v_λ the corresponding eigenvector with $v_\lambda(1) = 1$. Then*

$$Z_\lambda(t) = \sum_{k=1}^{N(t)} v_\lambda\big(Y_k(t)\big)\exp\{\lambda\,[X_k(t)+ct]\}.$$

is a true (not just a local) martingale, and $Z_\lambda(t)$ converges to a limit $Z_\lambda(\infty)$ almost surely and in $\mathcal{L}^1$. Moreover, $\mathbb{P}_{x,y}(Z_\lambda(\infty) > 0) = 1$ for all x and y.

For the proof of the uniqueness modulo translation of the monotone travelling wave from S to T, we need the following result. When there is only one stable monotone eigenvalue of $K_{c,\theta}(T)$ we have no other candidate for a wave (each stable monotone eigenvalue gives us a martingale which could correspond to a suitable wave), when there are two we use this result to rule out the *other* possibility.

Lemma 5.5 *Suppose that $c > c(\theta)$ and that there are two stable monotone eigenvalues of $K_{c,\theta}(T)$. Denote by β the eigenvalue further from 0, and by v_β the associated Perron-Frobenius eigenvector of $\beta B + \theta Q + R$ with $v_\beta(1) = 1$. Then, almost surely,*

$$Z_\beta(t) = \sum_{k=1}^{N(t)} v_\beta\big(Y_k(t)\big) \exp\big\{\beta\,[X_k(t) + ct]\big\} \to 0$$

as $t \to \infty$.

Proof. We prove this result by modifying an argument in Neveu [13]. Let $0 < p < 1$. Then for $u, v > 0$,

$$(u+v)^p \leq u^p + v^p.$$

Again, let J be the first jump time of Y_1 and let T be the first branch time of (N, X, Y). The decomposition

$$Z_\beta(\infty) = \begin{cases} \exp\big\{\beta[X_1(J) + cJ]\big\} Z_\beta^{(1)}(\infty) & \text{if } J < T, \\ \exp\big\{\beta[X_1(T) + cT]\big\}\big[Z_\beta^{(2)}(\infty) + Z_\beta^{(3)}(\infty)\big] & \text{if } T < J, \end{cases}$$

leads to the formula

$$\begin{aligned} g(y) := \mathbb{E}_{0,y}\,[Z_\beta(\infty)^p] \quad &\leq \quad \mathbb{E}_{0,y} \exp\big\{\alpha[X_1(J) + cJ]\big\} I_{J<T}\, g\big(Y_1(J)\big) \\ &+ \quad 2\mathbb{E}_{0,y} \exp\big\{\alpha[X_1(T) + cT]\big\} I_{T<J}\, g\big(Y_1(T)\big), \end{aligned}$$

where $\alpha := p\beta$. On evaluating these expectations, we obtain

$$g(y) \leq \frac{\Big\{\theta \sum_{z\neq y} Q(y,z) g(z)\Big\} + 2r(y)g(y)}{-\big(b(y) + c\big)\alpha + r(y) + \theta q(y)},$$

which rearranges to give

$$0 \leq \big((B + cI)\alpha + \theta Q + R\big) g.$$

We know that $g \geq 0$ on I. Lemma 3.3 shows that if we choose p sufficiently close to 1, then $g = 0$. The lemma is proved. □

5.4 Proof of Theorem 1.4

We now verify steps in the proof of Theorem 1.4, which lead to the important equality (8) of Theorem 1.3:

$$\lim t^{-1} L(t) = -c(\theta) \quad \text{(a.s.)} \quad \text{where } L(t) = \inf_{k \leq N(t)} X_k(t).$$

Part (i). Let $c > c(\theta)$, and let λ be the *probabilistic* eigenvalue of $K_{c,\theta}(T)$. Let Z_λ be the associated martingale. We see by considering the position of the left-most particle that

$$Z_\lambda(t) = \sum_{k=1}^{N(t)} v_\lambda\big(Y_k(t)\big) \exp\big\{\lambda\,[X_k(t) + ct]\big\} \geq \min\big(v_\lambda(1), v_\lambda(2)\big) \exp\big\{\lambda\,[L(t) + ct]\big\}.$$

Since $Z_\lambda(\infty)$ exists a.s., $\liminf [L(t)+ct] > -\infty$, a.s., so that

$$\liminf t^{-1}L(t) \geq -c = \lambda^{-1}\Lambda_{PF}(\lambda).$$

Part (ii). Before Theorem 1.4 was stated we noted that we would consider working with the $\mathbb{P} = \mathbb{P}_{0,1}$ law, so that $Z_\lambda(0) = 1$. Since Z_λ converges in $\mathcal{L}^1$, we can define a probability measure Q_λ on $\mathcal{F}_\infty$ via

$$dQ_\lambda/d\mathbb{P} = Z_\lambda(\infty) \text{ on } \mathcal{F}_\infty, \qquad \text{whence } \ dQ_\lambda/d\mathbb{P} = Z_\lambda(t) \text{ on } \mathcal{F}_t.$$

Define

$$M_\lambda(t) := Z_\lambda(t)^{-1}\frac{\partial}{\partial\lambda}Z_\lambda(t).$$

Because $(\partial/\partial\lambda)Z_\lambda(t)$ is a $\mathbb{P}$-martingale, M_λ is a Q_λ-martingale.
For $t \geq 0$ and $1 \leq k \leq N(t)$, define

$$H(t,k) := \frac{v_\lambda\big(Y_k(t)\big)\exp\big[\lambda X_k(t) - \Lambda_{PF}(\lambda)t\big]}{\sum_{j=1}^{N(t)} v_\lambda\big(Y_j(t)\big)\exp\big[\lambda X_j(t) - \Lambda_{PF}(\lambda)t\big]}.$$

Note that $H(t,k) \geq 0$ and $\sum_j H(t,j) = 1$. Now (denoting by prime differentiation with respect to λ),

$$M_\lambda(t) = \sum_{k=1}^{N(t)} H(t,k)\Big\{u_\lambda\big(Y_k(t)\big) + X_k(t) - \Lambda'_{PF}(\lambda)t\Big\},$$

where $u_\lambda(j) := v'_\lambda(j)/v_\lambda(j)$, so that

$$t^{-1}M_\lambda(t) \geq t^{-1}\{\min_{i\in I} u_\lambda(i)\} + t^{-1}L(t) - \Lambda'_{PF}(\lambda). \tag{25}$$

By Jensen's inequality,

$$M_\lambda(t)^2 \leq \sum H(t,k)\Big\{u_\lambda\big(Y_k(t)\big) + X_k(t) - \Lambda'_{PF}(\lambda)t\Big\}^2. \tag{26}$$

However,

$$Z_\lambda(t)^{-1}\frac{\partial^2}{\partial\lambda^2}Z_\lambda(t)$$

is a Q_λ-martingale, and clearly

$$\begin{aligned} Z_\lambda(t)^{-1}\frac{\partial^2}{\partial\lambda^2}Z_\lambda(t) &= \sum H(t,k)\Big\{u_\lambda\big(Y_k(t)\big) + X_k(t) - \Lambda'_{PF}(\lambda)t\Big\}^2 \\ &+ \sum H(t,k)\Big\{u'_\lambda\big(Y_k(t)\big) - \Lambda''_{PF}(\lambda)t\Big\}. \end{aligned}$$

Thus, (26) shows that the Q_λ-expectation of $M_\lambda(t)^2$ satisfies the inequality

$$Q_\lambda\left[M_\lambda(t)^2\right] \leq Q_\lambda\left[Z_\lambda(t)^{-1}\frac{\partial^2}{\partial\lambda^2}Z_\lambda(t)\right] - Q_\lambda\left[\sum H(t,k)\Big\{u'_\lambda\big(Y_k(t)\big) - \Lambda''_{PF}(\lambda)t\Big\}\right].$$

The first term on the right-hand side is constant by the martingale property so that

$$Q_\lambda\left[M_\lambda(t)^2\right] \leq K_1(\lambda) - Q_\lambda\left[\sum H(t,k)\Big\{\max\big(u'_\lambda(1), u'_\lambda(2)\big) - \Lambda''_{PF}(\lambda)t\Big\}\right].$$

Since $\sum_j H(t,j) = 1$ further simplification gives

$$Q_\lambda[M_\lambda(t)^2] \leq K_2(\lambda) + K_3(\lambda)t$$

for finite constants $K_1(\lambda)$, $K_2(\lambda)$ and $K_3(\lambda)$, independent of t. Hence, for $\epsilon > 0$,

$$Q_\lambda\left(\sup\left\{s^{-1}|M_\lambda(s)| : 2^{n-1} \le s \le 2^n\right\} \ge \epsilon\right)$$

$$\le Q_\lambda\left(\sup_{s\le 2^n} |M_\lambda(s)| \ge \epsilon 2^{n-1}\right) \le \left(\epsilon 2^{n-1}\right)^{-2}\left[K_2(\lambda) + 2^n K_3(\lambda)\right],$$

by Doob's submartingale inequality. By the Borel-Cantelli lemma, we have $t^{-1}M_\lambda(t) \to 0$, a.s., whence, from (25),

$$\limsup_{t\to\infty} t^{-1}L(t) \le \Lambda'_{PF}(\lambda).$$

Part (ii) of Lemma 3.2 clinches Part (iii) of Theorem 1.4, and the proof of (8) is complete. □

5.5 Heaviside initial conditions

We have the McKean representation

$$u(t,x,y) = \mathbb{E}_{x,y}\prod_{k=1}^{N(t)} u\big(0, X_k(t), Y_k(t)\big)$$

for the unique (by the work of section 2) solution of our coupled equation (1) when $0 \le u \le 1$ and the initial data $u(0,\cdot,\cdot)$ are sufficiently smooth. We would like to obtain a McKean representation in the case of the Heaviside initial data. We verify this directly below then, because $t^{-1}L(t) \to -c(\theta)$ (a.s.), the rest of Theorem 1.3 is obvious. In a separate paper (currently in preparation), Lyne and Williams will

- give a new direct analytic method of proving existence and uniqueness of *weak* solutions with bounded, measurable initial data,
- show that these solutions lead to martingales and hence to a McKean representation,
- give numerical and simulation studies of the solutions with Heaviside initial data, and
- study the long-term behaviour of these *Heaviside* solutions.

We now verify that, *defining* $u(t,x,y)$ to be $\mathbb{P}_{x,y}[L(t) > 0]$, gives a *weak* solution of the coupled equation (1) with Heaviside initial data. Conditioning on the first jump

$$u(t,x,y) = \exp\Big\{-\big(r(y)+\theta q(y)\big)t\Big\}u\big(0, x+b(y)t, y\big)$$

$$+\int_{s=0}^{t}\Big\{\exp\Big(-\big(r(y)+\theta q(y)\big)s\Big)\Big\}\Big\{r(y)u^2\big(t-s, x+b(y)s, y\big) + \theta q(y)u\big(t-s, x+b(y)s, \hat{y}\big)\Big\}ds$$

where $\{y, \hat{y}\} = \{1, 2\}$. Integrating by parts and comparing the terms with those obtained by differentiating this expression for $u(t,x,y)$ by each of t and x yields

$$u(t,x,y) = \frac{r(y)}{r(y)+\theta q(y)}u^2(t,x,y) + \frac{\theta q(y)}{r(y)+\theta q(y)}u(t,x,\hat{y}) + \frac{b(y)\partial_x - \partial_t}{r(y)+\theta q(y)}u(t,x,y)$$

hence $u(t,x,y)$ satisfies equation (1). Clearly this expression for u corresponds to the Heaviside initial conditions.

We can observe several features of the solution from this representation. Since the left-most particle must lie in the interval $[x + t\min(b_1,b_2), x + t\max(b_1,b_2)]$ at time t then the solution is identically zero for

$x \le -t\max(b_1, b_2)$ and is identically one for $x > -t\min(b_1, b_2)$. It is clear that the solution retains the left-continuity of the initial data. Also, $u(t, x, y)$ must be strictly increasing (in x) inside this region.

The case $b_1 = b_2 = b$, say, is very simple. All particles will travel at speed b, so that, at time t, all particles will be at $x + bt$. Hence, $L(t) = x + bt$, so $\mathbb{P}_{x,y}[L(t) > 0] = I_{\{x>-bt\}}$. This is simply the Heaviside function travelling at speed $-b$. This corresponds to the analysis in section 2.6.

To deal with $b_1 \ne b_2$ assume from now on, without loss of generality, that $b_1 > b_2$.

Note that the only way that the left-most particle can be at position $x + b_1 t$ at time $t > 0$ is if the first particle was of type 1 and there have been no mutations to type 2 by the particle or any of its descendants so far. The probability of this event occurring when the initial particle is of type 1 is

$$\frac{r_1 + \theta q_1}{r_1 + \theta q_1 \exp\big((r_1 + \theta q_1)t\big)},$$

which explains why $u(t, -t+)$ equals this in the calculation in section 2.4.

The only way that the left-most particle can be at position $x + b_2 t$ at time $t > 0$ is if the first particle was of type 2 and the embedded birth and death process on this line has not become extinct. This embedded process is that constructed by considering a particle to die when it mutates to type 2. Thus, on $x = -b_2 t$, for $t > 0$, u_1 is identically zero. The probability of this event occurring when the initial particle is of type 2 is exactly the expression calculated for $v(t, +t)$ in section 2.4, and the limit as $t \to \infty$ (which is $\min(\frac{\theta q_2}{r_2}, 1)$) is the probability of eventual extinction.

More generally it is natural to expect that the McKean representation will tie up with *weak* solutions of our coupled equation (1). This is because they are piecewise classical and any discontinuities (corresponding to atoms of probability in the distribution of the particles' positions) will naturally propagate only along characteristics (since the atoms of probability travel at b_1 and b_2). Thus a solution satisfying the McKean representation satisfies the Rankine-Hugoniot conditions, hence is a *weak* solution (see section 2.3).

5.6 Polishing off the probability

Theorem 5.6 *Assume $c > c(\theta)$, let λ be the* probabilistic *eigenvalue of $K_{c,\theta}(T)$, and as usual denote by v_λ be the corresponding eigenvector with $v_\lambda(1) = 1$. Then if u satisfies the coupled system* (1), *$u \in C^{1,1}$ and satisfies $0 \le u \le 1$, and if also, for $y \in I$,*

$$1 - u(0, x, y) \sim v_\lambda(y)e^{\lambda x} \qquad (x \to \infty),$$

then, as $t \to \infty$,

$$u(t, x + ct, y) \to w(x, y),$$

where

$$w(x, y) := \mathbb{E}_{x,y} \exp\big[-Z_\lambda(\infty)\big].$$

This function w satisfies the travelling-wave equation (2) *and is, modulo translations, the unique monotone wave of speed c from S to T.*

Proof. We are guided by McKean [11]. So, we are supposing that u solves (1), that $0 \le u \le 1$ and that

$$1 - u(0, r, y) \sim v_\lambda(y)e^{\lambda r} \qquad (r \to \infty). \tag{27}$$

For (temporarily) fixed $\epsilon > 0$, we have for large r,

$$\exp\left\{-(1+\epsilon)v_\lambda(y)e^{\lambda r}\right\} \le u(0, r, y) \le \exp\left\{-(1-\epsilon)v_\lambda(y)e^{\lambda r}\right\}.$$

Now since $L(t) + ct \to \infty$ (a.s.), we shall (a.s.) have for large t,

$$\exp\big\{-(1+\epsilon)Z_\lambda(t)\big\} \le \prod_{k=1}^{N(t)} u\big(0, X_k(t) + ct, Y_k(t)\big) \le \exp\big\{-(1-\epsilon)Z_\lambda(t)\big\}.$$

Thus we have that

$$\begin{aligned}
\liminf_{t\to\infty}\ \exp\big\{-(1+\epsilon)Z_\lambda(t)\big\} &\le \liminf_{t\to\infty}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \limsup_{t\to\infty}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \limsup_{t\to\infty}\ \exp\big\{-(1-\epsilon)Z_\lambda(t)\big\}.
\end{aligned}$$

Since $Z_\lambda(\infty)$ exists almost surely (and clearly all the terms in the sequence of inequalities are bounded below by 0 and above by 1) this becomes

$$\begin{aligned}
\exp\big\{-(1+\epsilon)Z_\lambda(\infty)\big\} &\le \liminf_{t\to\infty}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \limsup_{t\to\infty}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \exp\big\{-(1-\epsilon)Z_\lambda(\infty)\big\}.
\end{aligned}$$

After taking expectations and using Fatou's Lemma, this implies that

$$\begin{aligned}
\mathbb{E}_{x,y}\exp\big\{-(1+\epsilon)Z_\lambda(\infty)\big\} &\le \mathbb{E}_{x,y}\liminf_{t\to\infty}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \liminf_{t\to\infty}\mathbb{E}_{x,y}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \limsup_{t\to\infty}\mathbb{E}_{x,y}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \mathbb{E}_{x,y}\limsup_{t\to\infty}\prod_{k=1}^{N(t)} u\big(0, X_k(t)+ct, Y_k(t)\big)\\
&\le \mathbb{E}_{x,y}\exp\big\{-(1-\epsilon)Z_\lambda(\infty)\big\}.
\end{aligned}$$

Using the McKean representation (4.19) this yields that

$$\begin{aligned}
\mathbb{E}_{x,y}\exp\big\{-(1+\epsilon)Z_\lambda(\infty)\big\} &\le \liminf_{t\to\infty} u(t, x+ct, y)\\
&\le \limsup_{t\to\infty} u(t, x+ct, y)\\
&\le \mathbb{E}_{x,y}\exp\big\{-(1-\epsilon)Z_\lambda(\infty)\big\}.
\end{aligned}$$

On letting $\epsilon \downarrow 0$, we now obtain the desired result

$$u(t, x+ct, y) \to w(x,y) = \mathbb{E}_{x,y}\exp\{-Z_\lambda(\infty)\}. \tag{28}$$

Existence of a monotone travelling wave from S to T when $c > c(\theta)$. It is now intuitively obvious, and not that difficult to prove directly from the branching property, that the function $w(\cdot,\cdot)$ in (28) is a monotone travelling wave from S to T. Firstly it does in fact satisfy the travelling wave equation (2) because it can be written in the form

$$\begin{aligned}
w(x,y) = \int_{t=0}^{\infty}\Big(\exp\Big(-\big(r(y)+\theta q(y)\big)t\Big)\Big)\ &\Big\{r(y)w^2\big(x+(b(y)+c)t, y\big)\\
&+\theta q(y)w\big(x+(b(y)+c)t, \hat{y}\big)\Big\}dt
\end{aligned}\tag{29}$$

where $\{y, \hat{y}\} = \{1, 2\}$. Integrating by parts gives

$$w(x,y) = \frac{r(y)}{r(y)+\theta q(y)} w^2(x,y) + \frac{\theta q(y)}{r(y)+\theta q(y)} w(x,\hat{y}) + \frac{b(y)+c}{r(y)+\theta q(y)} w'(x,y)$$

since the last term in the integration by parts is simply a multiple of the x-derivative of equation (29). Then note that

$$w(x,y) = \mathbb{E}_{0,y} \exp\{-e^{\lambda x} Z_\lambda(\infty)\},$$

(from the definitions of $\mathbb{E}_{x,y}$ and Z), that $w(x,y) \to 0$ as $x \to -\infty$ (because $Z_\lambda > 0$ (a.s.) and $\lambda < 0$), and that

$$1 - w(x,y) \sim v_\lambda(y) e^{\lambda x} \qquad (x \to \infty) \tag{30}$$

because Z_λ converges $\mathcal{L}^1$ (by Theorem 5.4). Hence

$$\mathbb{E}_{x,y} Z_\lambda(\infty) = \mathbb{E}_{x,y} Z_\lambda(0) = e^{\lambda x} v_\lambda(y).$$

Hence our claim is proved.

Uniqueness modulo translation of the monotone travelling wave from S to T. Let $c > c(\theta)$, and let $\tilde{w}$ be a monotone travelling wave from S to T. We know from differential-equation theory that

either a suitable translate of $\tilde{w}$ satisfies (30), *or* a suitable translate of $\tilde{w}$ satisfies

$$1 - \tilde{w}(x,y) \sim v_\beta(y) e^{\beta x} \qquad (x \to \infty), \tag{31}$$

where β is the monotone eigenvalue of $K_{c,\theta}$ further from 0.

If $\tilde{w}$ satisfies (2) and (30), then $u(t,x,y) := \tilde{w}(x - ct, y)$ satisfies (1) and (27), so that from (28), we must have $\tilde{w} = w$. If $\tilde{w}$ satisfied (31), then we would have

$$\tilde{w}(x,y) = \mathbb{E}_{x,y} \exp\{-Z_\beta(\infty)\} = 1,$$

because $Z_\beta(\infty) = 0$ (a.s.) by Lemma 5.5; and $\tilde{w}$ would not go from S to T.

The proof of uniqueness is now complete, so we have proven Theorem 5.6. $\square$

It is interesting to compare the above probabilistic proofs of existence and uniqueness modulo translation of travelling waves with the analytic proofs given in section 4. We could, for example, use ODE results to obtain results on $\mathcal{L}^1$ convergence of our martingales.

The methods in section 4 dealt with the $c = c(\theta)$ case; the probability theory for this case should be amenable to the techniques of Neveu [13] — stopping lines will correspond to simple conditions on the occupation times of the two states of the Markov chain.

Acknowledgements

I would like to thank Michael Shearer for ideas in section 2, Elaine Crooks for ideas in section 3, and my supervisors — Professors John Toland and David Williams — for their invaluable assistance. I would also like to thank the EPSRC for financial support in the form of a Research Studentship, my PhD examiners, Jey Sivaloganathan and Alison Etheridge, for their comments on my thesis from which this work is taken, and Geoffrey Grimmett for commenting on drafts of this paper. Finally I would like to thank the anonymous referees who examined the paper so carefully.

References

[1] Beale, J. T. (1986) Large-Time Behavior of Discrete Velocity Boltzmann Equations. *Commun. Math. Phys.*, **106**, 659–678.

[2] Champneys, A., Harris, S., Toland, J., Warren, J. & Williams, D. (1995) Algebra, analysis and probability for a coupled system of reaction-diffusion equations. *Phil. Trans. R. Soc. Lond.*, A **350**, 69–112.

[3] Coddington, E. A. & Levinson, N. (1955) *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.

[4] Cohen, J.E. (1981) Convexity of the dominant eigenvalue of an essentially non-negative matrix. *Proc. Amer. Math. Soc.*, **81**, 657–658.

[5] Courant, R. & Hilbert, D. (1962) *Methods of Mathematical Physics, Volume II.* New York: Interscience.

[6] Crooks, E. C. M. (1996) On the Vol'pert theory of travelling-wave solutions for parabolic systems. *Nonlinear Analysis, Theory, Methods and Applications.* **26**, 1621–1642.

[7] Deuschel, J-D. & Stroock, D.W. (1989) *Large Deviations.* Boston: Academic Press.

[8] Dunbar, S.R. (1988) A branching random evolution and a nonlinear hyperbolic equation. *SIAM J. Appl. Math.*, **48**, 1510–1526.

[9] Hadeler, K. P. (1995) Travelling Fronts in Random Walk Systems. *Forma*, **10**, 223–233.

[10] Holmes, E. E. (1993) Are diffusion models too simple? A comparison with telegraph models of invasion. *Amer. Naturalist*, **142**, 779–795.

[11] McKean, H. P. (1975) Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Comm. Pure Appl. Math.*, **28**, 323–331.

[12] McKean, H. P. (1976) Correction to the above. *Comm. Pure Appl. Math.*, **29**, 553–554.

[13] Neveu, J. (1987) Multiplicative martingales for spatial branching processes. *Seminar on Stochastic Processes* (ed. E. Çinlar, K. L Chung and R. K. Getoor), Progress in Probability and Statistics **15**. pp. 223–241. Boston: Birkhäuser.

[14] Rogers, L. C. G. & Williams, D. (1994) *Diffusions, Markov Processes and Martingales, Volume 1: Foundations*, 2nd edition. Chichester: Wiley.

[15] Seneta, E. (1981) *Non-negative Matrices and Markov Chains.* Heidelberg, New York: Springer-Verlag.