CMU lecture notes
Topics in Branching Brownian motion and parabolic equations

Lenya Ryzhik

June 16, 2026

Abstract

These notes are not proofread sufficiently carefully. They have also not been checked by any LLM for consistency of language or notation, or for the correctness of the proofs. Use with caution.

Contents

1 The overall plan **3**

2 Lecture 1: BBM and semi-linear parabolic equations **3**

2.1 Overview of the lecture . . . 3
2.2 The definition of a branching Brownian motion . . . 4
2.3 The total number of particles of a BBM . . . 5
2.3.1 The average total number of particles . . . 5
2.3.2 A simplest martingale . . . 6
2.4 Linear parabolic equations and branching Brownian motion . . . 7
2.5 McKean's interpretation of the Fisher-KPP equation . . . 9
2.6 Not all Fisher-KPP nonlinearities are of the McKean type . . . 10
2.7 Voting schemes for semi-linear equations . . . 11
2.7.1 The Etheridge-Freeman-Penington model for the Allen-Cahn equation . . . 12
2.7.2 Random outcome probabilistic voting models . . . 13
2.7.3 The standard heat equation and unbiased voting . . . 14
2.7.4 Representing general nonlinearities . . . 15
2.7.5 A random threshold voting model . . . 16
2.7.6 Recursive up the tree propagation models for other nonlinearities . . . 18
2.8 Examples of voting models . . . 19
2.8.1 Random outcome voting models for the McKean nonlinearities . . . 19
2.8.2 Uniformly biased voting models . . . 20
2.8.3 Group voting models . . . 20
2.9 The up the tree voting models revisited and nonlinear integral equations . . . 21
2.9.1 The multiplication voting rule . . . 21
2.9.2 The moments and convergence to $\delta(\xi)$ . . . 22
2.9.3 Logarithmic moments . . . 23
2.9.4 Diffusion in the logarithmic variables . . . 23

3 Lecture 2: The frontier of BBM in one and higher dimensions **24**
3.1 Overview of the lecture . . . 24
3.2 Traveling waves for positive nonlinearities . . . 25
3.3 The Bramson shift and convergence to a wave . . . 26
3.3.1 Steepness and the KPP convergence result . . . 26
3.3.2 The Bramson shift and the law of the maximum of BBM . . . 27
3.3.3 The Bramson correction and other log-correlated random fields . . . 29
3.3.4 Strategy of the PDE proof of Theorem 3.3.3 . . . 31
3.4 The randomized Gumbel distribution of the maximum of BBM . . . 32
3.4.1 The traveling wave martingale . . . 33
3.4.2 The additive and derivative martingales . . . 34
3.4.3 The limit of the derivative martingale . . . 36
3.4.4 The law of the maximum and the derivative martingale . . . 38
3.4.5 The ergodic theorem and the limit of the derivative martingale . . . 39
3.5 Anisotropy in higher dimensions . . . 40
3.6 Proof outline: excursions and the derivative martingale . . . 42
3.7 Large excursions . . . 45
3.7.1 A connection to the Fisher–KPP equation . . . 45
3.7.2 A Liouville theorem for the ancient solution . . . 46

4 Lecture 3. Bounds for thermo-diffusive systems **49**
4.1 Thermo-diffusive systems . . . 49
4.2 A voting model for the thermo-diffusive system . . . 50
4.3 The special case $\kappa = \nu$ (a single equation reduction) . . . 51
4.3.1 Spreading in a single reaction-diffusion equation . . . 51
4.3.2 Spreading results for a non-local equation . . . 52
4.4 The surprisingly simple case $\nu \geq \kappa$ . . . 52
4.5 Known uniform bounds on the temperature . . . 54
4.6 The main result: a global in time uniform bound . . . 55
4.7 Different diffusivities but identical forcing: the parabolic BMO bound . . . 56
4.8 From bounded averages to bounded pointwise values . . . 57
4.9 Comparison to a single reaction-diffusion equation . . . 59
4.9.1 The Michel Pierre argument revisited . . . 59
4.9.2 A lower bound by a single equation if $\nu \leq \kappa$ . . . 59
4.10 The hair trigger effect . . . 60
4.11 A forward in time bound on the fuel concentration . . . 63
4.12 Backward in time bound on the fuel temperature averages . . . 64
4.13 A bound on the averages . . . 65
4.13.1 Reduction to a Hölder in time bound . . . 65
4.14 A bound on the spatial variations of the fuel in a "parabola" . . . 66

5 Lecture 4: Diffusion of knowledge models in macroeconomics **69**
5.1 The Lucas-Moll model . . . 69
5.2 The basic models of diffusion of knowledge . . . 70
5.3 The dynamics of agents . . . 70
5.3.1 Acquisition of knowledge . . . 71
5.3.2 The macroscopic dynamics . . . 72
5.3.3 Modeling the production . . . 73

5.3.4 Innovation and diffusion of knowledge . . . 73
5.3.5 Evolution of the cumulative distribution function . . . 73
5.3.6 A digression on a very special non-local case . . . 74
5.3.7 Similarity to N-BBM . . . 75
5.3.8 Possible prescriptions of a strategy . . . 76
5.4 The value function and its evolution . . . 78
5.4.1 An individual agent in a soup of agents . . . 78
5.4.2 The standard Hamilton-Jacobi control theory setup . . . 78
5.4.3 The best individual strategy for a given common strategy . . . 79
5.4.4 The Nash equilibrium strategy . . . 80
5.5 Balanced growth paths: a candidate for the long time limit . . . 80
5.6 Basic properties of the solutions to the Lucas-Moll system . . . 81
5.6.1 Monotonicity of the solutions . . . 83
5.6.2 Bounds to the right of the learning front . . . 83
5.6.3 Tightness of the learning front . . . 83
5.6.4 The intrinsic learning front location . . . 84
5.6.5 The median front relative to the learning front . . . 85
5.7 The state lottery society . . . 86
5.8 The effective learning regime . . . 88

1 The overall plan

The preliminary overall plan for these lectures is as follows.

Lecture 1. A basic introduction to the Fisher-KPP equation and its connection to branching Brownian motion. Voting models on the BBM genealogical trees and other semilinear parabolic partial differential equations.

Lecture 2. Some properties of the multi-dimensional branching Brownian motion. This lecture uses the basic definitions from Lecture 1.

Lecture 3. Bounds on the solutions to the thermo-diffusive system of reaction-diffusion equations. This lecture relies on Lecture 1 but only as a motivation.

Lecture 4. A very narrow PDE perspective on a toy version of diffusion models in machine learning. This lecture is independent from the first three lectures.

A note on references: we do not attempt to be exhaustive in our references, please consult the original papers we cite for further references. We apologize in advance for the omissions.

2 Lecture 1: BBM and semi-linear parabolic equations

2.1 Overview of the lecture

In this lecture, we will discuss representations of the solutions to semilinear parabolic equations

$$u_t = \Delta u + f(u), \tag{2.1.1}$$

in terms of the branching Brownian motion. Such representation was first discovered by Skorokhod and, independently, by Ikeda and Watanabe in the 1960s, and later rediscovered by McKean in 1975. All these works arrived at a special class of nonlinearities $f(u)$, that is now sometimes referred to as the McKean nonlinearities. In particular, they belong to the wider class of the Fisher-KPP

nonlinearities $f(u)$ that goes back to the original papers by Fisher [47] and Kolmogorov, Petrovskii and Piskunov [72] that both appeared in 1937.

McKean's representation for $u(t,x)$ uses only the positions $X_1(t),\dots,X_{N_t}(t)$ of the BBM particles present at the time t. Much more recently, Etheridge, Freeman and Penington [44, 45] presented a new and surprising interpretation of the solutions to (2.1.1) with the Allen-Cahn nonlinearity

$$f(u) = u(1-u)(u-1/2),$$

that is neither of the McKean nor of the Fisher-KPP class, in terms of the ternary BBM. Their representation uses both the positions $X_1(t),\dots,X_{N_t}(t)$ and also a voting model on the Galton-Watson tree of the BBM. We discuss in this lecture these representations and the extensions of the Etheridge-Freeman-Penington voting model to a large class of nonlinearities $f(u)$ in (2.1.1). We also present some explicit voting models for some "basic" nonlinearities, such as $f(u) = u^m - u^n$.

The material is mostly based on the presentation in [5] where an interested reader can find more details and examples, as well as some of the omitted proofs.

2.2 The definition of a branching Brownian motion

Models involving branching particles appear very naturally in the context of biological invasions in ecology, as well as in SIR-type models of epidemics. A simple and common process of this type is the binary branching Brownian motion. It is described as follows. A single particle starts at a position $x \in \mathbb{R}^d$ at $t = 0$ and performs a standard Brownian motion. The particle carries an exponential clock that runs at a rate $\beta > 0$, which simply means that the clock rings at a random time $\tau > 0$, with

$$\mathbb{P}(\tau > t) = e^{-\beta t}. \tag{2.2.1}$$

We will often, but not always, assume that the branching rate $\beta = 1$. At the time τ, when the clock rings, the particle splits into two particles that we will refer to as the children. The original particle is sometimes called the parent. The original particle is removed at the branching event. The two children perform independent standard Brownian motions for $t > \tau$, both of them starting at the position of the branching event. Each of the children carries its own exponential clock, and when the corresponding clock rings, the particle splits into two, and the process continues. Thus, at each time $t > 0$ we have a collection of particles $x_1(t),\dots,x_{N(t)}(t)$. Here, $N(t)$ is the random number of particles present at the time t.

The above model is usually referred to as a binary Brownian motion since the number of children is limited to two. A simple modification is a process where the particles may produce a random number k of children at each birth event, with the corresponding probabilities p_k, such that

$$\sum_{k=1}^{\infty} p_k = 1. \tag{2.2.2}$$

Then, the average number of off-spring is

$$\overline{N} = \sum_{k=1}^{\infty} k p_k. \tag{2.2.3}$$

We will usually assume that only finitely many p_k are not zero, for simplicity, though this is by no means necessary. In particular, some interesting effects happen if p_k decay slowly as $k \to +\infty$.

To summarize, a branching Brownian motion is characterized by the exponential clock rate $\beta > 0$, the probabilities p_k to have k children at each branching event, and, finally, the diffusivity $\sigma > 0$ at which the Brownian motions run.

Remark 2.2.1. Let us comment that the exponential form of the clocks in (2.2.1) is, in principle, not required for the results described in these note to hold. However, it is extraordinarily convenient, as can be seen by the difficulty of the results for branching random walks compared to the branching Brownian motion.

Remark 2.2.2. All Brownian motions will always have diffusivity $\sigma = \sqrt{2}$ so that the coefficient in front of the Laplacian in the corresponding parabolic PDE always equals to 1.

2.3 The total number of particles of a BBM

2.3.1 The average total number of particles

The total number $N(t)$ of particles present at the time $t > 0$ can be thought of as a pure birth process – $N(t)$ can go up but not down. As a warm-up, let us prove the following, using a method that will be very useful later on.

Proposition 2.3.1. *Suppose that $\beta = 1$ and let $N(t)$ be the number of particles present in the binary BBM with the exponential clock as in (2.2.1), then*

$$\mathbb{E}(N(t)) = \exp((\overline{N} - 1)t). \tag{2.3.1}$$

Proof. Let us fix some time $t > 0$ and let τ_1 be the first time the exponential clock rings. Then, we can write the following renewal relation for the expected total number of particles

$$\mathbb{E}(N(t)) = 1 \cdot \mathbb{P}(\tau_1 > t) + \sum_{k=2}^{\infty} k p_k \int_0^t \mathbb{E}(N(t-s))\mathbb{P}(\tau_1 \in ds) = e^{-t} + \overline{N} \int_0^t \mathbb{E}(N(t-s))e^{-s} ds. \tag{2.3.2}$$

Here, the first term in the right side accounts for the case when there is no branching until the time t, and the second for the event when the first branching happens in an interval $[s, s + ds]$ with some $0 < s < t$ and that k children are produced at the first branching. This recursive relation relies crucially on the independence of the off-spring of a given parent. It follows from (2.3.2) that the function

$$u(t) = \mathbb{E}(N(t)), \tag{2.3.3}$$

satisfies an integral equation

$$u(t) = e^{-t} + \overline{N} \int_0^t u(t-s)e^{-s} ds = e^{-t} + \overline{N} \int_0^t e^{-(t-s)} u(s) ds. \tag{2.3.4}$$

This, of course, is the Duhamel formula for the ODE

$$\frac{du}{dt} + u = \overline{N} u, \quad u(0) = 1. \tag{2.3.5}$$

Now, (2.3.1) follows. □

Equation (2.3.5) is maybe the simplest example of an interpretation of the solutions to a differential equation in terms of a branching Brownian motion. Observe that if our starting point is the ODE (2.3.5), with a given $\overline{N} > 0$, then we can choose any BBM to represent $u(t)$ by (2.3.3), as long as we choose the probabilities p_k so that (2.2.3) holds. This non-uniqueness of a stochastic model behind a deterministic differential equation is a generic phenomenon – there are often many representations for the solutions and it is not always clear which one is the "best" or "natural".

2.3.2 A simplest martingale

We can actually say more about the total number of particles. Very naively, since the number of particles is very large, one may think that, in the spirit of the law of large numbers, $N(t)$ behaves as $N_\infty e^t$, as $t \to +\infty$, where N_∞ would be a fixed deterministic constant.

Exercise 2.3.2. Show this is impossible because, for example, the time of the first branching event is random and the above guess would indicate that $N(t)$ should also behave as $2N_\infty e^{t-\tau_1}$, which would be a contradiction.

However, the guess is not very far from the correct picture, except that N_∞ is random.

Proposition 2.3.3. *Let $N(t)$ be the number of particles in the binary BBM, then*

$$M(t) = e^{-t}N(t) \tag{2.3.6}$$

is a martingale. Moreover, $M(t)$ converges, as $t \to +\infty$, almost surely and in L^1 to a random variable M_∞.

Proof. First, note that $M(t)$ is a positive random variable with $E(M(t)) = 1$. Hence, it is integrable. To check that $M(t)$ is a martingale, we note that for $t > s$ we have

$$\mathbb{E}(M(t)|\mathcal{F}_s) = \mathbb{E}(e^{-t}N(t)|\mathcal{F}_s) = e^{-t}\mathbb{E}(N(t)|\mathcal{F}_s). \tag{2.3.7}$$

Recall that the exponential clocks have the "lack of memory" property:

$$\mathbb{P}(\tau > s + t|\tau > t) = \mathbb{P}(\tau > s). \tag{2.3.8}$$

This is because

$$\begin{aligned}\mathbb{P}(\tau > s + t|\tau > t) &= \frac{\mathbb{P}(\tau > s + t \text{ and } \tau > t)}{\mathbb{P}(\tau > t)} = \frac{\mathbb{P}(\tau > s + t)}{\mathbb{P}(\tau > t)} = \frac{e^{-(t+s)}}{e^{-t}} \\ &= e^{-s} = \mathbb{P}(\tau > s).\end{aligned} \tag{2.3.9}$$

Note that the lack of memory property is very specific to the exponential clocks, and that is one reason why using an exponential clock is convenient.

The lack of memory property means that we can reset all the clocks at the time s to zero, without changing the law of the process for $t > s$, and deduce from Proposition 2.3.1 that

$$\mathbb{E}(N(t)|\mathcal{F}_s) = N(s)e^{(t-s)}. \tag{2.3.10}$$

Using this in (2.3.7) gives

$$\mathbb{E}(M(t)|\mathcal{F}_s) = e^{-t}\mathbb{E}(N(t)|\mathcal{F}_s) = e^{-t}N(s)e^{(t-s)} = e^{-s}N(s) = M(s). \tag{2.3.11}$$

Thus, $M(t)$ is a martingale. Doob's theorem implies that, as $M(t)$ is a martingale, it converges almost surely, as $t \to +\infty$ to a random limit M_∞. To prove convergence in L^1, it suffices to show that $M(t)$ is uniformly integrable. For that, we can look at

$$\mathbb{E}(M^2(t)) = e^{-2t}\mathbb{E}(N^2(t)). \tag{2.3.12}$$

We have the following recursion relation for the function $\phi_2(t) = \mathbb{E}(N^2(t))$, analogous to (2.3.2):

$$\begin{aligned}\phi_2(t) &= \mathbb{P}(\tau_1 < t) + \int_0^t \mathbb{E}(N_1(t-s) + N_2(t-s))^2 \mathbb{P}(\tau_1 \in ds)\mathbb{P}(\tau_1 < t) \\ &= \mathbb{P}(\tau_1 < t) + \int_0^t (2\phi_2(t-s) + 2e^{2(t-s)})\mathbb{P}(\tau_1 \in ds) \\ &= e^{-t} + 2\int_0^t (\phi_2(t-s) + e^{2(t-s)})e^{-s}ds.\end{aligned} \tag{2.3.13}$$

Here, $N_1(s)$ and $N_2(s)$ are to independent copies of the process $N(s)$ starting at $N(0) = 1$. We used above (2.3.1) with $\overline{N} = 2$. Hence, the function

$$\psi_2(t) = \phi_2(t) + e^{2t} \tag{2.3.14}$$

satisfies

$$\psi_2(t) = e^{-t} + e^{2t} + 2\int_0^t \psi_2(t-s)e^{-s}ds. \tag{2.3.15}$$

Differentiating in t gives

$$\begin{aligned}\frac{d\psi_2(t)}{dt} &= -e^{-t} + 2e^{2t} + 2e^{-t}\psi_2(0) + 2\int_0^t \frac{d\psi_2(t-s)}{dt}e^{-s}ds \\ &= 3e^{-t} + 2e^{2t} - 2\int_0^t e^{-s}\frac{d\psi_2(t-s)}{ds}ds = 3e^{-t} + 2e^{2t} - 2e^{-t}\psi_2(0) + 2\psi_2(t) \\ &- 2\int_0^t e^{-s}\psi_2(t-s).\end{aligned} \tag{2.3.16}$$

Inserting (2.3.15) into (2.3.16) and using the fact that $\psi_2(0) = 2$, gives

$$\frac{d\psi_2(t)}{dt} = 2e^{2t} - e^{-t} + 2\psi_2(t) - (\psi_2(t) - e^{-t} - e^{2t}) = 3e^{2t} + \psi_2(t). \tag{2.3.17}$$

It follows that

$$\psi_2(t) = 3e^{2t} - e^t, \tag{2.3.18}$$

and

$$\phi_2(t) = 2e^{2t} - e^t. \tag{2.3.19}$$

This gives

$$\gamma_2(t) := \mathbb{E}(M^2(t)) = e^{-2t}\phi_2(t) = 2 - e^{-t}. \tag{2.3.20}$$

Thus, $M(t)$ is uniformly integrable, and converges to M_∞ in L^1, as well. □

The martingale $N(t)$ is not particularly important in itself but other related martingales will play an important role in the stories below. It will also become more clear why it is a martingale when we revisit this issue.

2.4 Linear parabolic equations and branching Brownian motion

Let us next explain how we can obtain a probabilistic interpretation for a parabolic equation with a constant zero-order term:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + mu, \tag{2.4.1}$$

with some $m \in \mathbb{R}$ fixed.

Let us assume that the BBM starts at $t = 0$ at the position x and denote the locations of the BBM particles at the time $t > 0$ by $X_1(t), \dots, X_{N_t}(t)$. Here and below, unless specified otherwise, we will assume that the BBM exponential clock rate $\beta = 1$. Given a bounded function $g(x)$, consider

$$u(t,x) = \mathbb{E}_x \sum_{k=1}^{N_t} g(X_k(t)). \tag{2.4.2}$$

For example, to get the expected number of particles, as in (2.3.3), we simply set $g(x) \equiv 1$.

In order to get an equation for $u(t,x)$ let us write a renewal relation, using the independence of the off-spring particles. Looking at the first branching event gives the identity

$$\begin{aligned} u(t,x) &= \mathbb{E}_x(g(B_t))\mathbb{P}(\tau_1 > t) + \sum_{k=1}^{\infty} kp_k \int_0^t \mathbb{E}_x(u(t-s,B_s))\mathbb{P}(\tau_1 \in ds) \\ &= \mathbb{E}_x(g(B_t))e^{-t} + \sum_{k=1}^{\infty} kp_k \int_0^t \mathbb{E}_x(u(t-s,B_s))e^{-s}ds. \end{aligned} \tag{2.4.3}$$

Here, as usual, the notation $\mathbb{E}_x$ means that the Brownian motion B_s starts at the position $x \in \mathbb{R}^n$ at the time $t = 0$. Once again, the first term in the right side accounts for the event when there was no branching until the time $t > 0$, and the second for the event that the first branching happened in a time interval $[s, s + ds]$ with some $0 < s < t$. Note that the function

$$v(t,x) = \mathbb{E}_x(g(B_t))$$

is the solution to the heat equation (recall that the diffusivity of the Brownian motion is $\sqrt{2}$)

$$\frac{\partial v}{\partial t} = \Delta v, \tag{2.4.4}$$

with the initial condition $v(0,x) = g(x)$. Hence, it can be written as

$$v(t,x) = [e^{t\Delta} g(\cdot)](x). \tag{2.4.5}$$

In addition, for $0 < s < t$ fixed, the function

$$w(\tau,x) = \mathbb{E}_x(u(t-s,B_\tau)) \tag{2.4.6}$$

is the solution to the heat equation

$$\frac{\partial w}{\partial \tau} = \Delta w, \tag{2.4.7}$$

with the initial condition $w(0,x) = u(t-s,x)$. That is, we have

$$w(\tau,x) = [e^{\tau\Delta} u(t-s,\cdot)](x). \tag{2.4.8}$$

It follows that

$$\mathbb{E}_x(u(t-s,B_s)) = w(s,x) = [e^{s\Delta} u(t-s,\cdot)](x). \tag{2.4.9}$$

Hence, (2.4.3) has the form

$$u(t,x) = e^{-t}[e^{t\Delta} g(\cdot)](x) + \overline{N} \int_0^t [e^{s\Delta} u(t-s,\cdot)](x) e^{-s} ds. \tag{2.4.10}$$

This is simply the Duhamel formula for the initial value problem

$$\begin{aligned} &\frac{\partial u}{\partial t} = \Delta u + (\overline{N} - 1)u, \\ &u(0,x) = g(x). \end{aligned} \tag{2.4.11}$$

This is exactly (2.4.1), with $m = \overline{N} - 1$.

2.5 McKean's interpretation of the Fisher-KPP equation

The Fisher-KPP equation

$$\frac{\partial u}{\partial t} = \Delta u + u - u^2 \tag{2.5.1}$$

was introduced in the classical papers by Fisher [47] and Kolmogorov, Petrovskii and Piskunov [72] as a very basic PDE model of spreading in 1937. It was extensively studied and used, without any connection to the probability theory, in many applications where spreading is relevant, from biology to flame propagation. Then, a direct link between the binary branching Brownian motion and the Fisher-KPP equation was discovered independently by Skorohod in 1964, Watanabe in 1968, and McKean in 1975, that we now describe.

Given a bounded function $g(x)$, we consider a multiplicative rather than additive functional of the branching Brownian motion. That is, we replace (2.4.2) by

$$v(t,x) = \mathbb{E}_x\Big(\prod_{k=1}^{N_t} g(X_k(t))\Big). \tag{2.5.2}$$

In order to get an equation for $v(t,x)$ let us again write a renewal relation, very similar to what we have seen in (2.4.3). Because of the product structure of (2.5.2) and since the children at each branching event behave independently, looking at the first branching event gives

$$\begin{aligned} v(t,x) &= \mathbb{E}_x(g(B_t))\mathbb{P}(\tau_1 > t) + \sum_{k=1}^{\infty} p_k \int_0^t \mathbb{E}_x\Big(\big[v(t-s,B_s)\big]^k\Big)\mathbb{P}(\tau_1 \in ds) \\ &= \mathbb{E}_x(g(B_t))e^{-t} + \sum_{k=1}^{\infty} p_k \int_0^t \mathbb{E}_x\Big(\big[v(t-s,B_s)\big]^k\Big)e^{-s}ds. \end{aligned} \tag{2.5.3}$$

We now recall (2.4.5):

$$\mathbb{E}_x(g(B_t)) = [e^{t\Delta}g(\cdot)](x), \tag{2.5.4}$$

as well as (2.4.9), written as

$$\mathbb{E}_x\big[v(t-s,B_s)\big]^k = [e^{s\Delta}v^k(t-s,\cdot)](x). \tag{2.5.5}$$

Now, (2.5.3) becomes

$$\begin{aligned} v(t,x) &= [e^{t\Delta}g(\cdot)](x)e^{-t} + \sum_{k=1}^{\infty} p_k \int_0^t [e^{s\Delta}v^k(t-s,\cdot)](x)e^{-s}ds \\ &= [e^{t\Delta}g(\cdot)](x)e^{-t} + \int_0^t [e^{s\Delta}F(v(t-s,\cdot))](x)e^{-s}ds. \end{aligned} \tag{2.5.6}$$

This is the Duhamel representation for the initial value problem

$$\begin{aligned} &\frac{\partial v}{\partial t} = \Delta v - v + F(v), \\ &v(0,x) = g(x). \end{aligned} \tag{2.5.7}$$

Here, the nonlinearity $F(v)$ is given by the generating function for the branching process:

$$F(v) = \sum_{k=1}^{\infty} p_k v^k. \tag{2.5.8}$$

From the PDE point of view, it is often convenient to use instead the function

$$u(t,x) = 1 - v(t,x) = 1 - \mathbb{E}_x\Big(\prod_{k=1}^{N_t} g(X_k(t))\Big). \tag{2.5.9}$$

It satisfies the initial value problem

$$\begin{aligned} &\frac{\partial u}{\partial t} = \Delta u + f(u), \\ &u(0,x) = 1 - g(x). \end{aligned} \tag{2.5.10}$$

Here, we have defined

$$f(u) = 1 - u - F(1-u) = 1 - u - \sum_{k=1}^{\infty} p_k (1-u)^k. \tag{2.5.11}$$

In the case of the purely binary branching, when $p_2 = 0$ and all other $p_k = 0$, the function $f(u)$ takes the form

$$f(u) = 1 - u - (1-u)^2 = u(1-u).$$

Then, (2.5.10) becomes the classical Fisher-KPP equation (2.5.1):

$$\begin{aligned} &\frac{\partial u}{\partial t} = \Delta u + u - u^2, \\ &u(0,x) = 1 - g(x). \end{aligned} \tag{2.5.12}$$

2.6 Not all Fisher-KPP nonlinearities are of the McKean type

To understand what kind of parabolic equations are related to the branching Brownian motion by the McKean probabilistic interpretation, let us mention some very basic general properties of the nonlinearities $f(u)$ of the form (2.5.11):

$$f(u) = 1 - u - \sum_{k=1}^{\infty} p_k (1-u)^k. \tag{2.6.1}$$

First, because

$$\sum_{k=1}^{\infty} p_k = 1, \tag{2.6.2}$$

we have

$$f(0) = f(1) = 0, \quad f(u) > 0 \text{ for } 0 < u < 1. \tag{2.6.3}$$

Second, $f(u)$ is concave, so that, in particular it satisfies the so-called Fisher-KPP condition

$$f(u) \le f'(0)u, \quad \text{for all } u \in (0,1). \tag{2.6.4}$$

This property of $f(u)$ will be very important in the discussion of the long time behavior of the solutions to (2.5.10).

The above conditions (2.6.3) and (2.6.4) are, clearly, not sufficient for a nonlinearity $f(u)$ to be of the BBM origin but give a good idea of the special properties of nonlinearities in that class.

As we have noted, the McKean nonlinearities belong to the class of Fisher-KPP nonlinearities, in the sense that they all satisfy (2.6.3)-(2.6.4). On the other hand, it is also well known that solutions to parabolic equations (2.5.10) with an FKPP type $f(u)$ enjoy many special properties that we will discuss later in these notes. However, while the McKean nonlinearities lie in the FKPP class, they form a very special sub-class of that set which excludes many natural examples.

To explain the above point, consider the following exercise.

Exercise 2.6.1. Let $f(u)$ be of the form (2.6.1). Then, it can be written in the form

$$f(u) = \lambda(u - A(u)) = \lambda u(1 - \alpha(u)), \tag{2.6.5}$$

with a function $A(u)$ that is increasing and convex on $[0, 1]$, with

$$A(0) = 0, \qquad A(1) = 1, \qquad A'(0) = 0. \tag{2.6.6}$$

while $\alpha(u) = A(u)/u$ is increasing and concave on $[0, 1]$ and such that $\alpha(0) = 0$ and $\alpha(1) = 1$.

The classical Fisher-KPP equation (2.5.1) has $f(u) = u - u^2$ which is in the McKean class, as can be seen by setting $p_2 = 1$ and $p_k = 0$ for $k > 2$. This corresponds to $\lambda = 1$ and $\alpha(u) = u$. However, even the original example

$$f(u) = u(1 - u)^2 = u - 2u^2 + u^3 \tag{2.6.7}$$

in the KPP paper [72] is not of the McKean form because the function

$$A(u) = 2u^2 - u^3 \tag{2.6.8}$$

is not convex for $u \in (2/3, 1)$. Or, consider the functions $f_n(u) = u - u^n$ that are of the Fisher-KPP type but the corresponding functions $\alpha_n(u) = u^{n-1}$ are convex and not concave. This means that they are also not of the McKean type. In short, Skorohod-Watanabe-McKean's connection between semi-linear parabolic equations and branching Brownian motion, while incredibly elegant, does not cover all polynomial Fisher-KPP nonlinearities.

2.7 Voting schemes for semi-linear equations

We now describe a different probabilistic interpretation that covers all semilinear parabolic equations with polynomial nonlinearities, via a branching Brownian motion. The presentation is based on [5]. We should also mention a very remarkable (undergraduate!) thesis of O'Dowd [100] where many of these results were also obtained, in a slightly different form. The approach of [5] is very much inspired by the arguments in the papers by Etherdige, Freeman and Penington [44, 45] for the Allen-Cahn nonlinearity. We first recall that connection in Section 2.7.1. Then, we discuss the random outcome and random threshold voting models that allow us to give a probabilistic interpretation to solutions to equations of the form

$$u_t = \Delta u + f(u), \tag{2.7.1}$$

with any polynomial nonlinearity $f(u)$ such that

$$f(0) = f(1) = 0. \tag{2.7.2}$$

This is done in Sections 2.7.2 and 2.7.5, with the main results given in Theorems 2.7.3 and 2.7.4. Section 2.7.2 also describes an interpretation of the standard heat equation, in terms of a BBM and an unbiased voting model, formulated in Proposition 2.7.2. Finally, in Section 2.7.6 we drop the assumption (2.7.2) and describe a recursive procedure on the genealogical tree that gives a BBM-interpretation for the solutions to any equation of the form (2.7.1) with a polynomial nonlinearity $f(u)$. The result is described in Theorem 2.7.5.

2.7.1 The Etheridge-Freeman-Penington model for the Allen-Cahn equation

An alternative connection between semilinear parabolic equations and branching Brownian motion to McKean's was pointed out in a beautiful paper by Etheridge, Freeman, and Penington [44] (see also [41, 73, 97]). One of the main points of the approach of [44, 45] is to consider functionals that depend not just on the locations of the BBM particles, as was done by McKean in [89], but also on the structure of the (random) genealogical tree that results from the branching. These ideas also go back to Sznitman [109]. Maybe the main observation is that for semi-linear equations Sznitman's ideas have a very simple form.

There is a natural way to associate a random genealogical tree $\mathcal{T}(t)$ to each realization of the BBM running on a time interval $0 \leq s \leq t$. Each vertex of the tree corresponds to a branching event, while each of the edges coming out of a vertex represents an offspring particle born at that branching event. The root of the tree $\mathcal{T}(t)$ represents the original particle that started at the time $s = 0$ at a position $x \in \mathbb{R}^d$. We refer to [44, 45] for a formal definition of $\mathcal{T}(t)$.

Before introducing a generalization of their ideas, let us recall the example of [44]. Consider a ternary branching Brownian motion starting at the time $t = 0$ at a point $x \in \mathbb{R}^d$ – each branching event produces three children, with probability one. The process is run until a time $t > 0$, with the BBM particles at the time $t > 0$ located at the positions $X_1(t), \ldots, X_{N_t}(t)$. Then, each of the youngest generation particles $X_j(t)$, $j = 1, \ldots, N_t$, "votes" 0 or 1, with the probabilities

$$\mathbb{P}(V_j = 1) = g(X_j(t)) \quad \text{and} \quad \mathbb{P}(V_j = 0) = 1 - g(X_j(t)). \tag{2.7.3}$$

Here, $g(x)$ is a prescribed function such that $0 \leq g(x) \leq 1$ for all $x \in \mathbb{R}$, and V_j is the vote of the particle $X_j(t)$, $j = 1, \ldots, N_t$. This produces the votes of the youngest generation of particles. Next, we go back up the ternary branching tree $\mathcal{T}(t)$, with the rule that each parent accepts the vote of the majority of its three children. In this way, we obtain the votes of all particles on the genealogical tree.

Let V_{orig} be the resulting vote of the original ancestral particle that started at $t = 0$ at the position x, and consider the function

$$u(t, x) = \mathbb{P}_x(V_{\text{orig}} = 1). \tag{2.7.4}$$

We now derive an equation for $u(t, x)$ using a similar approach to (2.5.3). There are exactly two possible ways in which the original ancestor can vote 1: either all three of its children voted 1 or two of them voted 1 and one voted 0. In the latter case, there are three choices of the particle that voted 0. If there has been no branching before the time t then the only particle present is the original ancestor, and it takes the vote by itself. These simple observations give the renewal identity

$$\begin{aligned} u(t, x) &= \mathbb{E}_x\big[g(B_t)\big]\mathbb{P}(\tau_1 > t) \\ &\quad + \int_0^t \mathbb{E}_x(u^3(t-s, B_s) + 3u^2(t-s, B_s)(1 - u(t-s, B_s))\mathbb{P}(\tau_1 \in ds) \\ &= \mathbb{E}_x[g(B_t)]e^{-\beta t} + \beta \int_0^t \mathbb{E}_x(u^3(t-s, B_s) + 3u^2(t-s, B_s)(1 - u(t-s, B_s))e^{-\beta s} ds \\ &= e^{(\Delta - \beta)t} g(x) + \beta \int_0^t e^{(\Delta-\beta)s}\big[u^3(t-s, \cdot) + 3u^2(1-u)(t-s, \cdot)\big](x) ds. \end{aligned} \tag{2.7.5}$$

The corresponding semi-linear parabolic equation is

$$\frac{\partial u}{\partial t} = \Delta u + \beta(u^3 + 3u^2(1-u) - u). \tag{2.7.6}$$

A simple computation shows that, miraculously, this simplifies to

$$u^3 + 3u^2(1-u) - u = 3u^2 - 2u^3 - u = u(3u - 2u^2 - 1) = u(1-u)(2u-1). \tag{2.7.7}$$

Thus, the function $u(t,x)$ defined by (2.7.4) satisfies the Allen-Cahn equation

$$\begin{aligned} &\frac{\partial u}{\partial t} = \Delta u + \beta u(1-u)(2u-1), \\ &u(0,x) = g(x). \end{aligned} \tag{2.7.8}$$

This equation is probably the most standard example of a semi-linear parabolic equation that does not have a McKean connection to BBM and, prior to [44], was believed to have no probabilistic interpretation, to the best of our knowledge. Note that the nonlinearity

$$f(u) = u(1-u)(2u-1) \tag{2.7.9}$$

does not satisfy the Fisher-KPP properties we have discussed in Section 2.6. Indeed, it is not even non-negative for $u \in (0,1)$ but rather changes its sign. Thus, the Allen-Cahn equation (2.7.8) does not have an interpretation in terms of a McKean functional. The voting scheme idea of [44] adds a genuinely new aspect here and dramatically broadens the class of equations that have an interpretation in terms of the BBM.

Let us also mention another, more geometric,interpretation of the above voting rule. Assume, for simplicity that the initial condition $g(x) = \mathbb{1}(x \le 0)$ is a step function at the origin. Then, we have $V_{orig} = 1$ if and only if one can find a sub-collection $\{Y_k(t)\}$ of the youngest generation of particles, $X_j(t)$, $j = 1, \dots, N_t$, that forms a binary sub-tree of the full ternary tree $\mathcal{T}(t)$, and such that all $Y_k(t) \le 0$. This is further discussed in [79].

2.7.2 Random outcome probabilistic voting models

The voting procedure of [44] that we have described above is deterministic, in the sense that once the genealogical tree $\mathcal{T}(t)$ and the votes of the youngest generation particles $X_1(t), \dots, X_{N_t}(t)$ are fixed, the vote V_{orig} of the original particle is completely determined. However, one can also randomize the voting process itself, in at least two clear ways that we now discuss. We consider a general BBM, with the probabilities p_k to produce k offspring particles at each branching event, with $2 \le k \le N$. As before, we denote by $\mathcal{T}(t)$ the genealogical tree produced by branching events on the time interval $0 \le s \le t$, and by p_k the probability that a parent produces exactly k children at a given branching event.

Let us fix a continuous function $g(x)$ such that $0 \le g(x) \le 1$ for all $x \in \mathbb{R}^d$ and run a branching Brownian motion starting at a position $x \in \mathbb{R}^d$ until a time $t > 0$. At the time t, each of the BBM particles $X_1(t), \dots, X_{N_t}(t)$ votes randomly 0 or 1, with the probability to vote 1 given by

$$\mathbb{P}(\text{Vote}(X_k(t)) = 1) = g(X_k(t)), \quad \text{for each } 1 \le k \le N_t. \tag{2.7.10}$$

In a difference with [44], we also fix a collection of probabilities $0 \le \alpha_{kn} \le 1$, with $0 \le k \le n$, and $n \ge 2$, such that

$$\alpha_{0n} = 0, \quad \alpha_{nn} = 1, \quad \text{for all } n \ge 2. \tag{2.7.11}$$

Given the votes of the particles that are present at the time t, we propagate the vote up the genealogical tree $\mathcal{T}(t)$ as follows. If a parent particle has n children and k out of its n children voted 1, then the parent particle votes 1 with the probability α_{kn}. That is, the vote of the parent is no longer a deterministic function of the votes of its children. Using this rule to go up the tree

all the way to the root produces the vote $\text{Vote}_{\text{orig}}$ of the original ancestor particle, and we can, as before, define

$$u(t,x) = \mathbb{P}_x(\text{Vote}_{\text{orig}} = 1). \tag{2.7.12}$$

If there was no branching event until the time t, so that $N_t = 1$, then the vote of the original particle is 1 with the probability $g(X_1(t))$. Note that while we allow α_{kn} to be different from 0 and 1, we do impose (2.7.11) which says that if all children voted unanimously, then the parent accepts the vote of the children. We refer to the above as a random outcome voting model.

Similarly to (2.7.5), one can write a renewal equation for the function $u(t,x)$ defined in (2.7.12):

$$\begin{aligned}
u(t,x) &= \mathbb{E}(g(x+B_t))\mathbb{P}(\tau > t) \\
&\quad + \int_0^t \mathbb{E}\Big(\sum_{n=2}^N p_n \sum_{k=0}^n \binom{n}{k} \alpha_{kn} u^k(t-s, x+B_s)(1-u(t-s,x+B_s))^{n-k}\Big)\mathbb{P}(\tau \in ds) \\
&= \mathbb{E}(g(x+B_t))e^{-\beta t} \\
&\quad + \int_0^t \mathbb{E}\Big(\sum_{n=2}^N p_n \sum_{k=0}^n \binom{n}{k} \alpha_{kn} u^k(t-s, x+B_s)(1-u(t-s,x+B_s))^{n-k}\Big)\beta e^{-\beta s} ds \\
&= e^{(\Delta-\beta)t} g(x) + \beta \int_0^t \Big(\sum_{n=2}^N p_n \sum_{k=0}^n \binom{n}{k} \alpha_{kn} e^{(\Delta-\beta)s} u^k(t-s,\cdot)(1-u(t-s,\cdot))^{n-k}(x)\Big) ds.
\end{aligned} \tag{2.7.13}$$

The first term on the right in (2.7.13) comes from the event that there was no branching until the time t. The second accounts for the first branching event happening at a time $t \in [s, s+ds]$, with $0 < s < t$. The binomial coefficient counts the number of possibilities to choose the k children who voted 1 out of the n children. Note that (2.7.13) is the Duhamel formulation of the initial value problem

$$\begin{aligned}
u_t &= \Delta u + f(u), \\
u(0,x) &= g(x),
\end{aligned} \tag{2.7.14}$$

with the nonlinearity

$$f(u) = \beta \sum_{n=2}^N p_n \sum_{k=0}^n \binom{n}{k} \alpha_{kn} u^k (1-u)^{n-k} - \beta u = \beta \sum_{n=2}^N p_n \Big(\sum_{k=0}^n \binom{n}{k} \alpha_{kn} u^k (1-u)^{n-k} - u\Big). \tag{2.7.15}$$

As we have seen in the Etheridge-Freeman-Penington example, unlike in the McKean interpretation, the nonlinearities produced in this way need not be of the Fisher-KPP type. The advantage of the voting models is precisely in providing a probabilistic interpretation for nonlinear parabolic equations not accessible by the McKean formula.

2.7.3 The standard heat equation and unbiased voting

We now consider some concrete examples of parabolic equations coming from probabilistic voting models, starting with the standard heat equation. Let us first note an elementary identity.

Exercise 2.7.1. Show that any $n \geq 1$ and $0 \leq u \leq 1$ we have

$$\sum_{k=0}^n \binom{n}{k} \frac{k}{n} u^k (1-u)^{n-k} = u. \tag{2.7.16}$$

Using (2.7.16) in the representation (2.7.15) for $f(u) = 0$, we see that taking the probabilities

$$\alpha_{kn} = \frac{k}{n} \tag{2.7.17}$$

in the above voting scheme leads to the standard heat equation: (2.7.14) becomes

$$\begin{aligned} u_t &= \Delta u, \\ u(0,x) &= g(x). \end{aligned} \tag{2.7.18}$$

That is, consider any branching Brownian motion, regardless of the branching probabilities p_k, and introduce the voting scheme such that a parent with n children, out of which k voted 1, votes 1 with the "unbiased" probability $\alpha_{kn} = k/n$. Then, the function $u(t,x)$, the probability that the original ancestor particle votes 1, is the solution to the standard heat equation. Let us summarize this result as follows.

Proposition 2.7.2. *Let $g(x)$ be a continuous function that satisfies $0 \le g(x) \le 1$ for all $x \in \mathbb{R}^d$. Consider the random outcome voting model with the voting probabilities $\alpha_{kn} = k/n$, $0 \le k \le n$, for any branching Brownian motion. Then, the function $u(t,x) = \mathbb{P}_x(V_{\text{orig}} = 1)$ is the solution to the initial value problem*

$$\begin{aligned} &u_t = \Delta u, \quad t > 0, \ x \in \mathbb{R}^d, \\ &u(0,x) = g(x), \quad x \in \mathbb{R}^d. \end{aligned} \tag{2.7.19}$$

2.7.4 Representing general nonlinearities

Let now $f(u)$ be a polynomial of degree N:

$$f(u) = \sum_{k=0}^{N} f_k u^k, \tag{2.7.20}$$

that vanishes at $u = 0$ and $u = 1$:

$$f(0) = f(1) = 0. \tag{2.7.21}$$

Our goal is to find $\beta > 0$ and α_{kN}, $0 \le k \le N$ such that

$$\alpha_{0N} = 0, \quad \alpha_{NN} = 1, \quad 0 \le \alpha_{kN} \le 1, \quad \text{for all } 1 \le k \le N-1, \tag{2.7.22}$$

and so that representation (2.7.15) holds for $f(u)$. We set all $p_n = 0$ except for $p_N = 1$, so that each branching event produces exactly N offspring. We look for α_{kN} in the form

$$\alpha_{kN} = \frac{k}{N} + \mu_k \quad \text{for any } 1 \le k \le N-1, \tag{2.7.23}$$

with μ_k to be chosen later. Recalling (2.7.15) and (2.7.16), we see that we need to have

$$f(u) = \beta \sum_{k=1}^{N-1} \binom{N}{k} \mu_k u^k (1-u)^{N-k}. \tag{2.7.24}$$

On the other hand, the Bernstein polynomials

$$B_{k,N}(u) = \binom{N}{k} u^k (1-u)^{N-k}, \quad k = 0, \dots, N, \tag{2.7.25}$$

form a basis for the vector space of polynomials of degree at most N. Therefore, $f(u)$ has a representation

$$f(u) = \sum_{k=0}^{N} b_k[f] B_{k,N}(u), \tag{2.7.26}$$

with some coefficients $b_k[f]$. Note that

$$b_0[f] = f(0), \quad b_N[f] = f(1). \tag{2.7.27}$$

We deduce from (2.7.21) and (2.7.27) that

$$b_0[f] = b_N[f] = 0. \tag{2.7.28}$$

Next, comparing (2.7.24) and (2.7.26), we see that for (2.7.24) to hold, we need to have

$$\mu_k = \frac{b_k[f]}{\beta}, \quad \text{for all } 1 \le k \le N-1, \ \ \mu_0 = \mu_N = 0. \tag{2.7.29}$$

It remains to choose $\beta > 0$ so that α_{kN} given by (2.7.23) satisfy (2.7.22). Note that, since we have set $\mu_0 = \mu_N = 0$, we automatically have $\alpha_{0N} = 0$ and $\alpha_{NN} = 1$. The rest of the conditions in (2.7.22) translates into

$$0 \le \frac{k}{N} + \mu_k \le 1, \quad \text{for all } 1 \le k \le N-1. \tag{2.7.30}$$

We see from (2.7.29) that this is equivalent to

$$0 \le \frac{k}{N} + \frac{b_k[f]}{\beta} \le 1, \quad \text{for all } 1 \le k \le N-1. \tag{2.7.31}$$

This condition holds as long as we choose $\beta > 0$ sufficiently large, so that

$$\beta \ge N \max_k |b_k[f]|. \tag{2.7.32}$$

Therefore, we have proved the following.

Theorem 2.7.3. *Let $f(u)$ be a polynomial of degree N such that $f(0) = f(1) = 0$. Then, there exists a random outcome voting model representation in terms of a purely N-ary branching Brownian motion for the solution to the initial value problem (2.7.14) with the initial condition $g(x)$ that is continuous and satisfies $0 \le g(x) \le 1$ for all $x \in \mathbb{R}^d$.*

We note that Theorem 2.7.3 has also been observed in Z. O'Dowd's thesis [100], although with different terminology and presentation. One somewhat important point is that with BBM we have the ability to choose $\beta > 0$ sufficiently large to ensure that the probabilities α_{nk} do not exceed 1. This is not true for a branching random walk.

2.7.5 A random threshold voting model

An alternative probabilistic voting model has been recently suggested in [79]. For simplicity of notation, let us fix the number N of offspring produced at each branching event and consider a purely N-ary BBM. Then, at each branching event we choose a number $L \in \{1, \dots, N\}$, with the probability $\zeta_k = \mathbb{P}(L = k)$ so that

$$\sum_{k=1}^{N} \zeta_k = 1. \tag{2.7.33}$$

Thus, an integer $L(\nu)$ is assigned separately to each vertex ν of the genealogical tree $\mathcal{T}(t)$. The voting is done as follows. As before, at the time t the youngest generation of particles votes according to (2.7.10). The difference is in the way the votes are propagated up the genealogical tree. A parent at a vertex ν votes 1 if and only if at least $L(\nu)$ of its children voted 1. We refer to this process as a random threshold voting model.

The same argument as in (2.7.13) shows that the function

$$u(t,x) = \mathbb{P}_x(\mathrm{Vote}_{\mathrm{orig}} = 1) \tag{2.7.34}$$

is a solution to the initial value problem

$$\begin{aligned} &u_t = \Delta u + G(u), \\ &u(0,x) = g(x), \end{aligned} \tag{2.7.35}$$

with the nonlinearity

$$G(u) = \beta \sum_{j=0}^{N} \zeta_j \sum_{k=j}^{N} \binom{N}{k} u^k (1-u)^{N-k} - \beta u = \beta \sum_{k=0}^{N} \binom{N}{k} u^k (1-u)^{N-k} \sum_{j=0}^{k} \zeta_j - \beta u. \tag{2.7.36}$$

A simple observation is that if we start with the random threshold voting model and set

$$\alpha_{kN} = \sum_{j=0}^{k} \zeta_j, \tag{2.7.37}$$

in (2.7.15), then the nonlinearities $f(u)$ in (2.7.15), coming from the random outcome model with the probabilities α_{kN}, and $G(u)$ in (2.7.36) are the same. Note that (2.7.33) implies that $0 \le \alpha_{kN} \le 1$ and $\alpha_{NN} = 1$, so α_{kN} satisfy the assumptions that we needed in Section 2.7.2.

On the other hand, given a random outcome voting model of Section 2.7.2, with a collection of probabilities α_{kN} that additionally have the property that the probabilities α_{kN} are increasing in k, then we can obtain a random threshold model by setting

$$\zeta_k = \alpha_{kN} - \alpha_{k-1,N}. \tag{2.7.38}$$

Note that

$$\sum_{k=0}^{N} \zeta_k = \alpha_{NN} = 1. \tag{2.7.39}$$

Monotonicity of α_{kN} in k is a natural assumption, as it says that the larger number of children voted 1 the higher the probability that the parent votes 1. Moreover, it is easy to see that given a polynomial nonlinearity $f(u)$ satisfying (2.7.21), we can always find a collection of probabilities α_{kN} that is increasing in k and so that (2.7.15) holds. To see that, let us take a nonlinearity $f(u)$ that is a polynomial of degree N such that $f(0) = f(1) = 0$. The construction of the probabilities α_{kN} in the argument leading to Theorem 2.7.3 produces α_{kN} that are increasing in k as long as the branching rate β satisfies the condition

$$\beta \ge 2N \max_k |b_k[f]|, \tag{2.7.40}$$

that is slightly stronger than (2.7.32). This is because if α_{kN} are given by (2.7.23) and (2.7.29), then

$$\begin{aligned} \alpha_{k+1,N} &= \frac{k+1}{N} + \mu_k = \frac{k+1}{N} + \frac{b_{k+1}[f]}{\beta} = \frac{1}{N} + \alpha_{k,N} + \frac{b_{k+1}[f] - b_k[f]}{\beta} \\ &\ge \frac{1}{N} + \alpha_{k,N} - \frac{2}{\beta} \max_k |b_k[f]| \ge \alpha_{k,N}. \end{aligned} \tag{2.7.41}$$

We have proved the following.

Theorem 2.7.4. *Let $f(u)$ be a polynomial of degree N such that $f(0) = f(1) = 0$. Then, there exists a random threshold voting model representation in terms of a purely N-ary branching Brownian motion for the solution to the initial value problem (2.7.14) with the initial condition $g(x)$ that is continuous and satisfies $0 \leq g(x) \leq 1$ for all $x \in \mathbb{R}^d$.*

2.7.6 Recursive up the tree propagation models for other nonlinearities

The random outcome and random threshold voting models apply to equations of the form

$$u_t = \Delta u + f(u), \tag{2.7.42}$$

with nonlinearities $f(u)$ such that

$$f(0) = f(1) = 0. \tag{2.7.43}$$

The reason for this restriction is that the above assumption guarantees that if the initial condition $u(0, x) = g(x)$ satisfies $0 \leq g(x) \leq 1$ for all $x \in \mathbb{R}^d$, then

$$0 < u(t, x) < 1 \text{ for all } t > 0 \text{ and } x \in \mathbb{R}^d. \tag{2.7.44}$$

Thus, it is conceivable that $u(t, x)$ can be interpreted as a probability of some event. For a general polynomial $f(u)$ that does not satisfy (2.7.43), solutions to (2.7.42) do not necessarily satisfy (2.7.44), so there is no reason to expect that they can be interpreted as a probability. However, we can replace the voting model interpretation by a recursive propagation up the genealogical tree $\mathcal{T}(t)$ of the branching Brownian motion that we now describe.

Let

$$f(u) = f_0 + f_1 u + \cdots + f_N u^N \tag{2.7.45}$$

be a polynomial of degree N. Consider the corresponding symmetric polynomial of N variables

$$S_N(u_1, \ldots, u_N) = f_0 + \frac{f_1}{N}(u_1 + \cdots + u_N) + f_2 \binom{N}{2}^{-1} \sum_{k \neq j} u_k u_j + \cdots + f_N \prod_{i=1}^{N} u_i, \tag{2.7.46}$$

so that

$$f(u) = S_N(u, \ldots, u). \tag{2.7.47}$$

To build a solution to (2.7.42) with $f(u)$ as above, we run a purely N-ary BBM, with an exponential clock running at the rate $\beta = 1$, until a time $t > 0$. The particles $X_1, \ldots, X_{N_t}$ that are present at the time t are assigned the random values $u_k = g(X_k(t))$. Here, $g(x)$ is a given continuous function. Then, we propagate the values up the genealogical tree $\mathcal{T}(t)$ by assigning to each parent the value

$$u_{\text{parent}} = S_N(u_1, \ldots, u_N) + \frac{u_1 + \cdots + u_N}{N}. \tag{2.7.48}$$

Here, $u_1, \ldots, u_N$ are the values that have been previously assigned to the N children of the parent under consideration. This recursive procedure allows us to define the value u_{orig} at the root of the tree, the original particle that was present at the time $t = 0$ at the position $x \in \mathbb{R}^d$. We set

$$u(t, x) = \mathbb{E}_x[u_{\text{orig}}]. \tag{2.7.49}$$

Once again, the renewal argument using the independence of the offspring particles, nearly identical to that in (2.7.13)-(2.7.14), shows that $u(t, x)$ satisfies the initial value problem

$$\begin{aligned} u_t &= \Delta u + f(u), \\ u(0, x) &= g(x). \end{aligned} \tag{2.7.50}$$

This gives the following.

Theorem 2.7.5. *Let $f(u)$ be a polynomial of degree N. Then, there exists a recursive up the tree propagation representation in terms of a purely N-ary branching Brownian motion for the solution to the initial value problem (2.7.50) with the initial condition $g(x)$ that is continuous and bounded.*

We point out that this model is deterministic and so it is not a special case of either of the two random voting models above.

2.8 Examples of voting models

In this section, we consider some examples of voting models, beyond the heat equation and the Allen-Cahn equation we have considered above.

2.8.1 Random outcome voting models for the McKean nonlinearities

Let us go back to the McKean type nonlinearities of the form

$$f(u) = \beta\Big(1 - u - \sum_{k=2}^{N} p_k (1-u)^k\Big), \tag{2.8.1}$$

with

$$\sum_{k=2}^{N} p_k = 1. \tag{2.8.2}$$

First, we note the elementary identity:

$$1-(1-u)^n = (1-u+u)^n - (1-u)^n = \sum_{k=0}^{n} \binom{n}{k} u^{n-k}(1-u)^k - (1-u)^n = \sum_{k=0}^{n-1} \binom{n}{k} u^{n-k}(1-u)^k. \tag{2.8.3}$$

This allows us to write a nonlinearity of the form (2.8.1), using (2.8.2) and (2.8.3), as

$$\begin{aligned} f(u) &= \beta\Big(1 - u - \sum_{n=2}^{N} p_n (1-u)^n\Big) = \beta \sum_{n=2}^{N} p_n \Big(1 - (1-u)^n - u\Big) \\ &= \beta \sum_{n=2}^{N} p_n \Big(\sum_{k=0}^{n-1} \binom{n}{k} u^{n-k}(1-u)^k - u\Big). \end{aligned} \tag{2.8.4}$$

Comparing to (2.7.15), we see that this corresponds to the random outcome voting model that is not really random: we have $\alpha_{0n} = 0$ for all $n \geq 2$, and

$$\alpha_{kn} = 1, \quad \text{for all } 1 \leq k \leq n. \tag{2.8.5}$$

Therefore, the McKean nonlinearities come from a very simple voting rule: the parent particle votes 1 if and only if at least one of its children voted 1. This, of course, agrees with the familiar interpretation of the probability distribution of the maximum of BBM in terms of the solution to the Fisher-KPP equation [24, 31].

2.8.2 Uniformly biased voting models

Next, we introduce a uniform bias in the random outcome voting model (2.7.17) we have obtained for the standard heat equation. A parent with n children, out of which k voted 1, now votes 1 with a "uniformly biased" probability

$$\alpha_{kn} = \frac{(1+\gamma)k}{n}, \quad 0 \le k \le n-1, \quad \alpha_{nn} = 1. \tag{2.8.6}$$

Here, $\gamma \ge 0$ is a parameter measuring the "bias" toward voting 1 versus voting 0. As we need to have $\alpha_{kn} \le 1$ for all $1 \le k \le n-1$, the bias $\gamma > 0$ needs to satisfy

$$\gamma \le \frac{n}{n-1} - 1 = \frac{1}{n-1}. \tag{2.8.7}$$

In particular, if γ is fixed, only finitely many p_n may be non-zero. Using expression (2.8.6) for α_{kn} in (2.7.15) and recalling (2.7.16) gives the corresponding nonlinearity as

$$\begin{aligned} f(u) &= \beta \sum_{n=2}^{N} p_n \Big(\sum_{k=0}^{n} \binom{n}{k} \alpha_{kn} u^k (1-u)^{n-k} - u \Big) \\ &= \beta \sum_{n=2}^{N} p_n \Big(\sum_{k=0}^{n} \binom{n}{k} \frac{(1+\gamma)k}{n} u^k - \gamma u^n - u \Big) = \beta\gamma \sum_{n=2}^{N} p_n (u - u^n). \end{aligned} \tag{2.8.8}$$

Taking $\gamma = \beta^{-1}$ gives

$$f(u) = u - A(u) = \sum_{k=2}^{N} p_k (u - u^k), \quad A(u) = \sum_{k=2}^{N} p_k u^k. \tag{2.8.9}$$

As we have seen in Section 2.6, these nonlinearities are of the Fisher-KPP type but do not have a McKean representation. Instead, such nonlinearities come from voting models with a uniform bias toward voting 1, as in (2.8.6). They lead to convex functions $A(u)$ such that $\alpha(u) = A(u)/u$ is also convex, which is impossible for the McKean type.

2.8.3 Group voting models

The nonlinearities of the form (2.8.8) have the property that $f'(0) \ne 0$ except in the trivial cases when $\beta = 0$ or $\gamma = 0$. In order to obtain a voting model representation for nonlinearities with derivatives that vanish at $u = 0$, it is convenient to consider voting with a "group-based" bias. Let us fix some $m > 1$ and assume that branching can only happen into $n > m$ children; that is, $p_k = 0$ for all $k \le m$. The voting scheme is as follows: if a parent has n children, of which k vote 1, and $k < m$, then the parent votes 1 with the unbiased probability

$$\alpha_{n,m}^{(k)} = \frac{k}{n}, \quad \text{if } 0 \le k < m, \tag{2.8.10}$$

as in (2.7.17). However, if $m \le k \le n-1$, so that one can choose a group of m out of n children that all voted 1, then the parent votes 1 with the biased probability

$$\alpha_{n,m}^{(k)}(\gamma) = \frac{k}{n} + \gamma \binom{k}{m} \binom{n}{m}^{-1}, \quad \text{if } m \le k < n-1, \tag{2.8.11}$$

and, finally, if all children voted 1, then

$$\alpha_{n,m}^{(k)} = 1, \quad \text{if } k = n. \tag{2.8.12}$$

Thus, the bias for the parent to vote 1 relative to the unbiased probability k/n is proportional to the ratio of the number $\binom{k}{m}$ of m-tuples such that all particles in the m-tuple voted 1 to the total number $\binom{n}{m}$ of m-tuples of the n children. This is a generalization of the bias in the voting model (2.8.6) for the nonlinearity $u - u^n$, where the m-tuple is simply a single particle. This leads to the nonlinearity

$$\begin{aligned} f(u) &= \beta \sum_{n=m+1}^{N} p_n \Big(\sum_{k=0}^{n} \binom{n}{k} \alpha_{nm}^{(k)}(\gamma) u^k (1-u)^{n-k} - u \Big) \\ &= \beta\gamma \sum_{n=m+1}^{N} p_n \sum_{k=m}^{n-1} \binom{n}{k} \binom{k}{m} \binom{n}{m}^{-1} u^k (1-u)^{n-k}. \end{aligned} \tag{2.8.13}$$

Here, we used (2.7.16) and (2.8.10)-(2.8.12).

Exercise 2.8.1. Show that

$$f(u) = \beta\gamma \sum_{n=m+1}^{N} p_n (u^m - u^n). \tag{2.8.14}$$

Thus, the group voting models lead to this simple class of nonlinearities that vanish at least quadratically at the origin.

2.9 The up the tree voting models revisited and nonlinear integral equations

2.9.1 The multiplication voting rule

Let us consider an up the tree voting model with the rule that if the two children of a given particle voted $\eta \in \mathbb{R}$ and $\eta' \in \mathbb{R}$, then the parent particle votes

$$\xi(\eta, \eta') = \eta + \eta' - \eta\eta'. \tag{2.9.1}$$

Note that if either $\eta = 1$ or $\eta' = 1$ then $\xi = 1$:

$$\xi(1, \eta) = \xi(\eta, 1) = 1, \quad \text{for all } \eta \in \mathbb{R}. \tag{2.9.2}$$

If the only possible votes are 0 and 1 then the parent votes 1 if at least one of its children votes 1. This is exactly the voting model that corresponded to the maximum of a branching Brownian motion. Shifting the votes by 1 and flipping the sign, the voting rule (2.9.1) can be re-written as

$$-\xi + 1 = -\eta + 1 - \eta' + 1 - (-\eta + 1)(-\eta' + 1) = -\eta + 1 - \eta' + 1 - \eta\eta' + \eta' + \eta - 1, \tag{2.9.3}$$

which is simply the product rule

$$\xi = \eta\eta'. \tag{2.9.4}$$

In particular, if all votes of the children satisfy $|\xi| \le 1$, then the votes propagated up the tree satisfy the same condition.

This gives the following evolution equation for the probability density $p(t, x, \xi)$ for the vote of the original particle:

$$\frac{\partial p(t,x,\xi)}{\partial t} = \Delta_x p(t,x,\xi) - p(t,x,\xi) + \int p(t,x,\eta) p(t,x,\eta') \delta(\xi - \eta\eta') d\eta d\eta'. \tag{2.9.5}$$

We see that if the initial condition $p(0,x,\xi) = p_0(x,\xi)$ is supported in the interval $[-1,1]$ in ξ then the solution $p(t,x,\xi)$ is supported in $[-1,1]$ in ξ for all $t > 0$ as well. In particular, all moments

$$I_m(t,x) = \int |\xi|^m p(t,x,\xi) d\xi \tag{2.9.6}$$

of $p(t,x,\xi)$ remain finite for all $t > 0$ and $m \geq 0$ in that case. On the other hand, if $p_0(x,\xi)$ is supported outside of $[-1,1]$ in ξ, then $p(t,x,\xi)$ is supported outside of $[-1,1]$ as well, for all $t > 0$. Note that

$$M_0(t,x) = \int p(t,x,\xi) d\xi \tag{2.9.7}$$

satisfies

$$\frac{\partial M_0}{\partial t} - \Delta_x M_0 = -M_0 + M_0^2. \tag{2.9.8}$$

Therefore, if

$$\int p_0(x,\xi) d\xi = 1, \quad \text{for all } x \in \mathbb{R}^d, \tag{2.9.9}$$

then

$$M_0(t,x) := \int p(t,\xi) d\xi = 1, \quad \text{for all } t > 0, \tag{2.9.10}$$

as it should be.

2.9.2 The moments and convergence to $\delta(\xi)$

Let us now look at the moments $I_m(t,x)$ defined by (2.9.6). They satisfy

$$\begin{aligned} \frac{\partial I_m(t,x)}{\partial t} - \Delta_x I_m(t,x) &= -I_m(t,x) + \int |\xi|^m z(t,x,\eta) z(t,x,\eta') \delta(\xi - \eta\eta') d\eta d\eta' d\xi \\ &= -I_m(t,x) + \int |\eta|^m |\eta'|^m z(t,x,\eta) z(t,x,\eta') d\eta d\eta' = -I_m(t,x) + I_m^2(t,x). \end{aligned} \tag{2.9.11}$$

Thus, the equation for all $I_m(t,x)$ is

$$\frac{\partial I_m}{\partial t} = \Delta I_m - I_m + I_m^2. \tag{2.9.12}$$

We deduce from (2.9.12) that $I_m(t,x)$ remains finite for short times $0 < t < t_m$, with t_m that depends on $I_m(0,x)$. In particular, we have $t_m = +\infty$ if $I_m(0,x) \leq 1$ for all $x \in \mathbb{R}^d$. However, if $I_m(0,x) > 1$ on a sufficiently large region, the moment $I_m(t,x)$ blows up in a finite time.

Furthermore, if there exists $m > 0$ such that $I_m(0,x) < 1$ for all $x \in \mathbb{R}^d$, then $I_m(t,x) \to 0$ as $t \to +\infty$, uniformly on compact sets.

Exercise 2.9.1. Suppose that there exists $m > 0$ so that $I_m(0) < 1$. Then, the solution to (2.9.5) satisfies

$$p(t,x,\xi) \to \delta(\xi), \quad \text{as } t \to +\infty, \tag{2.9.13}$$

in the sense of distributions, locally uniformly in $x \in \mathbb{R}^d$.

2.9.3 Logarithmic moments

To understand the solutions for which

$$I_m(0) \geq 1, \quad \text{for all } m > 0, \tag{2.9.14}$$

let us ignore the spatial dependence and consider the logarithmic moments

$$L_m(t) = \int (\log|\xi|)^m z(t,\xi) d\xi. \tag{2.9.15}$$

We multiply (2.9.5) by $(\log|\xi|)^m$ and integrate. This gives

$$\begin{aligned} \frac{dL_m(t)}{dt} + L_m(t) &= \int (\log|\eta_1\eta_2|)^m p(t,\eta_1)p(t,\eta_2)d\eta_1 d\eta_2 \\ &\leq C_m \int \left[(\log|\eta_1|)^m + (\log|\eta_2|)^m\right] p(t,\eta)p(t,\eta_2)d\eta d\eta_2 = 2C_m L_m(t). \end{aligned} \tag{2.9.16}$$

We conclude that

$$L_m(t) < +\infty, \quad \text{for all } t > 0 \text{ and } m > 0 \text{ if } L_m(0) < +\infty. \tag{2.9.17}$$

In particular, the solution can not develop a $\delta(\xi)$ component in a finite time if it is not present at $t = 0$.

2.9.4 Diffusion in the logarithmic variables

Note that if $z_0(x,\xi)$ is supported on $\{\xi > 0\}$, the solution remains supported in that region. Thus, we can make a change of variable

$$p(t,x,\xi) = \frac{q(t,x,\log\xi)}{\xi}, \quad q(t,x,y) = e^y z(t,x,e^y), \ \xi > 0, \ -\infty < y < +\infty, \tag{2.9.18}$$

so that

$$\int_{-\infty}^{\infty} q(t,x,y)dy = \int_{-\infty}^{\infty} e^y z(t,x,e^y)dy = \int_0^{\infty} z(t,x,\xi)d\xi = 1. \tag{2.9.19}$$

A short computation shows that the equation for $q(t,y)$ is

$$\frac{\partial q(t,x,y)}{\partial t} = \Delta_x q(t,x,y) - q(t,x,y) + \int_{-\infty}^{\infty} q(t,x,y_1)q(t,x,y-y_1)dy_1. \tag{2.9.20}$$

We may also add diffusion in the vote and in space, leading to

$$\frac{\partial q(t,x,y)}{\partial t} = \Delta_x q(t,x,y) + \sigma^2 \Delta_y q(t,x,y) - q(t,x,y) + \int_{-\infty}^{\infty} q(t,x,y_1)q(t,x,y-y_1)dy_1. \tag{2.9.21}$$

An interesting question is to understand how the votes spread to large values. The reason why this is relevant will become more clear when we talk about systems of equations in the next lecture.

Note that the total pointwise mass

$$M(t,x) = \int q(t,x,y)dy \tag{2.9.22}$$

satisfies

$$\frac{\partial M(t,x)}{\partial t} = \Delta M(t,x) - M(t,x) + M^2(t,x). \tag{2.9.23}$$

Therefore, if

$$M(0,x) \equiv 1, \quad \text{for all } x \in \mathbb{R}, \tag{2.9.24}$$

then

$$M(t,x) \equiv 1, \quad \text{for all } t > 0 \text{ and } x \in \mathbb{R}. \tag{2.9.25}$$

Let us also define

$$u_\lambda(t,x) = \int_{-\infty}^{\infty} e^{\lambda y} q(t,x,y) dy. \tag{2.9.26}$$

This function satisfies

$$\begin{aligned} \frac{\partial u_\lambda}{\partial t} - \Delta_x u_\lambda + u_\lambda - \sigma\lambda^2 u_\lambda &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\lambda y} q(t,x,y_1) q(t,x,y-y_1) dy_1 dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\lambda(y_1+y_2)} q(t,x,y_1) q(t,x,y_2) dy_1 dy_2 = u_\lambda^2. \end{aligned} \tag{2.9.27}$$

Thus, we have

$$\frac{\partial u_\lambda}{\partial t} = \Delta_x u_\lambda - (1-\sigma^2\lambda^2) u_\lambda + u_\lambda^2. \tag{2.9.28}$$

We see the appearance of the Fisher-KPP equation for the function

$$v_\lambda(t,x) = 1 - \sigma^2\lambda^2 - u_\lambda(t,x). \tag{2.9.29}$$

The solution remains finite as long as

$$u_\lambda(0,x) = \int e^{\lambda\xi} p_0(x,\xi) d\xi \le 1 - \sigma^2\lambda^2. \tag{2.9.30}$$

In other words, as long as the initial exponential moment is sufficiently small, even though the solution will populate arbitrarily large vote values, the solution to the Fisher-KPP equation will remain finite. To the best of our knowledge, the long time behavior of the probability density $p(t,x,y)$ has not been studied in the regime when $u_\lambda(0,x)$ is above the threshold.

3 Lecture 2: The frontier of BBM in one and higher dimensions

3.1 Overview of the lecture

We will discuss in this lecture the long time behavior of the extrema of branching Brownian motion, both in one dimension and in higher dimensions. In one dimension, the situation has been quite well understood but the results in higher dimensions are much more recent. We will only consider the binary BBM for simplicity of notation.

In one dimension, consider a binary BBM that starts at $x = 0$ at the time $t = 0$, and let $X_1(t), \dots, X_{N_t}(t)$ be the BBM particles present at time $t > 0$. We have seen in the first lecture that the cumulative distribution function

$$v(t,x) = \mathbb{P}\Big(\max_{1 \le k \le N_t} X_k(t) \ge x \Big) \tag{3.1.1}$$

of the maximum particle of BBM satisfies the Fisher-KPP equation

$$\begin{aligned} &\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + v - v^2, \ t > 0, \ x \in \mathbb{R}, \\ &v(0,x) = \mathbb{1}(x \le 0). \end{aligned} \tag{3.1.2}$$

Thus, the law of the maximum of a binary BBM is directly related to the long time behavior of the solutions to (3.1.2). We will first review the asymptotics of the solutions to (3.1.2). Next, we will look at the higher dimensional case. More precisely, we will be interested in the following question: how far from each other can be the furthest locations of a multi-dimensional BBM in various directions?

3.2 Traveling waves for positive nonlinearities

We first recall some basic facts about traveling waves for equations of the form

$$u_t = u_{xx} + f(u),\ t > 0,\ x \in \mathbb{R}, \tag{3.2.1}$$

with a non-negative $f(u)$:

$$f(0) = f(1) = 0, \quad f(u) > 0 \quad \text{for all } 0 < u < 1. \tag{3.2.2}$$

This assumption means that for the ODE

$$\dot{u} = f(u),$$

the steady state $u = 0$ is unstable and $u = 1$ is a stable steady state.

Traveling waves are solutions to (3.2.1) of the form

$$u(t, x) = U_c(x - ct). \tag{3.2.3}$$

In order for such $u(t, x)$ to be a solution to (3.2.1), the function $U_c(x)$ has to satisfy the ODE

$$-cU_c' = U_c'' + f(U_c). \tag{3.2.4}$$

In addition, we require $U_c(x)$ to satisfy the following boundary conditions at infinity

$$U_c(x) \to 1, \quad \text{as } x \to -\infty, \quad U_c(x) \to 0, \quad \text{as } x \to -\infty. \tag{3.2.5}$$

Here is the key result on the existence of traveling waves for positive nonlinearities.

Proposition 3.2.1. *Assume that the nonlinearity $f(u)$ satisfies the positivity assumption (3.2.2). Then, there exists $c_* > 0$ such that equation (3.2.4) has positive solutions $U_c(x) > 0$ that satisfy the boundary conditions (3.2.5) if and only if $c \geq c_*$. Moreover, if, in addition to (3.2.2), $f(u)$ satisfies the Fisher-KPP condition*

$$f(0) = f(1) = 0, \quad f(u) > 0,\ f(u) \leq f'(0)u, \quad \textit{for all } 0 < u < 1, \tag{3.2.6}$$

then

$$c_* = 2\sqrt{f'(0)}. \tag{3.2.7}$$

For the Fisher-KPP nonlinearities, this result goes back to the original papers [47, 72], and for general positive nonlinearities it was proved in [60]. In this lecture, we will mostly assume that $f(u)$ is of the Fisher-KPP type.

An important role below will be played by the traveling wave profile $U_*(x) = U_{c_*}(x)$ that corresponds to the minimal speed c_*, and its asymptotics as $x \to +\infty$. Let us set

$$\lambda_c = \frac{c - \sqrt{c^2 - 4f'(0)}}{2}, \quad \text{for } c > c_*, \quad \lambda_* = \frac{c_*}{2} = \sqrt{f'(0)}. \tag{3.2.8}$$

Proposition 3.2.2. *Let $f(u)$ satisfy the Fisher-KPP assumption (3.2.6). If $c > c_*$ then there exists a constant $A_1 > 0$ so that*

$$U_c(x) \sim A_1 e^{-\lambda_c x}, \quad \text{as } x \to +\infty, \tag{3.2.9}$$

and there exists $A_2 > 0$ so that

$$U_*(x) \sim A_2 x e^{-\lambda_* x}, \quad \text{as } x \to +\infty. \tag{3.2.10}$$

Proposition 3.2.2 says that for $c > c_*$ the traveling waves $U_c(x)$ have a "purely exponential" decay as $x \to +\infty$ but the minimal speed traveling wave $U_*(x)$ has an extra factor of x in front of the exponential. This turns out to be surprisingly important in the long time evolution of the solutions to (3.2.1). Strangely, it is also a signature of a connection to log-correlated random fields.

Unless stated otherwise, we will fix the translation of the wave by requiring that $U_*(x)$ has the asymtptotics

$$U_*(x) \sim x e^{-\lambda_* x}, \quad \text{as } x \to +\infty, \tag{3.2.11}$$

with the pre-factor equal to 1.

3.3 The Bramson shift and convergence to a wave

3.3.1 Steepness and the KPP convergence result

Convergence of the solution to

$$u_t = u_{xx} + f(u), \ t > 0, \ x \in \mathbb{R}, \tag{3.3.1}$$

in shape to a traveling wave was originally proved in the KPP paper [72]. They assumed that the initial condition is a step function $u_0(x) = \mathbb{1}(x \le 0)$. This assumption may be greatly relaxed but the initial condition does need to decay faster than $e^{-\lambda_* x}$ as $x \to +\infty$: see [26, 27] for the sharp conditions on $u_0(x)$ that are needed. Let us also mention that initial conditions that decay exactly as $e^{-\lambda_* x}$ and as $xe^{-\lambda_* x}$ as $x \to +\infty$ are very relevant in the analysis of the additive and derivative martingales that will be introduced in Section 3.4 below. That is, both solutions with compactly supported initial conditions and those with initial conditions that decay at the critical rate $e^{-\lambda_* x}$ are directly related to the long time behavior of various objects associated naturally to BBM.

To discuss a "modern" version of the KPP "convergence to a traveling wave result" from [72], it is convenient to introduce the notion of steepness. A very nice introduction to the steepness notion is in a recent paper by Giletti and Matano [57]. Let us first consider smooth monotonically decreasing functions $u(x)$, $x \in \mathbb{R}$, such that

$$\lim_{x \to -\infty} u(x) = 1, \quad \lim_{x \to +\infty} u(x) = 0. \tag{3.3.2}$$

Given two such functions u_1 and u_2, we say that u_1 is steeper than u_2 if

$$|u_1'(u_1^{-1}(z))| > |u_2'(u_2^{-1}(z))|, \quad \text{for all } z \in (0,1). \tag{3.3.3}$$

In other words, the graph of $u_1(x)$ is steeper than the graph of $u_2(x)$ when compared at each fixed level $z \in (0,1)$, rather than at a fixed point $x \in \mathbb{R}$. This notion is translation invariant; if u_1 is steeper than u_2, it is also steeper than any translate $u_2(\cdot + h)$, with a fixed $h \in \mathbb{R}$.

The differentiability or even continuity of u_1 and u_2 are not really needed for the steepness comparison. To avoid the slope comparison in (3.3.3), we say that u_1 is steeper than u_2 if for any two translates $\tilde{u}_1(x) = u_1(x - \ell_1)$ and $\tilde{u}_2(x) = u_2(x - \ell_2)$ of u_1 and u_2 there exists x_0 such that $\tilde{u}_1(x) > \tilde{u}_2(x)$ for $x < x_0$ and $\tilde{u}_1(x) < \tilde{u}_2(x)$ for $x > x_0$. Here, we still assume that u_1 and u_2 obey the boundary conditions (3.3.2). For smooth functions, the two definitions are equivalent.

The key observation is that equation (3.3.1) preserves the steepness ordering of the solutions.

Proposition 3.3.1. *Let $u_1(t,x)$ and $u_2(t,x)$ be the solutions to (3.3.1) with the corresponding initial conditions u_{10} and u_{20} that satisfy (3.3.2). If u_{10} is steeper than u_{20}, then $u_1(t,\cdot)$ is steeper than $u_2(t,\cdot)$ for all $t>0$.*

This result was essentially proved for the classical Fisher-KPP equation in the original KPP paper [72]. Let us stress that the proof does not use the positivity assumption (3.2.2) and holds for essentially any smooth nonlinearity $f(u)$. We also mention that there is a corresponding result by Bachmann [23] for random walks with log-concave jump laws. Without the log-concavity assumption on the jump laws, the steepness comparison principle does not hold for random walks.

The following result was also already proved in the KPP paper [72].

Theorem 3.3.2. *Suppose that $f(u)$ satisfies (3.2.2), and let $u(t,x)$ be the solution to (3.3.1) with the initial condition u_{in} that satisfies (3.3.2) and is steeper than the minimal speed traveling wave $U_*(x)$. Then, there exists a function $m(t)$ such that*

$$\frac{dm(t)}{dt} \to c_*, \quad \text{as } t \to +\infty, \tag{3.3.4}$$

and

$$u(t, x+m(t)) \to U_*(x) \quad \text{as } t \to +\infty, \text{ uniformly on } \mathbb{R}. \tag{3.3.5}$$

We should emphasize that the requirement that the initial condition is steeper than the minimal speed traveling wave is absolutely not needed and can be replaced by an assumption that $u_0(x)$ tends to zero sufficiently fast at infinity. We make it here purely for convenience. It is, however, required in the intersection number type arguments such as those invented in the original KPP paper.

Theorem 3.3.2 says nothing about the location $m(t)$ of the front of the solution to the Fisher-KPP equation,

$$\begin{aligned} &u_t - u_{xx} = f(u), \quad t>0,\ x\in\mathbb{R},\\ &u(0,x) = u_0(x), \end{aligned} \tag{3.3.6}$$

except for the rough asymptotics (3.3.4)

$$m(t) = c_* t + o(t), \quad \text{as } t \to +\infty. \tag{3.3.7}$$

3.3.2 The Bramson shift and the law of the maximum of BBM

Fisher has already made an informal argument in [47] that the $o(t)$ term in (3.3.7) is of the order $O(\log t)$. More precisely, he claimed that it equals to $(1/2)\log t$ to the leading order. An important series of papers by Bramson [32], [33] proved the following corrected version of Fisher's prediction.

Theorem 3.3.3. *Suppose that $f(u)$ satisfies the Fisher-KPP property (3.2.2) and $u_0(x)$ is compactly supported on the right. There is a constant x_∞, depending on u_0, such that*

$$m(t) = c_* t - \frac{3}{2\lambda_*}\log t - x_\infty + o(1), \quad \text{as } t \to +\infty, \tag{3.3.8}$$

with $\lambda_* = c_*/2$.

This theorem has the following interpretation in terms of the branching Brownian motion. Let

$$M(t) = \max_{1\le k\le N_t} X_k(t), \tag{3.3.9}$$

be the running maximum of the binary BBM that started at $x = 0$ at the time $t = 0$. We have seen in the first lecture that the cumulative distribution function

$$v(t, x) = \mathbb{P}\Big(\max_{1 \le k \le N_t} X_k(t) \ge x \Big) \tag{3.3.10}$$

of the maximum particle of the binary BBM satisfies the Fisher-KPP equation

$$\begin{aligned} &\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + v - v^2, \ t > 0, \ x \in \mathbb{R}, \\ &v(0, x) = \mathbb{1}(x \le 0). \end{aligned} \tag{3.3.11}$$

Theorem 3.3.3 says that there exists x_∞ such that

$$\mathbb{P}\Big(M(t) > c_* t - \frac{3}{2\lambda_*} \log t - x_\infty + x \Big) \to U_*(x), \quad \text{as } t \to +\infty. \tag{3.3.12}$$

In other words, the location of the maximum is tight around the Bramson location $m(t)$ given by (3.3.8) and its cumulative probability distribution converges to a traveling wave.

Bramson's result in Theorem 3.3.3, applies more generally to equations of the form (3.3.6) with a McKean type nonlinearity $f(u)$ and initial conditions that tend to zero sufficiently fast as $x \to +\infty$. The logarithmic correction in (3.3.8), with the precise factor of $3/2$, is surprisingly important and shows up in many other models involving log-correlated random fields. Some examples of random processes in this class are briefly discussed in Section 3.3.3 below, without doing this issue any justice, see references mentioned in that section. The recent PDE proofs of Bramson's asymptptics and further extensions can be found in [26, 27, 58, 64, 98, 99].

Bramson's proof of Theorem 3.3.3 used probabilistic arguments coming from the connection to branching Brownian motion and applied only to the McKean nonlinearities. We will briefly describe below a simple and reasonably robust proof of Theorem 3.3.3 from [64, 98] that works for all Fisher-KPP nonlineariies and does not use the intersection number argument. These ideas are further developed to study the refined asymptotics of the solutions in [58, 99]. A completely different and totally mind boggling even if not rigorous approach to these further corrections was developed in [26, 27]. Previous PDE results on the Bramson shift are [74, 112], while there are many probabilistic papers addressing the maximum of a branching Brownian motion, or a branching random walk: see, for example, [7, 9, 10, 23, 34, 35, 40, 81, 103].

Let us mention two important cases when the initial conditions are not compactly supported on the right. First, if

$$u_0(x) \sim e^{-\lambda_* x}, \quad \text{as } x \to +\infty, \tag{3.3.13}$$

then the front location asymptotics instead of (3.3.8) is

$$m(t) = c_* t - \frac{1}{2\lambda_*} \log t - x_\infty + o(1), \ \text{as } t \to +\infty. \tag{3.3.14}$$

On the other hand, if

$$u_0(x) \sim x e^{-\lambda_* x}, \quad \text{as } x \to +\infty, \tag{3.3.15}$$

then the front location asymptotics instead of (3.3.8) is

$$m(t) = c_* t - x_\infty + o(1), \ \text{as } t \to +\infty, \tag{3.3.16}$$

without any logarithmic correction. Once again, these asymptotics are relevant in the analysis of the additive and derivative martingales in Section 3.4.

3.3.3 The Bramson correction and other log-correlated random fields

Very surprisingly, the $3/2$ factor in front of the $\log t$ term in (3.3.8) is much more than it seems. We are not going to discuss this issue in any detail in these lectures but let us just give a couple of examples of this phenomenon, with some further references.

Maximum of independent Gaussians. First, let discuss a case when the asymptotics in (3.3.12) has the factor $1/2$ predicted by Fisher and not Bramson's $3/2$. This happens in a simplified model with no correlations between particles. Let us think of $N \in \mathbb{N}$ as an analog of the time variable for the BBM and take a large number M of independent particles $Y_1, \dots, Y_M$, with M that depends on N. In order to mimic the concept that N is the time of the BBM, we assume that Y_k are mean zero Gaussian random variables with variance N. Thus, we can think of each Y_k as a snapshot of a Brownian motion at the time $t = N$. To make sure that the number M of these variables also mimics BBM, we assume that it grows exponentially in N, as is the case for BBM. We take $M = 2^N$, to ensure that M is an integer. The key difference with the BBM is that here, for an infinitely greater simplicity, we assume that all Y_k are simply independent. Let us consider the maximal particle

$$\overline{M}_N = \max(Y_1, \dots, Y_M). \tag{3.3.17}$$

In particular, we are interested in the median location m_N such that

$$\mathbb{P}(\overline{M}_N > m_N) = 1/2. \tag{3.3.18}$$

In this simple model, since all particles are identical, this can be analyzed by writing

$$\mathbb{P}(\overline{M}_N < y) = \big(\mathbb{P}(Y_1 < y)\big)^{2^N} = \big(1 - \mathbb{P}(Y_1 > y)\big)^{2^N}. \tag{3.3.19}$$

A straightforward if lengthy calculation shows that

$$m_N = c_* N - \frac{1}{2\lambda_*} \log N + x_0 + o(1), \quad \text{as } N \to +\infty, \tag{3.3.20}$$

with the corresponding c_* and $\lambda_* = c_*/2$ coming from the fact that $M = 2^N$ and not $\exp(N)$. This is exactly Fisher's prediction – the coefficient here is $1/2$ and not Bramson's $3/2$.

Furthermore, the same computation shows that in the limit $N \to +\infty$ we get the Gumbel distribution:

$$u_N(y) = \mathbb{P}\Big(\overline{M}_N < m_N + y\Big) \to \exp\Big(- \frac{1}{c_*\sqrt{2\pi}} e^{-\lambda_* y}\Big), \quad \text{as } N \to +\infty. \tag{3.3.21}$$

It is helpful to think of (3.3.21) as

$$u_N(y) \to \exp\Big(- e^{-\lambda_*(y - \ell_*)}\Big), \quad \text{as } N \to +\infty, \tag{3.3.22}$$

with the shift

$$\ell_* = -\log(c_*\sqrt{2\pi}). \tag{3.3.23}$$

We see from this calculation that Bramson's asymptotics should come from the correlations built into the positions of BBM – these correlations slow down the maximum compared to independent Gaussians. The correlations come from the genealogical tree structure and thus have a logarithmic nature. It turns out that Bramson's asymptotics for the extremal values of such log-correlated fields are quite ubiquitous.

Maxima of the Riemann zeta function. The first example of the Bramson-like behavior concerns the maxima of the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} \tag{3.3.24}$$

on the critical line. The Lindelöf hypothesis says that for any $\varepsilon > 0$ we have

$$\zeta(1/2 + iT) = o(T^{\varepsilon}), \quad \text{as } T \to +\infty. \tag{3.3.25}$$

To the best of my knowledge, the best result in this direction is by Bourgain [30] who showed that

$$\zeta(1/2 + iT) = o(T^{13/84+\varepsilon}). \tag{3.3.26}$$

Fyodorov, Hiary and Keating, in a series of papers [49, 50], considered the following related question. Consider the Riemann zeta function on the critical line, and pick an interval of length 1, unfiormly at random, on an interval of the form $[T, 2T]$, with some $T \gg 1$. That is, we consider

$$\tilde{\zeta}(t) = \max_{|t-s|\leq 1} \log|\zeta(1/2 + is)|, \tag{3.3.27}$$

with t chosen uniformly on $[T, 2T]$. We are interested in

$$u(T, x) = \text{Prob}\big[\tilde{\zeta}(t) > x\big]. \tag{3.3.28}$$

They conjectured that

$$u\big(T, x + \log\log T - \frac{3}{4}\log\log\log T\big) \sim F(x), \tag{3.3.29}$$

with a function $F(x)$ that has the same asymptotics as $x \to +\infty$ as the Fisher-KPP traveling wave:

$$F(y) \sim Aye^{-2y}, \quad \text{as } y \to +\infty. \tag{3.3.30}$$

After a simple rescaling $x \to x/2$, (3.3.29) is exactly the same asymptotics as in (3.3.8), and the profile asymptotics (3.3.30) is exactly the same as (3.2.10) for the Fisher-KPP traveling wave!

As far as the rigorous the results in this direction, Arguin, Belius, Bourgade, Radziwill, and Soundararajan proved in [13] that for any $\varepsilon > 0$,

$$\text{Prob}\Big\{(1-\varepsilon)\log\log T < |\tilde{\zeta}(t)| \leq (1+\varepsilon)\log\log T\Big\} \to 1, \quad \text{as } T \to +\infty. \tag{3.3.31}$$

This is the speed asymptotics, exactly as (3.3.7) of the original KPP paper.

The most recent result, to the best of my knowledge, is by Arguin, Bourgade, and Radziwill [14] who proved a Bramson-like upper bound

$$u(T, x + \log\log T - \frac{3}{4}\log\log\log T + y) \leq Cye^{-2y}. \tag{3.3.32}$$

A similar lower bound was obtained very recently by the same authors. Further references for these questions for the Riemann zeta function are, among others, [15, 17].

The characteristic polynomial of random unitary matrices. The second example concerns the circular β ensembles of random unitary matrices. Consider random unitary matrices, with the eigenvalue distribution on $\mathbb{S}^n$ given by

$$C_{n,\beta} \prod_{1\leq j<k\leq n} \big|e^{i\theta_j} - e^{i\theta_k}\big|^{\beta} d\theta_1 \dots d\theta_n, \tag{3.3.33}$$

and let $P_n(z)$ be the corresponding characteristic polynomial. The Fyodorov-Hiary-Keating conjecture for $\beta = 2$ is that

$$\max_{z\in\mathbb{S}^1} \log|P_n(z)| = \log n - \frac{3}{4}\log\log n + X_n. \tag{3.3.34}$$

Here, X_n is conjectured to have a limit and Fisher-KPP traveling wave-like law, as in (3.3.32). The best result, as far as I know, is by Chhaibi, Madaule, and Najnudel [38] who capture both the speed and the Bramson correction but not the $O(1)$ term:

$$\max_{z\in\mathbb{S}^1} \mathrm{Re}\log P_n(z) = \sqrt{\frac{2}{\beta}}\Big(\log n - \frac{3}{4}\log\log n + O(1)\Big). \tag{3.3.35}$$

Previous results in this direction are [12, 82, 101].

The advantage of the branching Brownian motion compared to the other examples in this class, known as the log-correlated processes, is that analytic techniques are available to study the Fisher-KPP equation that allow to prove such results in what seems to be a much simpler way, and also make conjectures for the other processes in the log-correlated class. This makes BBM a very interesting special case of the log-correlated processes.

3.3.4 Strategy of the PDE proof of Theorem 3.3.3

We now very briefly discuss the PDE strategy of the proof of the Bramson correction. For simplicity, we assume that the nonlinearity is $f(u) = u - u^2$ but the proof outlined below only relies on the Fisher-KPP property (3.2.2) of $f(u)$. There is a separate question of what happens when $f(u)$ is not of the Fisher-KPP type – we will not worry about this for the moment.

Consider the Cauchy problem (3.3.6)

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = u - u^2, \quad x\in\mathbb{R}, \quad t > 1, \tag{3.3.36}$$

and proceed with a sequence of changes of variables. We first go into the moving frame:

$$x \mapsto x - 2t + r\log t,$$

with $r\in\mathbb{R}$ to be determined, leading to

$$u_t - u_{xx} - (2 - \frac{r}{t})u_x = u - u^2. \tag{3.3.37}$$

Next, we take out the exponential factor: set

$$u(t,x) = e^{-x}v(t,x) \tag{3.3.38}$$

so that $v(t,x)$ satisfies

$$v_t - v_{xx} - \frac{r}{t}(v - v_x) + e^{-x}v^2 = 0, \qquad x\in\mathbb{R}, \quad t > 1. \tag{3.3.39}$$

We note that for $x\to+\infty$, the term $e^{-x}v^2$ in (3.3.39) is negligible, while for $x\to-\infty$ the same term will create a large absorption and force the solution to be close to zero. For this reason, the linear Dirichlet problem

$$\begin{aligned} &z_t - z_{xx} - \frac{r}{t}(z - z_x) = 0, \quad x > 0,\\ &z(t,0) = 0, \end{aligned} \tag{3.3.40}$$

is a reasonable approximation for (3.3.39) for $x \gg 1$. The philosophy of the proof is that if we choose the correct reference frame, where the front is located then $u(t,x)$ should remain of the size $O(1)$ as $t \to +\infty$, for $x \sim O(1)$. Then, so should be $v(t,x)$ and $z(t,x)$. Our goal is to find $r \in \mathbb{R}$ so that $z(t,x)$ would remain $O(1)$ as $t \to +\infty$.

If we "naively" drop the term of the size $O(1/t)$ in (3.3.40), we obtain the heat equation on the half line

$$\begin{aligned} &\tilde{z}_t - \tilde{z}_{xx} = 0, \quad x > 0, \\ &\tilde{z}(t,0) = 0. \end{aligned} \tag{3.3.41}$$

Its solution has the long time asymptotics

$$\tilde{z}(t,x) \sim \frac{Cx}{t^{3/2}} e^{-x^2/(4t)}, \quad \text{as } t \to +\infty. \tag{3.3.42}$$

With some technical work, one can show that the solution to (3.3.40) has the long time behavior

$$z(t,x) \sim \frac{Cx}{t^{3/2-r}} e^{-x^2/(4t)}, \quad \text{as } t \to +\infty. \tag{3.3.43}$$

As we want to keep $z(t,x)$ of the size $O(1)$ as $t \to +\infty$, we are forced to take $r = 3/2$. This is the analytical reason behind the Bramson' correction.

Making the connection between the approximate Dirichlet problem (3.3.40) and the original problem (3.3.37) more precise and quantitative is a key to the proof of Theorem 3.3.3. We omit the technical details that can be found in [64, 98] but the key step is the following lemma.

Lemma 3.3.4. *There exists a constant $r_\infty > 0$ with the following property. For any $\gamma > 0$ and $\varepsilon > 0$ we can find T_ε so that for all $t > T_\varepsilon$ and $x_\gamma = t^\gamma$ we have*

$$|u(t,x_\gamma) - r_\infty x_\gamma e^{-x_\gamma} e^{-x_\gamma^2/(4t)}| \le \varepsilon x_\gamma e^{-x_\gamma} e^{-x_\gamma^2/(6t)}. \tag{3.3.44}$$

Therefore, we have

$$u(t,x) = r_\infty x e^{-x} e^{-x^2/(4t)} + l.o.t. \quad \text{for } x \sim t^\gamma. \tag{3.3.45}$$

This determines the unique translation: if we accept that $u(t,x)$ converges to a translate $U(x - x_\infty)$ of $U_*(x)$, then for large x (in the moving frame) we have

$$u(t,x) \sim U_*(x - x_\infty) \sim x e^{-x + x_\infty}. \tag{3.3.46}$$

Comparing this with (3.3.45), we infer that

$$x_\infty = \log r_\infty.$$

This argument, however, assumes that the two approximations are both valid for $x \sim O(t^\gamma)$. This is quite delicate, the details can be found in [64, 98].

3.4 The randomized Gumbel distribution of the maximum of BBM

Let us go back to the limiting distribution of the maximum of a binary BBM given in (3.3.12): there exists x_∞ such that the probability distribution of the maximum of BBM satisfies

$$\mathbb{P}\Big(M(t) > c_* t - \frac{3}{2\lambda_*} \log t - x_\infty + x\Big) \to U_*(x), \quad \text{as } t \to +\infty. \tag{3.4.1}$$

In order to connect it to the limiting Gumbel distribution for the maximum M_N of e^N independent Gaussians given by (3.3.21)

$$\mathbb{P}\Big(\overline{M}_N < m_N + y\Big) \to \exp\Big(-\frac{1}{c_*\sqrt{2\pi}} e^{-\lambda_* y}\Big), \quad \text{as } N \to +\infty, \tag{3.4.2}$$

we will need to introduce some martingales related to BBM.

3.4.1 The traveling wave martingale

It will be convenient for us not to work with the Fisher-KPP equation (3.2.1) directly but to use the solution to

$$\begin{aligned} &v_t = v_{xx} - v + g(v),\ t > 0,\ x \in \mathbb{R}, \\ &v(0,x) = r(x). \end{aligned} \tag{3.4.3}$$

We will assume that $g(v)$ is of the McKean form

$$g(v) = \sum_{k=1}^{\infty} p_k v^k, \quad \sum_{k=1}^{\infty} p_k = 1, \quad p_k \geq 0. \tag{3.4.4}$$

One can always think of the example $g(v) = v^2$ without losing much generality. Recall that if v is a solution to (3.4.3), then $u = 1 - v$ satisfies the Fisher-KPP equation (3.2.1):

$$\begin{aligned} &u_t = u_{xx} + f(u),\ t > 0,\ x \in \mathbb{R}, \\ &u(0,x) = 1 - r(x), \end{aligned} \tag{3.4.5}$$

with

$$f(u) = 1 - u - g(1-u).$$

McKean's representation tells us that

$$v(t,x) = \mathbb{E} \prod_{k=1}^{N_t} r(X_k^x(t)) = \mathbb{E} \prod_{k=1}^{N_t} r(x + X_k(t)). \tag{3.4.6}$$

Here, $X_k(t)$ are the BBM particles starting at $x = 0$ at $t = 0$, and $X_k^x(t)$ are the BBM particles starting at $t = 0$ at the position $x \in \mathbb{R}$.

Let us go into the frame moving with the speed c_* and note that the function

$$\hat{v}(t,x) = v(t, x + c_* t) \tag{3.4.7}$$

satisfies the initial value problem

$$\begin{aligned} &\hat{v}_t - c_* \hat{v}_x = \hat{v}_{xx} - \hat{v} + g(\hat{v}),\ t > 0,\ x \in \mathbb{R}, \\ &\hat{v}(0,x) = r(x). \end{aligned} \tag{3.4.8}$$

Using (3.4.6) we obtain a representation

$$\hat{v}(t,x) = \mathbb{E} \prod_{k=1}^{N_t} r(x + c_* t + X_k(t)). \tag{3.4.9}$$

Let us introduce the minimal speed traveling wave $\omega(x)$ for (3.4.3), the solution to

$$\begin{aligned} &-c_* \omega' = \omega'' - \omega + g(\omega),\ x \in \mathbb{R}, \\ &\omega(-\infty) = 0, \quad \omega(+\infty) = 1. \end{aligned} \tag{3.4.10}$$

It is related to $U_*(x)$, the minimal speed traveling wave for (4.11.2) by $\omega(x) = 1 - U_*(x)$. We fix the translation of $\omega(x)$ by the normalization $\omega(0) = 1/2$. Note that $\omega(x)$ is a time-independent solution to (3.4.8). Hence, (3.4.9) implies an identity

$$\omega(x) = \mathbb{E}\Big[\prod_{k=1}^{N_t} \omega(x + c_* t + X_k(t)) \Big], \tag{3.4.11}$$

that holds for all $t \geq 0$. This is a manifestation of the following important fact.

Proposition 3.4.1. *For a fixed $x \in \mathbb{R}$, the function*

$$\mathcal{W}(t,x) = \prod_{k=1}^{N_t} \omega(x + c_* t + X_k(t)). \tag{3.4.12}$$

is a martingale. Moreover, $\mathcal{W}(t,x)$ converges almost surely and in L^1 to a limit $\mathcal{W}_\infty(x)$ such that

$$\mathbb{E}\mathcal{W}_\infty(x) = \omega(x). \tag{3.4.13}$$

Proof. Integrability of $\mathcal{W}$ is immediate from the fact that $0 < \omega(x) < 1$. To show the martingale property of $\mathcal{W}(t,x)$, let us take $t > 0$ and $s > 0$. The Markov nature of the exponential clocks and BBM allows us to write

$$\begin{aligned}
\mathbb{E}(\mathcal{W}(t+s,x)|\mathcal{F}_s) &= \mathbb{E}\Big(\prod_{m=1}^{N_{t+s}} \omega(x + c_*(t+s) + X_m(t+s)) \Big| \mathcal{F}_s \Big) \\
&= \prod_{k=1}^{N_s} \mathbb{E}\Big(\prod_{j=1}^{N_t} \omega(x + X_k(s) + c_*(t+s) + X_{kj}(t)) \Big| \mathcal{F}_s \Big) \\
&= \prod_{k=1}^{N_s} \omega(x + X_k(s) + c_* s) = \mathcal{W}(s,x).
\end{aligned} \tag{3.4.14}$$

We used (3.4.11) in the last step above, and also the "lack of memory" property (2.3.8) of the exponential clocks:

$$\mathbb{P}(\tau > s + t | \tau > t) = \mathbb{P}(\tau > s). \tag{3.4.15}$$

Since $\mathcal{W}(t,x)$ is a non-negative martingale, it converges as $t \to +\infty$, almost surely and in L^1 to a limit $\mathcal{W}_\infty(x)$. Taking the expectation in (3.4.12) and passing to the limit $t \to +\infty$ gives (3.4.13). □

Exercise 3.4.2. Use Proposition 3.4.1, the boundary conditions for $\omega(x)$ in (3.4.10), as well as the fact that $0 < \omega(x) < 1$ to understand informally where we should expect the minimum of $X_k(t)$ to be, relative to the position $(-c_* t)$.

3.4.2 The additive and derivative martingales

The next two martingales are constructed out of solutions not to the Fisher-KPP equation but to the linearized Fisher-KPP equation written in the frame moving with the speed c_*:

$$\begin{aligned}
&z_t - c_* z_x = z_{xx} + (\overline{N} - 1) z, \\
&z(0,x) = r(x).
\end{aligned} \tag{3.4.16}$$

Here, $\overline{N}$ is the expected number of the off-spring at each branching:

$$\overline{N} = \sum_{k=1}^{\infty} k p_k. \tag{3.4.17}$$

For a binary BBM we have $\overline{N} = 2$. As we have seen, solutions to (3.4.16) have the probabilistic representation similar to McKean's for the Fisher-KPP equation but with the product replaced by the sum:

$$z(t,x) = \mathbb{E} \sum_{k=1}^{N_t} r(x + c_* t + X_k(t)). \tag{3.4.18}$$

In particular, if $\overline{z}(x)$ is a steady solution to (3.4.16):

$$-c_*\overline{z}_x = \overline{z}_{xx} + (\overline{N}-1)\overline{z}, \tag{3.4.19}$$

then we have an identity, similar to (3.4.11):

$$\overline{z}(x) = \mathbb{E}\sum_{k=1}^{N_t}\overline{z}(x + c_*t + X_k(t)), \tag{3.4.20}$$

that holds for all $t \geq 0$.

Exercise 3.4.3. Go back to Proposition 2.3.3 and explain why $M(t) = e^{-t}N(t)$ is a martingale for a binary BBM in the context of this discussion.

There are two explicit steady solutions to (3.4.16). The first one is

$$\eta(x) = e^{-\lambda_* x}, \tag{3.4.21}$$

with λ_* given by

$$\lambda_* = \frac{c_*}{2}. \tag{3.4.22}$$

The second is

$$\overline{z}(x) = xe^{-\lambda_* x}. \tag{3.4.23}$$

They give us the following two objects:

$$W(t) = \sum_{k=1}^{N_t}\exp(-\lambda_*(c_*t + X_k(t))), \tag{3.4.24}$$

and

$$Z(t) = \sum_{k=1}^{N_t}(c_*t + X_k(t))\exp(-\lambda_*(c_*t + X_k(t))), \tag{3.4.25}$$

that we may expect to be martingales, for the same reason as $\mathcal{W}(t,x)$ is a martingale.

Exercise 3.4.4. Using the proof of Proposition 3.4.1, show that $W(t)$ is a martingale.

The martingale $W(t)$ is non-negative and

$$\mathbb{E}(W(t)) = 1, \tag{3.4.26}$$

for all $t \geq 0$. Hence, it converges as $t \to +\infty$, almost surely to a finite limit $W_\infty \geq 0$. This has the following important consequence.

Exercise 3.4.5. Use the fact that $W(t)$ has a finite limit to show that

$$\lim_{t\to+\infty}\left(c_*t + \min_{1\leq k\leq N_t} X_k(t)\right) = +\infty, \tag{3.4.27}$$

almost surely. Hint: all arguments in the exponentials in the definition of $Y(t)$ must tend to negative infinity as $t \to +\infty$ in order for the limit to exist.

Similarly to (3.4.27), we also have

$$\lim_{t\to+\infty}\Big(c_*t-\max_{1\le k\le N_t}X_k(t)\Big)=+\infty,\ \text{a.s.} \tag{3.4.28}$$

This, roughly speaking, says that there are no BBM particles to the right of the position c_*t, or to the left of $(-c_*t)$, with a very high probability.

Exercise 3.4.6. Use the proof of Proposition 3.4.1, and (3.4.27) and (3.4.28) to show that $Z(t)$ is a martingale.

We will refer to $W(t)$ and $Z(t)$ as, respectively, the additive martingale and the derivative martingale.

3.4.3 The limit of the derivative martingale

The martingale $Z(t,x)$ is not necessarily positive, hence we do not a priori know that it has a limit as $t\to+\infty$. On the other hand, (3.4.27) hints that it should become positive as $t\to+\infty$, at least with a very high probability. To show that, indeed, it does have a limit, let us look again at the traveling wave martingale $\mathcal{W}(t,x)$ and represent it as

$$\begin{aligned}\mathcal{W}(t,x)&=\prod_{k=1}^{N_t}\omega(x+c_*t+X_k(t))=\prod_{k=1}^{N_t}\exp(\log\omega(x+c_*t+X_k(t)))\\&=\exp\Big(\sum_{k=1}^{N_t}\log\omega(x+c_*t+X_k(t))\Big).\end{aligned} \tag{3.4.29}$$

The arguments of ω in the right side of (3.4.29) are all very large as $t\to+\infty$, because of (3.4.27). Let us recall the corresponding asymptotics (3.2.10):

$$U_*(x)\sim xe^{-\lambda_*x},\quad \text{as } x\to+\infty, \tag{3.4.30}$$

which translates into

$$1-\omega(x)\sim xe^{-\lambda_*x},\quad \text{as } x\to+\infty. \tag{3.4.31}$$

With (3.4.27) in mind, we see that

$$\begin{aligned}\log\omega(x+c_*t+X_k(t))&\approx\log\big[1-(x+c_*t+X_k(t))\exp(-\lambda_*(x+c_*t+X_k(t)))\big]\\&\approx-(x+c_*t+X_k(t))\exp(-\lambda_*(x+c_*t+X_k(t))).\end{aligned} \tag{3.4.32}$$

This gives the asymptotics

$$\begin{aligned}&\sum_{k=1}^{N_t}\log\omega(x+c_*t+X_k(t)))\approx-\sum_{k=1}^{N_t}(x+c_*t+X_k(t))\exp(-\lambda_*(x+c_*t+X_k(t)))\\&=-(xW(t)+Z(t))e^{-\lambda_*x}.\end{aligned} \tag{3.4.33}$$

Using this in (3.4.29) gives

$$\mathcal{W}(t,x)\approx\exp\big(-e^{-\lambda_*x}(xW(t)+Z(t))\big). \tag{3.4.34}$$

Exercise 3.4.7. Make the above argument rigorous and show that almost surely we have

$$\mathcal{W}_\infty(x)=\lim_{t\to+\infty}\mathcal{W}(t,x)=\lim_{t\to+\infty}\exp\big(-e^{-\lambda_*x}(xW(t)+Z(t))\big). \tag{3.4.35}$$

Since we know that the limits $\mathcal{W}_\infty(x)$ and W_∞ of $W(t,x)$ and $W(t)$ as $t \to +\infty$ exist almost surely, it follows from (3.4.34) that the limit

$$Z_\infty = \lim_{t\to+\infty} Z(t) \tag{3.4.36}$$

exists and is finite. On the other hand, as we also have (3.4.27), existence of the limit of $Z(t)$ means that we must have

$$W_\infty = 0, \quad \text{a.s.} \tag{3.4.37}$$

Exercise 3.4.8. Use the Laplace transform of $W(t)$ and the Fisher-KPP equation to show that $W(t) \sim t^{-1/2}$ as $t \to +\infty$. Note that

$$1 - e^{-\lambda e^{-x}} \sim \lambda e^{-x}, \quad \text{as } x \to +\infty, \tag{3.4.38}$$

and

$$1 - e^{-\lambda e^{-x}} \sim 1, \quad \text{as } x \to -\infty. \tag{3.4.39}$$

Then you can use (3.3.14).

Thus, passing to the limit $t \to +\infty$ in (3.4.34) we obtain

$$\mathcal{W}_\infty(x) = \exp\big(-e^{-\lambda_* x} Z_\infty\big). \tag{3.4.40}$$

Recalling (3.4.13), we deduce a representation for the traveling wave as the Laplace transform of the limit of the derivative martingale

$$\omega(x) = \mathbb{E}\exp\big(-e^{-\lambda_* x} Z_\infty\big). \tag{3.4.41}$$

Viewed differently, this identifies the law of the derivative martingale via its Laplace transform in terms of the traveling wave: for every $\kappa > 0$ we have

$$\mathbb{E}\exp\big(-\kappa Z_\infty\big) = \omega\Big(-\frac{1}{\lambda_*}\log\kappa\Big). \tag{3.4.42}$$

As we will see, the derivative martingale plays an important qualitative role in the description of the limiting law of the maximum of BBM.

We will not prove this here but will use later the observation that [2, 69]

$$\sqrt{t}W(t) \to CZ_\infty \text{ in probability as } t \to \infty. \tag{3.4.43}$$

Exercise 3.4.9. Use the McKean approach to compute the Laplace transform of $\sqrt{t}W(t)$ and of $Z(t)$ and make a convincing argument why (3.4.43) is plausible. It may helpful to think of $\sqrt{t}W(t)$ as

$$\sqrt{t}W(t) = \sum_{k=1}^{N_t} \exp(-\lambda_*(c_* t - \frac{1}{2\lambda_*}\log t + X_k(t))), \tag{3.4.44}$$

The convergence in (3.4.43) is another reason why $W(t)$ vanishes as $t \to +\infty$. This also suggests that the derivative martingale $Z(t)$ in (3.5.3) is dominated by particles roughly $\sqrt{t}$ behind the BBM frontier. This provides considerable freedom in analyzing $Z(t)$ that we will use in the multi-dimensional case below: we can drop particles outside this window and modify the prefactor by $o(\sqrt{t})$ without changing the long-time limit. Such modifications are not true martingales, but this plays no role our analysis.

3.4.4 The law of the maximum and the derivative martingale

We now connect the law of the maximum of BBM to the derivative martingale. Let us recall that in Section 3.3.3 we have considered the maximum of 2^N independent Gaussian random variables $G_1, \dots, G_{2^N}$, with each G_k having variance N. We have shown that if we set

$$m_N = c_* N - \frac{1}{2\lambda_*} \log N, \tag{3.4.45}$$

with $c_* = \lambda_* = 2\sqrt{2}$, and define

$$u_N(y) = \mathbb{P}\Big(\max_{1 \le k \le 2^N} G_k < m_N + y \Big), \tag{3.4.46}$$

then in the limit $N \to +\infty$ we get the Gumbel distribution

$$u_N(y) \to \exp\Big(- \frac{1}{c_* \sqrt{2\pi}} e^{-\lambda_* y} \Big), \quad \text{as } N \to +\infty. \tag{3.4.47}$$

It is helpful to think of (4.11.6) as

$$u_N(y) \to \exp\Big(- e^{-\lambda_*(y - \ell_*)} \Big), \quad \text{as } N \to +\infty, \tag{3.4.48}$$

with the shift

$$\ell_* = - \log(c_* \sqrt{2\pi}). \tag{3.4.49}$$

We now describe a similar result for the branching Brownian motion – the obvious difference with the above situation is that the locations of the BBM are not uncorrelated. Let $X_1(t), \dots, X_{N_t}(t)$ be the BBM particles present at time $t > 0$. Recall that the function

$$v(t,x) = \mathbb{P}\Big(\max_{1 \le k \le N_t} X_k(t) \le x \Big) \tag{3.4.50}$$

satisfies (3.4.3)

$$\begin{aligned} &v_t = v_{xx} - v + g(v), \ t > 0, \ x \in \mathbb{R}, \\ &v(0,x) = \mathbb{1}(x \ge 0). \end{aligned} \tag{3.4.51}$$

We know from Theorem 3.3.3 that if we fix $m(t)$ by

$$m(t) = c_* t - \frac{3}{2} \log(t+1) \tag{3.4.52}$$

then there exists x_0 so that

$$v(t, x + m(t)) \to \omega(x + x_0) \ \text{ as } t \to +\infty, \text{ uniformly on } \mathbb{R}. \tag{3.4.53}$$

Let us now set

$$M(t) = \max_{1 \le k \le N_t} X_k(t). \tag{3.4.54}$$

Then (3.4.53) says that

$$\mathbb{P}\Big(M(t) - m(t) \le x \Big) \to \omega(x + x_0) \ \text{ as } t \to +\infty, \text{ uniformly on } \mathbb{R}. \tag{3.4.55}$$

In order to compare to (4.11.6) and see the Gumbel distribution, we go back to the traveling wave representation (3.4.41), which gives

$$\mathbb{P}\Big(M(t) - m(t) \le x\Big) \to \mathbb{E}\exp\big(-e^{-\lambda_*(x+x_0)}Z_\infty\big), \quad \text{as } t \to +\infty, \text{ uniformly on } \mathbb{R}. \tag{3.4.56}$$

Comparing it to (4.11.6), we see that the probability distribution of the maximum of BBM in the reference frame $m(t)$ is a "randomized Gumbel distribution" that is re-centered by a random shift

$$x \to x + x_0 - \frac{1}{\lambda_*}\log Z_\infty. \tag{3.4.57}$$

In other words, if we set

$$L_\infty = \frac{1}{\lambda_*}\log(Z_\infty) - x_0, \tag{3.4.58}$$

then we can write (3.4.56) as a random shift of (3.4.48):

$$\mathbb{P}\Big(M(t) - m(t) \le x\Big) \to \mathbb{E}\exp\big(-e^{-\lambda_*(x-L_\infty)}\big), \quad \text{as } t \to +\infty, \text{ uniformly on } \mathbb{R}. \tag{3.4.59}$$

The law of the random shift L_∞ is determined by (3.4.58) and the Laplace transform of Z_∞ in (3.4.42). We should also mention that the Bramson centering

$$m(t) = c_* t - \frac{3}{2\lambda_*}\log(t+1), \tag{3.4.60}$$

for the location $m(t)$ is different from (4.11.7) for the independent Gaussians – the pre-factor in front of $\log t$ changes from $1/2$ to $3/2$. The reason is the slowdown of BBM relative to independent Gaussians because the BBM particles are correlated. The intuitive reason for the random shift L_∞ is different: it is accumulated from the randomness at the initial stage of the branching process.

3.4.5 The ergodic theorem and the limit of the derivative martingale

As we have seen, Bramson's seminal work [32, 33] shows that the maximum of the binary BBM has the asymptotics

$$M_t = 2t - \frac{3}{2}\log t + \mathcal{O}_{\mathbb{P}}(1), \tag{3.4.61}$$

where $\mathcal{O}_{\mathbb{P}}(1)$ denotes a tight random variable depending on t. At long times, this tight remainder can be expressed as the sum of a time-independent random shift and an asymptotically stationary mean-zero fluctuation. Lalley and Sellke [81] first showed the essence of this decomposition, and asymptotic stationary can be traced to the convergence of the extremal point process due independently to Arguin, Bovier, and Kistler [10] and Aïdékon, Berestycki, Brunet, and Shi [1].

We are interested in the constant-in-time random shift, which determines the persistent "bias" of the BBM. One can isolate this shift by averaging over time to suppress the stationary fluctuation. Flath recently established the following ergodic theorem [48]: almost surely,

$$\frac{1}{t}\int_0^t \Big[M_s - 2s + \frac{3}{2}\log s\Big]\, \mathrm{d}s \to \log Z_\infty + \mathfrak{c} \quad \text{as } t \to \infty. \tag{3.4.62}$$

We note that (3.4.62) is an integrated form of an ergodic theorem conjectured in [81] and established by Arguin, Bovier, and Kistler [11]. They have shown that

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T \mathbb{1}\Big[M_s(\theta) - 2s + \frac{3}{2}\log s \le x\Big] ds \to \exp\big(-CZ_\infty e^{-x}\big) \quad \text{a.s., for any } x \in \mathbb{R}. \tag{3.4.63}$$

Thus, one can think of Z_∞ is the persistent random shift of the maximum of BBM. It is associated with the randomness that happens at the early stages of BBM. In that sense, it is not unlike the random N_∞ that we have seen back in Proposition 2.3.3. This is, if you wish, the "physical meaning" of the limit of the derivative martingale.

3.5 Anisotropy in higher dimensions

Recently, there has been a surge of interest in the extremal particles of branching Brownian motion in multiple spatial dimensions [88, 108, 70, 28, 71]. In this spirit, we will discuss in this section a fundamentally multi-dimensional phenomenon: the angular variation in the BBM frontier. We will show that when $d \geq 2$, BBM advances considerably farther in some directions than in others. This material is based on [59].

For the moment, fix a direction $\theta \in S^{d-1}$. Then $\theta \cdot X$ is the projection of the branching Brownian motion in $\mathbb{R}^d$ onto direction θ. It agrees in law with a one-dimensional BBM. We are interested in extremal behavior, so consider the maximum

$$M_t(\theta) := \max\{\theta \cdot X_t(p) : p \in \mathcal{P}_t\}.$$

Geometrically, $M_t(\theta)$ is the displacement of the supporting hyperplane in direction θ of the point set $\{X_t(p)\}_{p\in\mathcal{P}_t}$.

Of course, in each individual direction we have the Bramson asymptotics

$$M_t(\theta) = 2t - \frac{3}{2}\log t + \mathcal{O}_{\mathbb{P}}(1), \tag{3.5.1}$$

where $\mathcal{O}_{\mathbb{P}}(1)$ denotes a tight random variable depending on t and θ. The ergodic theorem [48] shows that, almost surely, we have

$$\frac{1}{t}\int_0^t \left[M_s(\theta) - 2s + \frac{3}{2}\log s\right] \mathrm{d}s \to \log Z_\infty(\theta) + \mathfrak{c} \quad \text{as } t \to \infty. \tag{3.5.2}$$

Here $\mathfrak{c}$ is a deterministic constant and $Z_\infty(\theta)$ is the almost-surely positive limit of the critical derivative martingale in direction θ:

$$Z_t(\theta) := \sum_{p\in\mathcal{P}_t} [2t - \theta \cdot X_t(p)]\mathrm{e}^{-[2t-\theta\cdot X_t(p)]}. \tag{3.5.3}$$

Note that (3.5.2) holds for each fixed θ due to equality in law with the projected process $\theta \cdot X$. It then follows from Fubini's theorem that with probability one, there exists a random full-measure subset of $\mathbb{S}^{d-1}$ within which (3.5.2) holds for all θ. We can likewise ensure that Z_t converges to Z_∞ in the same subset.

Due to (3.5.2), the derivative martingale limit $Z_\infty(\theta)$ determines the persistent bias of the BBM frontier in direction θ. That is, the size and regularity of Z_∞ governs the relative advance of the branching process in different directions. Stasiński, J. Berestycki, and Mallein have shown that $Z_\infty \in L^1(S^{d-1})$ almost surely [108]. Our main result in [59] shows that that Z_∞ is not substantially more regular.

Theorem 3.5.1. *If $d \geq 3$, then almost surely, for every nonempty open $U \subset S^{d-1}$, $\|Z_\infty\|_{L^p(U)} = \infty$ for all $p \geq \frac{d-1}{d-2}$. If $d = 2$, the same holds for $p = \infty$.*

The sharpness of the threshold $p = \frac{d-1}{d-2}$ remains an interesting open question.

Taking $p = \infty$, Theorem 3.5.1 implies that Z_∞ is nowhere locally bounded in any dimension $d \geq 2$. Thus in any open set of directions, BBM advances arbitrarily farther in one direction relative to another:

Corollary 3.5.2. *Let $d \geq 2$. Almost surely, for each $L, \varepsilon > 0$ and nonempty open $U \subset S^{d-1}$, there exist $\theta_1, \theta_2 \in U$ such that the long-term fraction of time during which $M_t(\theta_1) \geq M_t(\theta_2) + L$ is at least $1 - \varepsilon$. That is,*

$$\liminf_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{1}(M_s(\theta_1) \geq M_s(\theta_2) + L) \, \mathrm{d}s \geq 1 - \varepsilon. \tag{3.5.4}$$

In this sense, the BBM frontier is quite irregular.

Our approach builds on the following intuition: if a particle makes an unusual radial excursion, its descendants impart a permanent boost to Z_t in nearby directions. To prove Theorem 3.5.1, we show that the BBM makes suitably large excursions in a dense set of directions. We take inspiration from the work of Maillard and Pain [87], who used such excursions in one dimension to characterize the $\mathcal{O}(t^{-1/2})$ fluctuations of Z_t. In higher dimensions, there are more directions and thus more excursions, so this effect contributes at leading rather than vanishing order.

We are interested in particles that travel well beyond the typical frontier. By (3.5.1), in any fixed direction, the BBM frontier is typically at distance $2t - \frac{3}{2}\log t$ from the origin. However, Mallein [88] showed that there are random, time-varying directions in which particles are considerably farther, located at distance $2t + \frac{d-4}{2}\log t$. If we wait for an exceptional *time*, particles will travel farther still. Indeed in one dimension, Hu and Shi [69] and Aïdékon and Shi [2] showed that particles occasionally travel $\log t$ beyond the typical frontier.

This suggests that in higher dimensions, we should occasionally find particles at the radius

$$2t + \frac{d-4}{2}\log t + \log t = 2t + \frac{d-2}{2}\log t. \tag{3.5.5}$$

As a start, it is relatively straightforward to show that particles occasionally reach $2t + \alpha \log t$ for any $\alpha < \frac{d-2}{2}$. These particles' contributions to Z_∞ are sufficiently large to prove Theorem 3.5.1 when $d \geq 3$ and $p > \frac{d-1}{d-2}$. However, to reach the critical dimension 2 (or the threshold $p = \frac{d-1}{d-2}$), we must work considerably harder. We show that particles do reach the radius in (3.5.5), and in fact go slightly farther. This amounts to a law of the iterated logarithm for branching Brownian motion in multiple dimensions. To state it, we introduce the notation $\big(R_t(p), \Theta_t(p)\big)$ for the radial and angular coordinates of the position $X_t(p)$ of particle p at time t.

Theorem 3.5.3. *Let $\beta \in \mathcal{C}^2\big([0,\infty); \mathbb{R}\big)$ satisfy $t^2\ddot{\beta}_t \to 0$ and $\dot{\beta}_t \to 0$ as $t \to \infty$. If*

$$\int_1^\infty \frac{\mathrm{d}r}{r \exp \beta_r} = \infty, \tag{3.5.6}$$

then almost surely, for every nonempty open $U \subset S^{d-1}$, there exist sequences $t_n \nearrow \infty$ and $p_n \in \mathcal{P}_{t_n}$ such that for all $n \in \mathbb{N}$,

$$R_{t_n}(p_n) = 2t_n + \frac{d-2}{2}\log t_n + \beta_{t_n} \quad \text{and} \quad \Theta_{t_n}(p_n) \in U. \tag{3.5.7}$$

If (3.5.6) does not hold, then almost surely there exists $\tau > 0$ such that for all $t > \tau$,

$$\max_{p\in\mathcal{P}_t} R_t(p) < 2t + \frac{d-2}{2}\log t + \beta_t.$$

The criterion (3.5.6) was identified by Hu [68] for 1D branching random walks, and we extend it here to BBM in $\mathbb{R}^d$. The choice $\beta_t = \log\log t$ satisfies (3.5.6) and allows us to prove the borderline cases of Theorem 3.5.1. We once again emphasize that the fine result in Theorem 3.5.3 is only necessary

to treat $d = 2$ or $p = \frac{d-1}{d-2}$. When $d \geq 3$ and $p > \frac{d-1}{d-2}$, much shorter arguments suffice, as discussed after (3.6.17) and in Remark 3.7.4 below.

Previous works on exceptional excursions [69, 2, 68] have largely relied on spine decompositions [84] and the second moment method. We take a different approach, by expressing probabilities of interest in terms of solutions to the Fisher–KPP equation, a deterministic parabolic PDE. As we have seen, this strategy is well known in one dimension, and led to Bramson's result in (3.5.1). However, the connection between branching processes and PDEs has received little attention in multiple dimensions. To our knowledge, this is the first work that uses intrinsically multi-dimensional Fisher–KPP solutions to study branching Brownian motion. Our proof of Theorem 3.5.3 combines such solutions with results on the derivative martingale from [28]. We thus take a hybrid approach, drawing on both PDE constructions and purely probabilistic estimates to produce large excursions.

The duality between BBM and the Fisher–KPP equation involves a time reversal, so the long-time behavior of BBM is tied to ancient solutions of Fisher–KPP. In order to treat arbitrary open sets of directions as in Theorem 3.5.1, these solutions are necessarily anisotropic and do not have a direct one-dimensional analog. For example, our proof of the localization of mass in Z_t rests on a detailed analysis of the distinct behavior of a Fisher–KPP solution in one direction relative to others. This line of inquiry bears some resemblance to PDE literature such as constructions of conical traveling waves in [61, 62] and the partial classification of entire solutions to Fisher–KPP in [63], but the connection is indirect.

3.6 Proof outline: excursions and the derivative martingale

In this section, we explain the structure of our argument and prove Theorem 3.5.1 and Corollary 3.5.2, assuming certain auxiliary results that we defer to later sections. The proof rests on the observation that large radial excursions produce large values in Z_∞ with high probability. We now offer an informal explanation of why this is the case.

Define a radial threshold

$$b_t := 2t + \frac{d-2}{2} \log t + \beta_t \tag{3.6.1}$$

fine-tuned by a slowly varying function $t \mapsto \beta_t$ satisfying the hypotheses of Theorem 3.5.3 and (3.5.6). We will ultimately take $\beta_t = \log\log t$, but for the moment we view it as a (relatively) free parameter. Given a nonempty open set of directions $U \subset S^{d-1}$, Theorem 3.5.3 ensures that particles cross the radial threshold b_t infinitely often through the sector U. In particular, if we fix a large time $T \gg 1$, then almost surely there exists a random earliest time $\hat{t} \geq T$ and a particle $\hat{p} \in \mathcal{P}_{\hat{t}}$ making a large excursion in U:

$$R_{\hat{t}}(\hat{p}) = b_{\hat{t}} = 2\hat{t} + \frac{d-2}{2} \log \hat{t} + \beta_{\hat{t}} \quad \text{and} \quad \Theta_{\hat{t}}(\hat{p}) \in U. \tag{3.6.2}$$

Now recall the derivative martingale $Z_t(\theta)$ defined in (3.5.3), and its additive analog

$$W_t(\theta) := \sum_{p \in \mathcal{P}_t} \mathrm{e}^{-[2t - \theta \cdot X_t(p)]}. \tag{3.6.3}$$

We also recall (3.4.43):

$$\sqrt{t} W_t(\theta) \to C Z_\infty(\theta) \text{ in probability as } t \to \infty. \tag{3.6.4}$$

As we have mentioned, this suggests that the derivative martingale Z_t in (3.5.3) is dominated by particles roughly $\sqrt{t}$ behind the BBM frontier. This provides considerable leeway in analyzing Z_t: we can drop particles outside this window (as in Y^{win} in (3.7.22) below) and modify the prefactor

by $o(\sqrt{t})$ (as in Y and $\hat{Y}$ below) without changing the long-time limit. We make frequent use of this flexibility, routinely modifying Z_t to produce convenient properties while preserving its essential nature. Such modifications are not true martingales, but this plays no role our analysis.

As a first example, we work with a variant of Z_t similar to one introduced in [28]. Define the random measure

$$Y_t(\theta) := \pi^{\frac{d-1}{4}} \sum_{p \in \mathcal{P}_t} ([b_t - R_t(p)]_+ + 1) \mathrm{e}^{-[2t + \frac{d-1}{2} \log t - R_t(p)]} \delta_{\Theta_t(p)}(\theta). \tag{3.6.5}$$

The intuition is as follows: if we integrate $Z_t(\theta)$ from (3.5.3) against a smooth function of θ, we can use Laplace's method to identify the asymptotic contribution of a given particle to the integral. This produces a factor of $(\sqrt{\pi}/t)^{(d-1)/2}$, which appears in the point measure (3.6.5). Thus (3.5.3) and (3.6.5) behave similarly when integrated against test functions. In addition, we have modified the linear prefactor in a manner that preserves the leading contribution of particles roughly $\sqrt{t}$ behind the frontier. As discussed above, this change has a negligible effect at long times:

Theorem 3.6.1 ([28]). *As $t \to \infty$, the measure $Y_t(\theta)$ converges weakly in probability to the limit $Z_\infty(\theta)$ of the projected derivative martingale in (3.5.3).*

This is a minor variation on Theorem 1.4 in [28].

We wish to show that $Y_t(\theta)$ is consistently large in certain directions. Because its summands are nonnegative, it suffices to consider only those p that descend from the particle $\hat{p}$ defined above. We let $\hat{\mathcal{P}}_t$ denote this subpopulation and introduce

$$\hat{Y}_t(\theta) := \pi^{\frac{d-1}{4}} \sum_{p \in \hat{\mathcal{P}}_t} ([\hat{b}_t - R_t(p)]_+ + 1) \mathrm{e}^{-[\hat{b}_t - R_t(p)]} \delta_{\Theta_t(p)}(\theta) \quad \text{for } t \geq \hat{t}, \tag{3.6.6}$$

where

$$\hat{b}_t := 2t + \frac{d-1}{2} \log t - \frac{1}{2} \log \hat{t} + \beta_{\hat{t}} = b_t + \frac{1}{2} \log t - \frac{1}{2} \log \hat{t} + \beta_{\hat{t}} - \beta_t. \tag{3.6.7}$$

Note that the curves b and $\hat{b}$ agree at time $\hat{t}$. The exponent in (3.6.6) differs from that in (3.6.5) by $-\frac{1}{2} \log \hat{t} + \beta_{\hat{t}}$, which is chosen so that $\hat{Y}_t$ has mass of order 1 at its initial time $\hat{t}$. The difference in the linear prefactor in (3.6.6) plays a negligible role, so

$$Y_t \geq \hat{t}^{-1/2} \mathrm{e}^{\beta_{\hat{t}}} \hat{Y}_t + \mathcal{O}_{t \to \infty}(1). \tag{3.6.8}$$

At time $\hat{t}$, the set $\hat{\mathcal{P}}_{\hat{t}}$ consists of the single particle $\hat{p}$, and $R_{\hat{t}}(\hat{p}) = b_{\hat{t}} = \hat{b}_{\hat{t}}$. Hence $\hat{Y}_{\hat{t}}(S^{d-1}) = \pi^{(d-1)/4}$. By design, both Y and $\hat{Y}$ are approximately martingales. We thus expect the total mass of $\hat{Y}$ to remain bounded away from zero. To express this carefully, let $\mathcal{G}$ denote the σ-algebra containing all information about particles until they first cross the threshold b_t through U after time T.

Lemma 3.6.2. *For all $\varepsilon > 0$, there exists $m > 0$ such that for all $t \geq \hat{t}$,*

$$\mathbb{P}\big[\hat{Y}_t(S^{d-1}) \geq m \mid \mathcal{G}\big] \geq 1 - \varepsilon.$$

It turns out that this statement can be expressed in terms of the Laplace transform of $\hat{Y}_t$, which we analyze using a radially-symmetric solution of the Fisher–KPP equation.

Due to (3.6.8), this bound implies that

$$Y_t(S^{d-1}) \gtrsim \hat{t}^{-1/2} \mathrm{e}^{\beta_{\hat{t}}} \sim \frac{\log \hat{t}}{\hat{t}^{1/2}}, \tag{3.6.9}$$

which does *not* suffice to show that $\|Z_\infty\|_{L^\infty} = \infty$. Indeed, no argument based on total mass alone can work, for [108] showed that $\|Z_\infty\|_{L^1} < \infty$ a.s. Thus, to show that the L^∞ norm diverges, we must argue that some of the mass of Z_∞ remains localized in a sufficiently small region around the direction $\hat{\Theta}_{\hat{t}}(\hat{p})$. This is plausible, because $\hat{Y}_t$ is dominated by particles near the BBM frontier at time t. In order to reach the frontier from the location of their parent $\hat{p}$, they cannot afford to move much in the angular coordinate. In fact, $\hat{Y}_t$ remains localized to scale $\hat{t}^{-1/2}$ around the angle $\hat{\theta} := \hat{\Theta}_{\hat{t}}(\hat{p}) \in U$.

Proposition 3.6.3. *Let Γ denote a ball in S^{d-1} centered at $\hat{\theta}$. For all $K > 0$, there exist $A(K)$ and $C(d, K) > 0$ such that if Γ has radius $A\hat{t}^{-1/2}\log^{1/2}\hat{t}$, then for all $T \geq C(d, K)$ and $t \geq \hat{t}$,*

$$\mathbb{P}\big[\hat{Y}_t(\Gamma^{\mathrm{c}}) \geq \hat{t}^{-K} \mid \mathcal{G}\big] \leq T^{-K}. \tag{3.6.10}$$

We show this localization using a Fisher–KPP solution resembling a radially inbound traveling wave that is missing mass in the angular sector Γ.

We now have the pieces in place to show that Z_∞ has large values in U. We use the notation $f \gtrsim g$ to indicate that $f \geq cg$ for some constant $c > 0$, and $f \asymp g$ indicates that $f \gtrsim g$ and $g \gtrsim f$. Combining Lemma 3.6.2 and Proposition 3.6.3, we see that

$$\hat{Y}_t(\Gamma) \gtrsim 1, \tag{3.6.11}$$

for a random neighborhood Γ of volume

$$\mathrm{vol}_{S^{d-1}}(\Gamma) \asymp (\log\hat{t})^{\frac{d-1}{2}}\hat{t}^{-\frac{d-1}{2}}. \tag{3.6.12}$$

By (3.6.8), we also know that

$$Y_t(\Gamma) \gtrsim \hat{t}^{-1/2}\mathrm{e}^{\beta_{\hat{t}}}\hat{Y}_t(\Gamma) \gtrsim \hat{t}^{-1/2}\mathrm{e}^{\beta_{\hat{t}}}. \tag{3.6.13}$$

Combining (3.6.12) and (3.6.13), we see that the *mean* of Y_t on Γ satisfies

$$\mathrm{vol}_{S^{d-1}}(\Gamma)^{-1}Y_t(\Gamma) \gtrsim \mathrm{e}^{\beta_{\hat{t}}}(\log\hat{t})^{-\frac{d-1}{2}}\hat{t}^{\frac{d-2}{2}}.$$

Sending $t \to \infty$ and using Theorem 3.6.1, the same holds for Z_∞. The mean is a lower bound on the essential supremum and $\Gamma \subset U$, so

$$\|Z_\infty\|_{L^\infty(U)} \gtrsim \mathrm{e}^{\beta_{\hat{t}}}(\log\hat{t})^{-\frac{d-1}{2}}\hat{t}^{\frac{d-2}{2}}. \tag{3.6.14}$$

If $d \geq 3$, it suffices to take

$$\beta_t = -\log\log t, \tag{3.6.15}$$

and in $d = 2$ we can take

$$\beta_t = \log\log t, \tag{3.6.16}$$

to ensure that in both cases, the right side of (3.6.14) is increasing in $\hat{t}$. Since $\hat{t}$ is the earliest excursion time after T, we have $\hat{t} \geq T$. This yields

$$\|Z_\infty\|_{L^\infty(U)} \gtrsim \mathrm{e}^{\beta_T}(\log T)^{-\frac{d-1}{2}}T^{\frac{d-2}{2}}. \tag{3.6.17}$$

With the above choices of β_t, we can send $T \to +\infty$ to conclude that

$$\|Z_\infty\|_{L^\infty(U)} = \infty.$$

The difference in sign between the extra drifts in $d \geq 3$ (3.6.15) and $d = 2$ (3.6.15) means that we require the full strength of Theorem 3.5.3 in $d = 2$. This greatly complicates the paper, as it is much more difficult to produce the excursions in (3.5.7) when $\beta_t = +\log\log t$ rather than $-\log\log t$. We explain this added difficulty in Remark 3.7.4 below.

3.7 Large excursions

In this section, we explain the proof of the first part of Theorem 3.5.3, which states that particles cross the radial threshold b_t defined in (3.6.1) infinitely often through every nonempty open sector $U \subset S^{d-1}$, provided β_t satisfies (3.5.6). We assume β satisfies the regularity condition

$$|\ddot{\beta}_t| \ll t^{-2} \tag{3.7.1}$$

in Theorem 3.5.3, as well as $\dot{\beta}_t \to 0$ as $t \to \infty$. Integrating twice, these imply that

$$|\beta_t| \ll \log t \quad \text{and} \quad \left|\dot{\beta}_t\right| \ll t^{-1}. \tag{3.7.2}$$

3.7.1 A connection to the Fisher–KPP equation

We first obtain a Fisher–KPP representation of these excursions. Fix times $T < S$ and define the event

$$E_{T,S}(U) := \big\{\exists \tau \in [T,S], p \in \mathcal{P}_\tau : R_\tau(p) = b_\tau, \Theta_\tau(p) \in U\big\}. \tag{3.7.3}$$

We are interested in the union

$$E_T(U) := \bigcup_{S \geq T} E_{T,S}(U), \tag{3.7.4}$$

which captures excursions after time T. We can write its probability as

$$\overline{h}_T(t,x) := \mathbb{P}_{t,x}[E_T(U)] = \lim_{S \to \infty} \mathbb{P}_{t,x}[E_{T,S}(U)], \tag{3.7.5}$$

where $\mathbb{P}_{t,x}$ denotes the law of a BBM started from (t,x). We also define the set

$$\mathcal{A}(t) = \{(R,\theta) \in \mathbb{R}^d : R = b_t, \theta \in U\}, \tag{3.7.6}$$

with the convention that

$$\mathcal{A}(t) = \emptyset \quad \text{for } t < T. \tag{3.7.7}$$

Lemma 3.7.1. *The function $\overline{h}_T(t,x)$ from (3.7.5) is the minimal positive solution to the global-in-time backward problem*

$$\begin{cases} \partial_t \overline{h}_T + \Delta \overline{h}_T + \overline{h}_T - \overline{h}_T^2 = 0, & t \in \mathbb{R}, \quad x \in \mathcal{A}^{\mathrm{c}}(t), \\ \overline{h}_T(t,x) = 1, & t > T, \ x \in \mathcal{A}(t). \end{cases} \tag{3.7.8}$$

This is similar to Lemma A.1 in [86] in the one-dimensional setting.

One important observation is that in one dimension the McKean argument coupled with a time-reversal symmetry allowed us to avoid dealing with backward in time solutions to the Fisher-KPP equation. In higher dimensions, this seems inevitable. In addition, in higher dimensions we have a rich family of ancient solutions that look like traveling waves coming from infinity with "angular holes"

Proof. Given $t < S$ and $x \in \mathcal{A}^{\mathrm{c}}(t)$, let

$$h_{T,S}(t,x) := \mathbb{P}_{t,x}[E_{T,S}(U)] \tag{3.7.9}$$

be the probability that a BBM started at time t and position x crosses the threshold $\mathcal{A}(\tau)$ at some time $\tau \in [T,S]$. Let

$$z_{T,S}(t,x) := 1 - h_{T,S}(t,x) = \mathbb{P}_{t,x}[E^{\mathrm{c}}_{T,S}(U)]. \tag{3.7.10}$$

The branching property implies the following renewal identity for any $\tau \in (t, T)$:

$$\begin{aligned} z_{T,S}(t,x) &= \mathrm{e}^{-(\tau-t)}\mathbb{E}[z_{T,S}(\tau, x + B_{\tau-t})] + \int_t^\tau \mathrm{e}^{-(s-t)}\mathbb{E}[z_{T,S}^2(s, x + B_{s-t})]\,\mathrm{d}s \\ &= \mathrm{e}^{(\tau-t)(\Delta-1)}[z_{T,S}(\tau,\,\cdot\,)](x) + \int_t^\tau \mathrm{e}^{(s-t)(\Delta-1)}[z_{T,S}^2(s,\,\cdot\,)](x)\,\mathrm{d}s. \end{aligned} \tag{3.7.11}$$

Here, B_s is a Brownian motion with diffusivity $\sqrt{2}\,\mathrm{Id}$ that starts at time $t < S$ and position $x \in \mathbb{R}^d$, and dies if it reaches the moving boundary $\mathcal{A}$ during the time interval $[T, S]$. If $t < s < T$, there is no killing.

The first term on the right side of (3.7.11) corresponds to the event that the initial particle does not branch in the time interval $[t, \tau]$, which has probability $\mathrm{e}^{-(\tau-t)}$. The second term accounts for a first branching at time $s \in [t, \tau]$. In this case, the descendants of each child must avoid the threshold, giving $z_{T,S}^2$ by independence.

The identity (3.7.11) is the mild form of the backward-in-time Fisher–KPP equation

$$\partial_t z_{T,S} + \Delta z_{T,S} - z_{T,S} + z_{T,S}^2 = 0, \quad t < S,\, x \in \mathcal{A}^{\mathrm{c}}(t). \tag{3.7.12}$$

In addition, we have the terminal condition

$$z_{T,S}(S, x) = 1 \quad \text{for } x \in \mathcal{A}^{\mathrm{c}}(S) \tag{3.7.13}$$

and the boundary condition

$$z_{T,S}(t, x) = 0 \quad \text{for } T \le t \le S, |x| = b_t, \theta \in U. \tag{3.7.14}$$

This boundary condition disappears at time T, leaving $z_{T,S}$ to solve Fisher–KPP in the whole space when $t < T$. One can view this as a whole-space problem for $t < T$ with terminal data $z_{T,S}(T, x)$ at $t = T$.

Using (3.7.10), $h_{T,S}$ now satisfies

$$\begin{cases} \partial_t h_{T,S} + \Delta h_{T,S} + h_{T,S} - h_{T,S}^2 = 0 & t < S,\ x \in \mathcal{A}^{\mathrm{c}}(t), \\ h_{T,S}(S, x) = 0, & x \in \mathcal{A}^{\mathrm{c}}(S), \\ h_{T,S}(t, x) = 1, & x \in \mathcal{A}(t),\ T < t < S. \end{cases} \tag{3.7.15}$$

By its definition (3.7.9), the family $(h_{T,S})_{S \ge T}$ is increasing in S. By (3.7.5), its limit as $S \to \infty$ is $\overline{h}_T(t, x)$. It follows from parabolic estimates that $\overline{h}_T(t, x)$ solves the global-in-time backward problem (3.7.8), recalling the convention (3.7.7). Moreover, the terminal condition in (3.7.15) implies that for every $S > T$, $h_{T,S}$ lies below every global-in-time positive solution of (3.7.8). It follows that the limit $\overline{h}_T$ is the least such solution. □

3.7.2 A Liouville theorem for the ancient solution

Assuming (3.5.6), we wish to show that if β_t satisfies assumptions (3.7.2), then for any $T \ge 2$, the minimal positive entire solution of (3.7.8) is

$$\overline{h}_T(t, x) \equiv 1. \tag{3.7.16}$$

We find it convenient to reverse time in (3.7.8) by setting

$$h_T(t, x) := \overline{h}_T(-t, x). \tag{3.7.17}$$

This is an ancient solution of the forward-in-time Fisher–KPP problem

$$\begin{cases} \partial_t h_T = \Delta h_T + h_T - h_T^2, & t \in \mathbb{R}, \quad x \in \mathcal{A}^{\mathrm{c}}(-t) \\ h_T(t,x) = 1, & t < -T, \ x \in \mathcal{A}(-t). \end{cases} \tag{3.7.18}$$

The key estimate in the proof of (3.7.16) states that far in the past, $h \gtrsim x\mathrm{e}^{-x}$ relative to $b_{|t|}$.

Proposition 3.7.2. *Let β_t satisfy (3.5.6). Then for every nonempty open $U' \Subset U$, there exists $t_{\max}(d,U') \le -T$ and $c(d,U') > 0$ such that*

$$h_T(t,r,\theta) \ge c[b_{|t|} - r]\mathrm{e}^{-[b_{|t|}-r]} \tag{3.7.19}$$

for all $t \le t_{\max}$, $\theta \in U'$, and $b_{|t|} - r \in [4\log|t|, |t|/8]$.

The proof of Proposition 3.7.2 is rather technical, and we refer to [59] for details. We only mention that it is based on purely PDE arguments.

Here, we use a result of [28] to show that Proposition 3.7.2 forces h_T to be 1. The key element of the estimate (3.7.19) is the linear pre-factor $b_{|t|} - r$, which enables a connection with the approximate derivative martingale Y_t.

Proposition 3.7.3. *For all $T \ge 2$, $h_T \equiv 1$.*

Proof. By definition $0 \le h_T \le 1$, so the strong maximum principle for the Fisher–KPP equation implies that it suffices to show that

$$h_T(0,0) = 1. \tag{3.7.20}$$

Indeed, this would imply that $h_T(t,x) = 1$ for all $t < 0$ and $x \in \mathbb{R}^d$, which would immediately entail that $h_T(t,x) = 1$ for all $t > 0$ as well.

Define the "windowed" BBM population

$$\mathcal{P}_t^{\mathrm{win}} := \{p \in \mathcal{P}_t : t^{1/6} \le 2t - R_t(p) \le t^{2/3}\} \tag{3.7.21}$$

and the windowed derivative martingale

$$Y_t^{\mathrm{win}} := \left(\frac{\sqrt{\pi}}{t}\right)^{\frac{d-1}{2}} \sum_{p \in \mathcal{P}_t^{\mathrm{win}}} [2t - R_t(p)]\mathrm{e}^{-[2t-R_t(p)]}\delta_{\Theta_p(t)}. \tag{3.7.22}$$

By McKean's formula, for all $t \ge 0$, we have the representation

$$\begin{aligned} h_T(0,0) &= 1 - \mathbb{E}_{-t,0}\prod_{p\in\mathcal{P}_0}\big[1 - h\big(-t, X_0(p)\big)\big] = 1 - \mathbb{E}_{0,0}\prod_{p\in\mathcal{P}_t}\big[1 - h\big(-t, X_t(p)\big)\big] \\ &\ge 1 - \mathbb{E}\exp\Big(\sum_{p\in\mathcal{P}_t^{\mathrm{win}}}\log\big[1 - h\big(-t, X_t(p)\big)\big]\Big). \end{aligned} \tag{3.7.23}$$

Here we use the shorthand $\mathbb{E}$ for $\mathbb{E}_{0,0}$. It follows from (3.6.1) and (3.7.1) that there exists T_1 such that for all $|t| > T_1$,

$$b_{|t|} - r \in [4\log|t|, |t|/8] \quad \text{if } 2|t| - r \in [|t|^{1/6}, |t|^{2/3}]. \tag{3.7.24}$$

Fix a nonempty open $U' \Subset U$. Then Proposition 3.7.2, (3.7.23), and (3.7.24) yield

$$h_T(0,0) \ge 1 - \mathbb{E}\exp\Big(\sum_{p\in\mathcal{P}_t^{\mathrm{win}}}\log\big\{1 - c[b_t - R_t(p)]\mathrm{e}^{-[b_t-R_t(p)]}\mathbf{1}_{U'}\big(\Theta_p(t)\big)\big\}\Big) \tag{3.7.25}$$

for all $t \geq \max\{-t_{\max}(d, U'), T_1\}$.

Note that, simply from (3.6.1), we can re-write the exponential in (3.7.25) as

$$\mathrm{e}^{-[b_t - R_t(p)]} = t^{-(d-2)/2}\mathrm{e}^{-\beta_t}\mathrm{e}^{-[2t - R_t(p)]}. \tag{3.7.26}$$

For the windowed population, $2t - R_t(p) \geq t^{1/6}$ while $|b_t - 2t| \lesssim \log t$. It follows that the pre-factor in (3.7.25) can be bounded as

$$b_t - R_t(p) \geq \frac{2t - R_t(p)}{2} \quad \text{for } p \in \mathcal{P}_t^{\mathrm{win}}. \tag{3.7.27}$$

Using (3.7.26) and (3.7.27) in (3.7.25) shows that $h_T(0,0)$ is bounded from below by

$$1 - \mathbb{E}\exp\Big(\sum_{p\in\mathcal{P}_t^{\mathrm{win}}} \log\big[1 - \tfrac{c}{2}t^{-\frac{d-2}{2}}\mathrm{e}^{-\beta_t}[2t - R_t(p)]\mathrm{e}^{-[2t-R_t(p)]}\mathbf{1}_{U'}\big(\Theta_p(t)\big)\big]\Big).$$

Recall the windowed version Y_t^{win} of the derivative martingale defined in (3.6.5). Because $\log(1-a) \leq -a$, we obtain

$$\begin{aligned} h_T(0,0) &\geq 1 - \mathbb{E}\exp\Big(-\tfrac{c}{2}t^{-\frac{d-2}{2}}\mathrm{e}^{-\beta_t}\sum_{p\in\mathcal{P}_t^{\mathrm{win}}}[2t - R_t(p)]\mathrm{e}^{-[2t-R_t(p)]}\mathbf{1}_{U'}\big(\Theta_p(t)\big)\Big) \\ &\geq 1 - \mathbb{E}\exp\Big[-\frac{c}{2\pi^{(d-1)/4}}t^{1/2}\mathrm{e}^{-\beta_t}Y_t^{\mathrm{win}}(U')\Big]. \end{aligned} \tag{3.7.28}$$

for all sufficiently large t.

Now, Theorem 1.4 of [28] states that $Y_t^{\mathrm{win}} \to Z_\infty$ a.s. in the vague topology as $t \to \infty$, and Z_∞ is a.s. positive on nonempty open sets. The portmanteau lemma implies that

$$\liminf_{t\to\infty} Y_t^{\mathrm{win}}(U') \geq Z_\infty(U') > 0 \quad \text{a.s.}$$

By (3.7.2), $t^{1/2}\mathrm{e}^{-\beta_t} \to \infty$ as $t \to \infty$. It follows that

$$\liminf_{t\to\infty} \frac{c}{2\pi^{(d-1)/4}}t^{1/2}\mathrm{e}^{-\beta_t}Y_t^{\mathrm{win}}(U') = \infty \quad \text{a.s.}$$

Sending $t \to \infty$ in the right side of (3.7.28), we see that $h_T(0,0) \geq 1$. Of course we also know that $h_T \leq 1$, whence $h_T(0,0) = 1$, which is (3.7.20). □

Remark 3.7.4. We can now see why the "subcritical" regime $d \geq 3$ and $p > \frac{d-1}{d-2}$ permits a much shorter proof. As discussed after (3.6.17) for $p = \infty$, subcriticality allows us to use $\beta_t = -\log\log t$ in the threshold (3.6.1). This relaxation allows us to omit the linear pre-factor in (3.7.19). Indeed, given a pure exponential lower bound in (3.7.19), we would obtain an analogue A_t of an *additive* martingale rather than the *derivative* martingale Y_t appearing above. The Seneta–Heyde scaling [69, 2] implies that $t^{1/2}A_t \to CZ_\infty$, so the bracketed expression in (3.7.28) is then of order $\mathrm{e}^{-\beta_t}$. This diverges if $\beta_t = -\log\log t$, and the proof of Proposition 3.7.3 goes through. It is much easier to prove (3.7.19) without the linear prefactor but this part is outside the scope of these notes.

Conversely, to treat the critical regime such as $d = 2$, we must use a more stringent threshold with $\beta_t = +\log\log t$. This requires the linear prefactor in (3.7.19) to overcome the decaying factor $\mathrm{e}^{-\beta_t}$ in (3.7.28), and the prefactor demands a considerably more sophisticated argument.

We can now prove the positive part of Theorem 3.5.3.

Proof of first part of Theorem 3.5.3. Assume β_t satisfies (3.5.6) and (3.7.1). By (3.7.5) and Proposition 3.7.3,

$$\mathbb{P}\Big[\bigcap_{n\in\mathbb{N}} E_n(U)\Big] = 1$$

for $E_T(U)$ defined in (3.7.4). Working in this intersection, let t_n be the earliest time contributing to $E_n(U)$, with corresponding particle $p_n \in \mathcal{P}_{t_n}$. The random sequence $(t_n)_{n\in\mathbb{N}}$ is increasing and satisfies $t_n \geq n$, so $t_n \nearrow \infty$. Moreover, (3.5.7) holds by the definition of $E_n(U)$. □

4 Lecture 3. Bounds for thermo-diffusive systems

4.1 Thermo-diffusive systems

We will consider in this lecture the thermo-diffusive system, one of the simplest combustion models:

$$\begin{aligned} \partial_t u &= \nu\Delta u - ug(v), \\ \partial_t v &= \kappa\Delta v + ug(v). \end{aligned} \tag{4.1.1}$$

Such models also appear in, for example, biology and chemistry – see Murray's books on mathematical biology for some examples. One can consider systems with more general nonlinearities of the form

$$\begin{aligned} \partial_t u &= \nu\Delta u - f_1(u,v), \\ \partial_t v &= \kappa\Delta v + f_2(u,v), \end{aligned} \tag{4.1.2}$$

and with a greater number of unknowns but we want to keep things as simple as possible to clarify the main points.

We think of $v(t,x)$ in (4.1.1) as the fuel temperature and of $u(t,x)$ as the fuel concentration. This interpretation "implies" that the reaction rate $g(v)$ is positive:

$$g(v) \geq 0, \quad \text{for all } v \geq 0, \tag{4.1.3}$$

This assumption ensures that

$$u(t,x) > 0 \text{ and } v(t,x) > 0 \text{ for all } x \in \mathbb{R}^n \text{ and } t > 0, \tag{4.1.4}$$

as long as

$$u_0(x) \geq 0 \text{ and } v_0(x) \geq 0 \text{ for all } x \in \mathbb{R}^n. \tag{4.1.5}$$

We will always make these assumptions. Typically, we will think of the situation where $v_0(x)$ is non-negative and compactly supported, and the initial concentration satisfies

$$u_0(x) = 1, \quad \text{for all } |x| \geq R. \tag{4.1.6}$$

On the other hand, the situation when $u_0(x)$ is, say, integrable, or otherwise decays at infinity, is "much less interesting" because the most likely scenario is that all concentration will be burnt out and there would be no fuel left. This, in turn, would lead to decay of the temperature $v(t,x)$, so that both the concentration and the temperature would vanish in the long time limit. Our interest is in the opposite regime, where the concentration does not decay at infinity, and a high temperature front may spread.

4.2 A voting model for the thermo-diffusive system

We now describe a voting model that corresponds to the thermo-diffusive system with $g(v) = v$. Generalizations to other functions $g(v)$ are straightforward. Consider a branching Brownian motion where every parent produces one child of type A and one child of type B. Particles of type A perform a Brownian motion with diffusivity κ and those with type B do a Brownian motion with diffusivity ν. This is the main difference with the standard BBM that we have considered in the previous lectures. The voting on the genealogical tree is done as follows. Assume that the two children of types A and B voted ξ_A and ξ_B, respectively. If the parent is type A then its vote is

$$y_A = \xi_A + \xi_A \xi_B, \tag{4.2.1}$$

while if the parent is of type B then its vote is

$$y_B = \xi_B - \xi_A \xi_B. \tag{4.2.2}$$

Note that this voting model does not preserve positivity of the votes: even if ξ_A and ξ_B are positive, the type B parent vote y_B will be negative if $\xi_A > 1$.

Let $p(t, x, \xi)$ be the probability density of the vote of the root of the tree when the original particle is of type A, and $q(t, x, \xi)$ be the corresponding probability density when the original particle is of type B. Then, assuming that the exponential clocks have rate 1, we have the following system of nonlinear integro-differential equations

$$\begin{aligned} \frac{\partial p(t,x,\xi)}{\partial t} &= \kappa \Delta_x p(t,x,\xi) - p(t,\xi) + \int p(t,x,\eta_1) q(t,\eta_2) \delta(\xi - \eta_1 - \eta_1\eta_2) d\eta_1 d\eta_2, \\ \frac{\partial q(t,x,\xi)}{\partial t} &= \nu \Delta_x q(t,x,\xi) - q(t,\xi) + \int p(t,x,\eta_1) q(t,x,\eta_2) \delta(\xi - \eta_2 + \eta_1\eta_2) d\eta_1 d\eta_2. \end{aligned} \tag{4.2.3}$$

These solutions satisfy

$$\int p(t,x,\xi) d\xi = \int p_0(x,\xi) d\xi = 1, \quad \int q(t,x,\xi) d\xi = \int q_0(x,\xi) d\xi = 1, \quad \text{for all } t \geq 0 \text{ and } x \in \mathbb{R}^d, \tag{4.2.4}$$

as long as this normalization holds at $t = 0$. Even though these equations are nonlinear, global in time existence of the solutions to (4.2.3) follows immediately from the probabilistic interpretation.

In order to connect (4.2.3) to the thermo-diffusive system, let us define the averages

$$M(t,x) = \int \xi p(t,x,\xi) d\xi, \quad N(t,x) = \int \xi q(t,x,\xi) d\xi. \tag{4.2.5}$$

Note that, as we do not know that $p(t, x, \xi)$ or $q(t, x, \xi)$ are supported on $\xi \geq 0$, the integrals in (4.2.5) may involve a lot of cancellation, and their careful definition may be an issue in itself. Still, formally, we have

$$\begin{aligned} \frac{\partial M(t,x)}{\partial t} - \kappa \Delta_x M(t,x) + M(t,x) &= \int \eta_1 (1 + \eta_2) p(t,x,\eta_1) q(t,x,\eta_2) d\eta_1 d\eta_2 \\ &= M(t,x) + M(t,x) N(t,x). \end{aligned} \tag{4.2.6}$$

Similarly, for $N(t, x)$ we have

$$\begin{aligned} \frac{\partial N(t,x)}{\partial t} - \nu \Delta_x N(t,x) + N(t,x) &= \int \eta_2 (1 - \eta_1) p(t,x,\eta_1) q(t,x,\eta_2) d\eta_1 d\eta_2 \\ &= N(t,x) - M(t,x) N(t,x). \end{aligned} \tag{4.2.7}$$

Thus, the functions $N(t,x)$ and $M(t,x)$ satisfy the thermo-diffusive system

$$\begin{aligned} \partial_t N &= \nu \Delta N - NM, \\ \partial_t M &= \kappa \Delta M + NM. \end{aligned} \tag{4.2.8}$$

Our interest in this lecture is the L^∞-bounds on the solutions to (4.2.8) when the initial conditions $M_0(x)$ and $N_0(x)$ are non-negative and uniformly bounded. These assumptions hold if, for example, the initial conditions $p_0(x,\xi)$ and $q_0(x,\xi)$ are supported on a closed interval $0 \le \xi \le L$ with some $L < +\infty$. Their supports can spread to $\xi < 0$ for $t > 0$ but $M(t,x)$ and $N(t,x)$ remain positive for all $t > 0$ and $\xi \in \mathbb{R}^d$. The question of uniform bounds on N and M is quintessentially the issue of how much the voting procedure populates the large values of ξ. In other words, it should be answered by analyzing the tail behavior of $p(t,x,\xi)$ and $q(t,x,\xi)$ as $\xi \to +\infty$. Instead, we will address this issue below by purely PDE methods. Nevertheless, there has to be a probabilistic proof!

4.3 The special case $\kappa = \nu$ (a single equation reduction)

4.3.1 Spreading in a single reaction-diffusion equation

Before describing the recent results for the thermo-diffusive system (4.1.1), that are the main subject of this lecture, let us recall what is known for a single reaction-diffusion equation of the form

$$v_t = \kappa \Delta v + g(v)(1-v). \tag{4.3.1}$$

This equation can be obtained from (4.1.1) when $\kappa = \nu$ and $u_0(x) = 1 - v_0(x)$. Indeed, when $\nu = \kappa$, it is easy to see that the sum $z = u + v$ satisfies the heat equation

$$z_t = \kappa \Delta z. \tag{4.3.2}$$

Therefore, if $z_0(x) \equiv 1$, then $z(t,x) = 1$ as well, and $u(t,x) = 1 - v(t,x)$ for all $t > 0$ and $x \in \mathbb{R}^d$. This leads to (4.3.1).

More generally, as long as $u_0 \in L^\infty(\mathbb{R}^d)$ and $v_0 \in L^\infty(\mathbb{R}^d)$, and $\nu = \kappa$, it follows from (4.3.2) that $v(t,x)$ remains bounded for all $t > 0$:

$$0 \le v(t,x) \le \|u_0 + v_0\|_{L^\infty}, \text{ for all } t > 0 \text{ and } x \in \mathbb{R}^n. \tag{4.3.3}$$

However, this is a very special case and can not be generalized to $\nu \neq \kappa$ in any simple way.

We recall the following picture in dimension $n = 1$ considered in the previous lecture. There exists a reference frame $m(t)$ such that the solution to (4.3.1) converges to a traveling wave:

$$|v(t, x + m(t)) - U(x)| \to 0, \quad \text{as } t \to +\infty, \text{ uniformly in } x \in \mathbb{R}. \tag{4.3.4}$$

The asymptotics of the front location $m(t)$ is also well understood, beyond the Fishr-KPP nonlinearities we have mostly considered so far. Generally, it can be of three types: if $g(v)$ is of the pushed type then

$$m(t) = ct + x_0 + o(1), \quad \text{as } t \to +\infty, \tag{4.3.5}$$

if $g(v)$ is of the pulled type then

$$m(t) = ct - \frac{3}{2\lambda} \log t + x_0 + o(1), \quad \text{as } t \to +\infty, \tag{4.3.6}$$

and, finally, if $g(v)$ is of the pushmi-pullyu type then

$$m(t) = ct - \frac{1}{2\lambda} \log t + x_0 + o(1), \quad \text{as } t \to +\infty. \tag{4.3.7}$$

In higher dimensions, the corresponding results are similar, with radially symmetric spreading, and some explicit modifications of the long time asymptotics of the front location $m(t)$. This is all quite well understood even though we keep writing papers about various aspects of the long time behavior, and intriguing questions keep popping up, especially in terms of the strong connections to the branching Brownian motion.

4.3.2 Spreading results for a non-local equation

The next step from a single semilinear equation toward the thermo-diffusive system is a non-local equation of the form

$$v_t = \kappa \Delta v + g(v)(1 - \phi \star v), \tag{4.3.8}$$

with some non-negative kernel $\phi(x)$ such that

$$\int_{\mathbb{R}^n} \phi(x) dx = 1. \tag{4.3.9}$$

This equation can not, to the best of our knowledge, be obtained from the thermo-diffusive system but serves as a useful model. The main difficulty compared to the local equation (4.3.1) is that the comparison principle does not apply to (4.3.8), which is also true for the thermo-diffusive system.

The long time behavior of the solutions to (4.3.8) is really not understood well at all, even in the Fisher-KPP case $g(v) = v$. Let us summarize some of what is known. For the non-local Fisher-KPP equation, traveling waves exist but the steady state $u \equiv 1$ may or may not be stable. Even if it is unstable there may be a wave connecting $u = 0$ to $u = 1$. Generally, traveling waves are not monotonic, periodic solutions and wave trains may exist, and it is all a mess (Britton'90s, Gourley'00, Geniyes-Volpert-Ager'07, Fang-Zhao'11; G. Nadin, B. Perthame, L. Rossi, LR in various permutations).

In terms of bounds, a uniform in time L^∞ bound on the solutions to a non-local Fisher-KPP equation was obtained in Hamel-R'14. Later, the front location was shown to satisfy the Bramson asymptotics

$$m(t) = 2t - \frac{3}{2} \log t + O(1), \quad \text{as } t \to +\infty, \tag{4.3.10}$$

if the kernel $\phi(x)$ is rapidly decaying. If $\phi(x)$ decays slowly, then the front delay may be algebraic in time: if $\phi(x) \sim |x|^{-r}$ and $1 < r < 3$ then

$$m(t) = 2t - C_\phi t^{(3-r)/(1+r)} + O(1), \quad \text{as } t \to +\infty, \tag{4.3.11}$$

The front location results were shown by Sarah Penington'18 using probabilistic tools and Bouin-Henderson-R.'20 using PDE methods. However, convergence to a traveling wave, or a pulsating wave, or anything else is not known. This part of the long time behavior is completely open, mostly due to our lack of understanding of the convergence mechanism in the absence of a comparison principle.

4.4 The surprisingly simple case $\nu \geq \kappa$

Let us go back to the thermo-diffusive system (4.1.1):

$$\begin{aligned} \partial_t u &= \nu \Delta u - u g(v), \\ \partial_t v &= \kappa \Delta v + u g(v). \end{aligned} \tag{4.4.1}$$

The basic issues we would like to understand are the same as in the two above cases: global bounds on the solution, location of the front and convergence to some asymptotic solution as $t \to +\infty$. Here, we know even less than for the non-local reaction diffusion equations.

We will always suppose that the initial conditions are bounded: there exists $K > 0$ so that

$$0 \le u_0(x) \le K \text{ and } 0 \le v_0(x) \le K \text{ for all } x \in \mathbb{R}^n. \tag{4.4.2}$$

The positivity of $g(v)$ and the comparison principle imply immediately that $u(t,x)$ remains bounded for all times:

$$0 \le u(t,x) \le K, \quad \text{for all } x \in \mathbb{R}^n \text{ and } t > 0. \tag{4.4.3}$$

However, the uniform bounds on the temperature $v(t,x)$ are much less obvious except in the aforementioned case $\nu = \kappa$ that can not be generalized very much.

In terms of the history of the problem, the question of global existence and uniform bound for combustion models was originally posed by Robert Martin in 1970s. For the polynomial nonlinearities, this was resolved by Alikakos'79 for $g(v) = v^\beta$, $1 \le \beta < (n+2)/n$, and Masuda'83 for an arbitrary $\beta > 0$, with further works by Haraux and Youkana'88, Barabanova'94, and a number of interesting papers and expositions by Michel Pierre who is a true champion of these problems. These papers considered the problem on a bounded domain $\Omega \subset \mathbb{R}^n$.

Even on a bounded domain interesting things can happen. Here is a quote from a very nice review by Martin and Pierre'91: "(a) global existence for all nonnegative initial data in O.D.E. does not imply global existence in associated R.D.S. (b) Weak solutions of R.D.S. can be bounded (and even C^∞) at two different times t_1, t_2, but not in (t_1, t_2)." Of course, it is well known that diffusion can destabilize things in reaction-diffusion systems but these are still striking observations.

The uniform bounds on the whole space are much more difficult to obtain because there is, potentially, an infinite supply of fuel, unlike in a bounded domain, which, in principle, could lead to an unbounded growth of the temperature. However, here is another beautiful elementary observation from the same paper by Martin and Pierre.

Theorem 4.4.1. *Consider the system (4.1.1) posed in the whole space $\mathbb{R}^n$ under the assumptions (4.1.3) and (4.4.2). If $\nu \ge \kappa$ then the solution to (4.1.1) exists for all $t \ge 0$ and there exists a constant C_0 such that*

$$0 \le u(t,x) \le C_0 \text{ and } 0 \le v(t,x) \le C_0 \text{ for all } t > 0 \text{ and } x \in \mathbb{R}^n. \tag{4.4.4}$$

The proof is very short and uses two simple but crucial ideas: first, the concentration $u(t,x)$ is bounded from above and remains non-negative by the comparison principle. Second, if $\nu \ge \kappa$ then the heat kernels have the comparison

$$\nu^{n/2} G_\nu(t,x) \ge \kappa^{n/2} G_\kappa(t,x), \quad \text{for all } t > 0 \text{ and } x \in \mathbb{R}^n. \tag{4.4.5}$$

These two facts together with the Duhamel formula imply the result more or less immediately. Indeed, we can write

$$u(t,x) = e^{\nu t \Delta} u_0 - \int_0^t e^{\nu(t-s)\Delta} [u g(v)](s) ds, \tag{4.4.6}$$

and

$$v(t,x) = e^{\kappa t \Delta} v_0 + \int_0^t e^{\kappa(t-s)\Delta} [u g(v)](s) ds. \tag{4.4.7}$$

We see from (4.4.6) that

$$\int_0^t e^{\nu(t-s)\Delta} [u g(v)](s) ds \le K. \tag{4.4.8}$$

The comparison (4.4.5) implies gives a uniform in time upper bound on the temperature $v(t,x)$:

$$v(t,x) \le K + \frac{\nu^{n/2}}{\kappa^{n/2}} \int_0^t e^{\nu(t-s)\Delta}[ug(v)](s)ds \le K + K\frac{\nu^{n/2}}{\kappa^{n/2}}. \tag{4.4.9}$$

Thus, the problem of uniform bounds for the solutions to (4.1.1) is magically trivial in the intuitively harder case $\nu \ge \kappa$ which seemingly can allow the concentration to arrive very fast into the places where the temperature is high, promoting its growth. It is, strangely, quite difficult in the physically supposedly simpler situation when $\nu \le \kappa$. We will solely focus on that situation in the rest of this lecture. One may hope, however, that a probabilistic proof, based on a voting model, would resolve this seeming contradiction.

4.5 Known uniform bounds on the temperature

A localized Lyapunov functional was adopted by Collet and Xin'96 to obtain the global existence of solutions to (4.1.1) in $\mathbb{R}^n$ with bounded non-negative initial conditions for polynomial nonlinearities of the form $g(v) = v^m$. They established an L^∞-bound of the form

$$\|v(t,\cdot)\|_{L^\infty} \le C \log\log t, \quad \text{as } t \to +\infty. \tag{4.5.1}$$

Existence of the global in time solution for the exponential nonlinearities $g(v) = \exp(v)$ in the same setting was proved in a paper by Herrero, Lacey and Velasquez in 1998. They do not state explicitly any bound on the growth of the solution, and their formulation of the main result is a bit vague, but, to the best of our understanding, their proof produces an estimate of the type

$$\|v(t,\cdot)\|_{L^\infty} \le C \log t, \quad \text{as } t \to +\infty. \tag{4.5.2}$$

Xinfu Chen and Yuanwei Qi considered in 2008 the Fisher-KPP system, with $g(v) = v$:

$$\begin{aligned} \partial_t u &= \nu u_{xx} - uv, \\ \partial_t v &= \kappa v_{xx} + uv, \end{aligned} \tag{4.5.3}$$

in one dimension, with the initial condition $u_0(x) \equiv 1$. They were able to show not only that the solution is uniformly bounded in time but also that the front has a finite width and is located at the position

$$m(t) = 2c_* t - \frac{3}{2\lambda_*}\log t + x_0, \quad \text{as } x \to +\infty, \tag{4.5.4}$$

with $c_* = 2\sqrt{\kappa}$, $\lambda_* = 1/\sqrt{\kappa}$. This is exactly the Bramson asymptotics for the location of the front of the solution to the single Fisher-KPP equation

$$v_t = \kappa v_{xx} + v - v^2, \tag{4.5.5}$$

with a compactly supported initial condition $v_0(x)$. Indeed, Chen and Qi very relied heavily on the probabilistic results by Bramson for the single equation and also on the fact that the initial condition $u_0(x)$ is constant and that $g(v) = v$. In that sense, their argument is not quite PDE as it uses the probabilistic results by Bramson as a black box. Because of their use of the Bramson result, their strategy is limited to both $g(v) = v$ and $u_0(x) \equiv 1$. Their key observation is, nevertheless, absolutely ingenious and insightful. Chen and Qi do not show convergence to a traveling wave, as it is not covered by their magic trick, and that question remains open

4.6 The main result: a global in time uniform bound

We consider the system (4.4.1):

$$\begin{aligned} \partial_t u &= \nu \Delta u - u g(v), \\ \partial_t v &= \kappa \Delta v + u g(v). \end{aligned} \tag{4.6.1}$$

For the initial conditions, as we have mentioned, we assume that they are uniformly bounded and are non-negative: there exists $K > 0$ so that at $t = 0$ we have

$$0 \le u_0(x) \le K, \quad 0 \le v_0(x) \le K, \quad \text{for all } x \in \mathbb{R}^n. \tag{4.6.2}$$

In addition, we assume that there is a uniformly positive lower bound on the sum $u_0(x) + v_0(x)$:

$$0 < K^{-1} \le u_0(x) + v_0(x) \le K, \quad \text{for all } x \in \mathbb{R}^n. \tag{4.6.3}$$

This assumption is both technical, in the sense that it should not be necessary, and critical for us. It is not as stringent from the physical point of view as it seems. For example, a commonly considered initial condition is $u_0(x) \equiv \text{const}$ – the fuel is uniform in space initially, so that (4.6.3) holds automatically. It is very feasible that one could replace (4.6.3) by a much weaker condition, or get rid of it completely, but this is not our goal here.

As for the nonlinearity, we assume that $g(v)$ satisfies the following assumptions: $g(0) = 0$ and there exist $c_1 > 0$, $c_2 > 0$, $\rho \in [0, 1)$ and $Z > 0$ so that

$$c_1 v \le g(v) \le c_2 \exp(Z v^\rho) \text{ for all } v \ge 0. \tag{4.6.4}$$

Global in time existence of the solutions to (4.6.1) under these assumptions follows from the paper by Herrero, Lacey and Velasquez. Our main result in [80] is the following uniform bound.

Theorem 4.6.1. *Let $u(t,x), v(t,x)$ be a solution to (4.6.1) with the initial conditions $u(0,x) = u_0(x)$ and $v(0,x) = v_0(x)$ that satisfy (4.6.2) and (4.6.3), with some $K > 0$. Suppose that $g(v)$ satisfies the assumption (4.6.4). Then, $u(t,x)$ and $v(t,x)$ are globally bounded in time: there exists $C > 0$ that depends on the constants in the above assumptions, so that*

$$0 \le u(t,x), \le C, \quad 0 \le v(t,x) \le C \tag{4.6.5}$$

for all $t > 0$ and $x \in \mathbb{R}^n$.

Let us comment on the assumptions (4.6.4) we have made on the nonlinearity. The condition $g(0) = 0$ is not needed, strictly speaking, but is physical and allows us not to consider separately the much simpler case $g(0) > 0$, hence we adopt it here, mostly for convenience. The restriction that $\rho < 1$, meaning that $g(v)$ grows sub-exponentially as $v \to +\infty$ is done to simplify some steps in the proof and can likely be removed in the same way as in the paper by Herrero, Lacey and Velasquez.

On the other hand, the lower bound in (4.6.4) is absolutely crucial for us, as will become clear from the outline of the proof. In other words, it is not the growth of $g(v)$ for large v that matters the most in the proof of the uniform bound on $v(t,x)$ but rather how $g(v)$ behaves near $v = 0$. In particular, a very interesting open question is to replace the lower bound even by the assumption $g(v) \ge cv^2$ for $0 \le v \le 1$.

Let us also comment that the assumption $g(v) \ge c_1 v$ is not just a weakening of the Chen-Qi assumption $g(v) = v$ – it is really a very different regime from the Fisher-KPP situation since the

behavior of the solutions is not dominated by the linearized problem at infinity. In particular, the Bramson asymptotics (4.5.4) do not have to hold for a single equation of the form

$$v_t = \kappa \Delta v + g(v)(1-v). \tag{4.6.6}$$

What we really need from $g(v)$ is that (4.6.6) has what is known as the hair trigger effect. This effect is much more general than the Fisher-KPP property, and the results and methods of Bramson and others that lead to the precise front location asymptotics (4.5.4) do not apply or hold.

We should also mention that the questions of the long time behavior of the solutions, the front locations asymptotics and convergence to some kind of a special solution, such as a traveling wave or a wave train, or a pulsating wave, are still completely open.

In the rest of the lecture we describe the main ingredients in the proof of Theorem 4.6.1.

4.7 Different diffusivities but identical forcing: the parabolic BMO bound

Let us recall (4.1.1)

$$\begin{aligned} \partial_t u &= \nu \Delta u - u g(v), \\ \partial_t v &= \kappa \Delta v + u g(v), \end{aligned} \tag{4.7.1}$$

and also that we know immediately that $u(t,x)$ is uniformly bounded. This leads to the following simple question. Let $F(t,x)$ and $H(t,x)$ be the solutions to a pair of standard forced heat equations

$$\partial_t F = \nu \Delta F + Z(t,x), \tag{4.7.2}$$

and

$$\partial_t H = \kappa \Delta H + Z(t,x), \tag{4.7.3}$$

with the initial conditions $F(0,x) = H(0,x) = 0$. Note that (4.7.2) and (4.7.3) have two different diffusivities $\kappa \neq \nu$ but the forcing term $Z(t,x)$ in the two equations is the same. Suppose that we know that $Z(t,x) \geq 0$ for all $t \geq 0$ and $x \in \mathbb{R}^d$ and that $F(t,x)$ satisfies a uniform bound

$$0 \leq F(t,x) \leq K, \quad \text{for all } t \geq 0 \text{ and } x \in \mathbb{R}^d. \tag{4.7.4}$$

The question is what kind of bound on the function $H(t,x)$ one can infer from (4.7.4). This is the situation we have in (4.7.1).

The main observation of the Herrero-Lacey-Velasquez paper is that one can not deduce a uniform bound on $H(t,x)$ simply from the boundedness of $F(t,x)$ but the function $H(t,x)$ does obey a parabolic BMO bound. Before we state the result, let us set some notation and recall the definition of the parabolic BMO semi-norm. A parabolic cylinder of size $R > 0$ centered at a point $(t,x) \in \mathbb{R} \times \mathbb{R}^n$ is a set of the form

$$Q = Q_R(t,x) = \{(s,y) : \ |x-y| < R, \ 0 < t-s < R^2\} \subset \mathbb{R} \times \mathbb{R}^n. \tag{4.7.5}$$

For $f \in L^1_{loc}(\mathbb{R} \times \mathbb{R}^n)$, we denote the average of f over a measurable set Q by

$$(f)_Q = \frac{1}{|Q|} \int_Q f(x,t) dx dt.$$

Here, $|Q|$ is the $(n+1)$-dimensional Lebesgue measure of Q. The parabolic BMO semi-norm is defined by

$$\|f\|_{\text{p-BMO}} = \sup \Big\{ \frac{1}{|Q|} \int_Q |f - (f)_Q| dx dt \Big\},$$

with the supremum is taken over all parabolic cylinders $Q \subset \mathbb{R}^n \times \mathbb{R}_+$.

The key point is that the solution H to (4.7.3) can be written as a space-time singular integral of the bounded function F. A simple algebra shows that

$$H = F + (\kappa - \nu)\mathcal{T}_\kappa F. \tag{4.7.6}$$

Here, we have set

$$\mathcal{T}_\kappa F := \Delta(\partial_t - \kappa\Delta)^{-1} F. \tag{4.7.7}$$

That is, $\Psi = \mathcal{T}_\kappa F$ is the solution to

$$\Psi_t - \kappa\Delta\Psi = \Delta F, \quad \Psi(0,x) = 0. \tag{4.7.8}$$

The key observation is the following p-BMO bound.

Proposition 4.7.1. *There exists a constant $C_{\mathcal{T}}$, which depends on the diffusivity $\kappa > 0$ and the dimension n such that for all $F \in L^\infty(\mathbb{R}_+ \times \mathbb{R}^n)$, we have*

$$\|\mathcal{T}_\kappa F\|_{\text{p-BMO}} \le C_{\mathcal{T}} \|F\|_{L^\infty(\mathbb{R}_+ \times \mathbb{R}^n)}. \tag{4.7.9}$$

This result corresponds to Lemma 3.3 of the Herrero-Lacey-Velasquez paper. The proof presented in our paper is very different and is based on the estimates on the singular kernel of the operator $\mathcal{T}_\kappa$. It is a relatively straightforward application of the harmonic analysis.

Here is an immediate consequence of Proposition 4.7.1.

Corollary 4.7.2. *Let $u(t,x)$, $v(t,x)$ be the solution to (4.7.1) with initial conditions $u(0,x) = u_0(x)$ and $v(0,x) = v_0(x)$ that satisfy (4.6.2). Assume, in addition, that $g(v) \ge 0$ for all $v \ge 0$. There exists a constant C_1 that depends on the constant K in (4.6.2), and also on the dimension n and the diffusivities $\kappa > 0$ and $\nu > 0$ but not on the function $g(v)$ such that*

$$\|v\|_{\text{p-BMO}(\mathbb{R}_+ \times \mathbb{R}^n)} \le C_1. \tag{4.7.10}$$

Let us emphasize that the proof of Corollary 4.7.2 uses neither the upper sub-exponential bound nor the lower bound in (4.6.4) on the reaction rate $g(v)$. Its conclusion holds as long as the solution to (4.7.1) exists and the production rate $g(v)$ is non-negative for $v \ge 0$.

4.8 From bounded averages to bounded pointwise values

The next step is to show that the point-wise values of the function $v(t,x)$ are controlled by its local averages. This is a consequence of the parabolic BMO bound in Corollary 4.7.2 and the ABP maximum principle. Here is the main estimate.

Proposition 4.8.1. *For any $\varepsilon \in (0,1)$ there is $C_\varepsilon > 0$ such that for any $t_0 \ge 1$ and $x_0 \in \mathbb{R}^n$ we have*

$$\sup_{Q_1(t_0,x_0)} v(t,x) \le C_\varepsilon (v)^{1-\varepsilon}_{Q_2(t_0,x_0)} \exp\left(C_{p,\varepsilon}(v)_{Q_2(t_0,x_0)}\right). \tag{4.8.1}$$

The proof of Proposition 4.8.1 uses the parabolic BMO bound in Corollary 4.7.2 and the parabolic reglularity theory applied to the equation

$$v_t = \kappa\Delta v + ug(v). \tag{4.8.2}$$

The first step in the proof of Proposition 4.8.1 is the following exponential moment bound that follows from the BMO bound in Proposition 4.7.1.

Proposition 4.8.2. *For any $Z \geq 0$ and $\rho \in (0,1)$ there exists a constant $C_{Z,\rho} > 0$ that depends on all of our constants so that for all $t_0 > 1$ and $x_0 \in \mathbb{R}^n$, we have*

$$\frac{1}{|Q_1(t_0,x_0)|}\int_{Q_1(t_0,x_0)} e^{Zv^\rho(t,x)}dxdt \leq C_{Z,\rho}\exp\big(C_{Z,\rho}(v)_{Q_1}\big). \tag{4.8.3}$$

We omit the details of the proof of this estimate but mention that it uses the parabolic John-Nirenberg inequality that bounds the deviation of a p-BMO function from its local average.

Proposition 4.8.3. *(John-Nirenberg inequality) There exist constants $A, B > 0$, such that for all parabolic cylinders $Q = Q_R(t_0,x_0)$ and all $\lambda > 0$, we have*

$$\big|\{(t,x) \in Q : \ |f(t,x) - (f)_Q| \geq \lambda\}\big| \leq B\exp\Big(-\frac{A\lambda}{\|f\|_{\text{p-BMO}}}\Big)|Q|, \tag{4.8.4}$$

for all $g \in$ p-BMO. Here, $|Q|$ is the $(n+1)$-dimensional Lebesgue measure of Q.

Proposition 4.8.2 gives the following.

Corollary 4.8.4. *For any $\varepsilon > 0$ and $p \geq 1$ there exists a constant $C_{p,\varepsilon} > 0$ that depends on all of our constants so that for all $t_0 > 1$ and $x_0 \in \mathbb{R}^n$, we have*

$$\frac{1}{|Q_1(t_0,x_0)|}\int_{Q_1(t_0,x_0)} [g(v(t,x))]^p dxdt \leq C_{p,\varepsilon}[(v)_Q]^{1-\varepsilon}\exp\big(C_{p,\varepsilon}(v)_Q\big). \tag{4.8.5}$$

Proof. Let us write, using Hölder's inequality

$$\begin{aligned}
&\frac{1}{|Q_1(t_0,x_0)|}\int_{Q_1(t_0,x_0)} [g(v(t,x))]^p dxdt \\
&\leq \frac{1}{|Q_1(t_0,x_0)|}\Big(\int_{Q_1(t_0,x_0)} v(t,x)dx\Big)^{1/p}\Big(\int_{Q_1(t_0,x_0)}\Big(\frac{g^p(v(t,x))}{v^{1-1/q}(t,x)}\Big)^q dx\Big)^{1/q} \\
&\leq \frac{1}{|Q_1(t_0,x_0)|^{1/q}}[(v)_Q]^{1/p}\Big(\int_{Q_1(t_0,x_0)} e^{Mv^\rho(t,x)}dx\Big)^{1/q}.
\end{aligned} \tag{4.8.6}$$

Here, we used the assumption (4.6.4):

$$c_1 v \leq g(v) \leq c_2\exp(Zv^\rho) \text{ for all } v \geq 0. \tag{4.8.7}$$

Now, (4.8.5) follows after taking p large enough and taking (4.8.3) into account.. □

Note that the above proof of Corollary 4.8.4 does use the fact that $g(0) = 0$ and $g(v)$ is smooth. We also note that it is a purely harmonic analysis statement and does not use the parabolic equation for the function v at all except for the BMO bound in (4.7.10).

Next, we recall the following version of the parabolic ABP maximum principle.

Proposition 4.8.5. *Let $h(t,x) \geq 0$ be the solution to*

$$h_t = \kappa\Delta h + f(t,x). \tag{4.8.8}$$

Then, we have, for any $p \geq 1$, $t_0 \geq 3$ and $x \in \mathbb{R}^n$:

$$\sup_{Q_1(t_0,x_0)} h \leq C\big(\|h\|_{L^p(Q_2(t_0,x_0))} + \|f\|_{L^{n+1}(Q_2(t_0,x_0))}\big). \tag{4.8.9}$$

Let us go back to the equation for $v(t,x)$:

$$v_t = \kappa\Delta v + ug(v). \tag{4.8.10}$$

Proposition 4.8.5 implies that $v(t,x)$ obeys a bound

$$\sup_{Q_1(t_0,x_0)} v(t,x) \le C\big(\|v\|_{L^p(Q_2(t_0,x_0))} + \|ug(v)\|_{L^{n+1}(Q_2(t_0,x_0))}\big), \tag{4.8.11}$$

and the conclusion of Proposition 4.8.1 now follows from Corollary 4.8.4 that bounds the first term in the right side and Proposition 4.8.2 that bounds the second because of the upper bound on $g(v)$ in (4.8.7).

Thus, the problem of a global uniform bound on $v(t,x)$ is reduced to bounding the local averages of v and this is the main new ingredient in the proof.

4.9 Comparison to a single reaction-diffusion equation

4.9.1 The Michel Pierre argument revisited

Let us first recall again how a uniform in time bound for the solutions to (4.1.1) with $\nu \ge \kappa$ on $\mathbb{R}^n$ can be proved by comparing the heat kernels with the diffusivities ν and κ, with no restrictions on the function $g(v)$ whatsoever using a beautiful idea by M. Pierre. Namely, we have

$$\nu^{n/2}G_\nu(t,x) \ge \kappa^{n/2}G_\kappa(t,x), \quad \text{for all } t>0 \text{ and } x\in\mathbb{R}^n \text{ if } \nu\ge\kappa>0. \tag{4.9.1}$$

It follows from (4.9.1) that if $\nu \ge \kappa$ we can write

$$u(t,x) = \overline{u}(t,x) - \int_0^t\int G_\nu(t-s,x-y)u(s,y)g(v(s,y))dyds, \tag{4.9.2}$$

and

$$\begin{aligned} v(t,x) &= \overline{v}(t,x) + \int_0^t\int G_\kappa(t-s,x-y)u(s,y)g(v(s,y))dyds \\ &\le \overline{v}(t,x) + \frac{\nu^{n/2}}{\kappa^{n/2}}\int_0^t\int G_\nu(t-s,x-y)u(s,y)g(v(s,y))dyds \\ &\le \overline{v}(t,x) + \frac{\nu^{n/2}}{\kappa^{n/2}}(\overline{u}(t,x)-u(t,x)) \le CK. \end{aligned} \tag{4.9.3}$$

Here, $\overline{u}$ and $\overline{v}$ are the solutions to the standard heat equation:

$$\begin{aligned} \overline{u}_t &= \Delta\overline{u}, \quad \overline{u}(0,x) = u_0(x), \\ \overline{v}_t &= \Delta\overline{v}, \quad \overline{v}(0,x) = v_0(x). \end{aligned} \tag{4.9.4}$$

Thus, in the case $\nu \ge \kappa$ the upper bound on $v(t,x)$ is very simple.

4.9.2 A lower bound by a single equation if $\nu \le \kappa$

If $\nu \le \kappa$, the above argument gives instead a lower bound that played a crucial role in the analysis of Chen and Qi in the special case $g(v) = v$, $u_0(x) \equiv 1$, and $n = 1$

$$v(t,x) \ge \overline{v}(t,x) + \frac{\nu^{n/2}}{\kappa^{n/2}}(\overline{u}(t,x)-u(t,x)), \tag{4.9.5}$$

or, equivalently,

$$u(t,x) \geq \overline{u}(t,x) + \frac{\kappa^{n/2}}{\nu^{n/2}}\overline{v}(t,x) - \frac{\kappa^{n/2}}{\nu^{n/2}}v(t,x). \tag{4.9.6}$$

Note that if the lower bound in (4.6.3) holds:

$$0 < K^{-1} \leq u_0(x) + v_0(x) \leq K, \quad \text{for all } x \in \mathbb{R}^n, \tag{4.9.7}$$

and $\nu \leq \kappa$ then we have

$$\begin{aligned} \overline{u}(t,x) + \frac{\kappa^{n/2}}{\nu^{n/2}}\overline{v}(t,x) &= \frac{\kappa^{n/2}}{\nu^{n/2}}G_\kappa(t,\cdot) \star v_0 + G_\nu(t,\cdot) \star u_0 \\ &\geq G_\nu(t,\cdot) \star v_0 + G_\nu(t,\cdot) \star u_0 \geq G_\nu(t,\cdot) \star (u_0 + v_0) \geq K^{-1} := d_0. \end{aligned} \tag{4.9.8}$$

We deduce that

$$u(t,x) \geq d_0 - d_1 v, \tag{4.9.9}$$

with

$$d_1 = \frac{\kappa^{n/2}}{\nu^{n/2}}. \tag{4.9.10}$$

Let us summarize this as follows.

Proposition 4.9.1. *Let $v(t,x)$ and $u(t,x)$ be the solutions to (4.1.1) with the initial conditions $u_0(x)$ and $v_0(x)$ that satisfy (4.6.2) and (4.6.3). If $\nu \leq \kappa$ then $v(t,x)$ is a super-solution to*

$$v_t \geq \kappa\Delta v + (d_0 - d_1 v)g(v), \tag{4.9.11}$$

with $d_0 = \inf_{x\in\mathbb{R}^d}(u_0 + v_0)$ and $d_1 = \kappa^{n/2}/\nu^{n/2}$.

We should stress that, first, the conclusion of this proposition does not use anything about the nonlinearity $g(v)$, and, second, it relies crucially on the lower bound on the sum $u_0(x) + v_0(x)$ in (4.6.3) that makes the constant d_0 in (4.9.8) be positive. Of course, if the set where this sum vanishes is sufficiently sparse, one could for some time, until it becomes positive everywhere.

Note that when $u_0(x) \equiv 1$ we can take $d_0 = 1$. Then, in the special case $g(v) = v$ and in the dimension $n = 1$, Chen and Qi used Proposition 4.9.1 and Bramson's results for the single equation to immediately identify the potential reaction zone and show that $v(t,x)$ remains bounded and, moreover, the front width remains $O(1)$ as $t \to +\infty$.

4.10 The hair trigger effect

The main consequence of Proposition 4.9.1 is that it controls the function $v(t,x)$ from below on a set that grows linearly in time. For that, we rely crucially on the lower bound in (4.6.4) for the function $g(v)$ for $v > 0$ small:

$$c_1 v \leq g(v), \text{ for all } v \geq 0. \tag{4.10.1}$$

This leads to the hair trigger effect.

Proposition 4.10.1. *There exist $\gamma_0 > 0$, $\lambda_0 > 0$, $c_0 > 0$ and $C > 0$ that depend on the constant K in (4.6.2) and (4.6.3), the diffusivities κ and ν, and the constant c_1 so that the following holds. Assume that at a time $t_0 > 0$ we have, for some $x_0 \in \mathbb{R}^n$,*

$$\int_{|x-x_0|\leq 1} v(t_0,x)dx \geq \delta_0, \tag{4.10.2}$$

with some $\delta_0 \in (0,1)$. *Then, we have*

$$v(t,x) \geq \min\left(\gamma_0, C^{-1}\delta_0 e^{\lambda_0(t-t_0)}\right), \quad \textit{for all } t \geq t_0+1 \textit{ and } |x-x_0| \leq 1, \tag{4.10.3}$$

and

$$v(t,x) \geq \gamma_0, \quad \textit{for all } t \geq t_1 := t_0 + \frac{1}{\lambda_0}|\log \delta_0| + C \textit{ and } |x-x_0| \leq c_0(t-t_1). \tag{4.10.4}$$

We should mention that the upper bound in (4.6.4) is not required here, we only rely on the lower bound on $g(v)$ in (4.10.1).

Proof. Step 1. A sub-solution for $v(t,x)$**.** First, we construct a sub-solution for $v(t,x)$. Observe that, due to the assumption (4.10.1), if we choose $\alpha_0 = d_0/(2d_1)$, then there exists $\beta_0 > 0$ so that

$$(d_0 - d_1 s)g(s) \geq \beta_0 s, \quad \text{for all } 0 \leq s \leq \alpha_0. \tag{4.10.5}$$

Let us now choose L sufficiently large, so that the principal eigenvalue of the Dirichlet problem

$$\begin{aligned} -\kappa\Delta\varphi_L &= \lambda_L \varphi_L, \quad \text{for } x \in B(0,L), \\ \varphi_L &= 0, \quad \text{on } \partial B(0,L), \end{aligned} \tag{4.10.6}$$

satisfies

$$0 < \lambda_L < \beta_0/2. \tag{4.10.7}$$

We normalize the non-negative principal eigenfunction $\varphi_L(x)$ so that

$$\varphi_L(0) = \sup_{x \in B(0,L)} \varphi_L(x) = 1.$$

The point is that the function φ_L satisfies the differential inequality

$$\kappa\Delta\varphi_L + \beta_0\varphi_L \geq 0, \quad \text{for } x \in B(0,L). \tag{4.10.8}$$

We see from (4.10.5) and (4.10.8) that for $0 < \delta \leq \alpha_0$ the function $\varphi_{\delta,L}(x) = \delta\varphi_L(x-x_0)$ obeys the differential inequality

$$\kappa\Delta\varphi_{\delta,L} + (d_0 - d_1\varphi_{\delta,L})g(\varphi_{\delta,L}) \geq 0, \quad \text{in } B(x_0,L). \tag{4.10.9}$$

Thus, the function

$$\tilde{\omega}_\delta(x) = \delta\varphi_L(x-x_0)\mathbb{1}(|x-x_0| \leq L)$$

is a time-independent sub-solution to (4.10.10), as a maximum of two sub-solutions, $\phi_{\delta,L}(x)$ and the function identically equal to zero.

Consider the function $\omega_\delta(t,x)$, the solution to

$$\partial_t\omega_\delta = \kappa\Delta\omega_\delta + (d_0 - d_1\omega_\delta)g(\omega_\delta), \quad t \geq t_0 + 1/2, \tag{4.10.10}$$

with the initial condition

$$\omega_\delta(t_0+1/2, x) = \tilde{\omega}_\delta(x). \tag{4.10.11}$$

As the initial condition $\tilde{\omega}_\delta(x)$ is a time-independent sub-solution to (4.10.10), the comparison principle shows that $\omega_\delta(t_0+1/2+h,x) - \tilde{\omega}_\delta(x)$ for any $h > 0$. The function $z(t,x) = \omega_\delta(t+h,x) > \omega_\delta(t,x)$ satisfies a parabolic equation of the form

$$z_t = \kappa\Delta z(t,x) + c(t,x)z, \tag{4.10.12}$$

with a bounded function $c(t,x)$ given by

$$c(t,x) = \frac{(d_0 - d_1\omega_\delta(t+h,x))g(\omega_\delta(t+h,x)) - (d_0 - d_1\omega_\delta(x))g(\omega_\delta(x))}{\omega_\delta(t+h,x) > \omega_\delta(t,x)}, \tag{4.10.13}$$

and a positive initial condition

$$z(t_0 + 1/2, x) = \omega_\delta(t_0 + 1/2 + h, x) - \tilde{\omega}_\delta(x) \geq 0. \tag{4.10.14}$$

The comparison principle implies that $z(t,x) > 0$ for all $x \in \mathbb{R}^n$ and $t > 0$, for any $h > 0$. Thus, the function $\omega_\delta(t,x)$ is increasing in time.

In addition, the function

$$\mu(t,x) = \delta e^{(\beta_0 - \lambda_L)(t - t_0 - 1/2)}\varphi_L(x - x_0), \tag{4.10.15}$$

satisfies

$$\mu_t - \kappa\Delta\mu = \beta_0\mu \leq (d_0 - d_1\mu)g(\mu), \quad \text{in } B(x_0, L), \text{ as long as } \|\mu(t,\cdot)\|_{L^\infty} \leq \alpha_0, \tag{4.10.16}$$

with the boundary condition $\mu(t,x) = 0$ for $x \in \partial B(x_0, L)$. We deduce from the comparison principle that

$$\omega_\delta(t,x) \geq \delta e^{(\beta_0 - \lambda_L)(t - t_0 - 1/2)}\varphi_L(x - x_0), \quad \text{as long as } \delta e^{(\beta_0 - \lambda_L)(t - t_0)} \leq C, \tag{4.10.17}$$

with a constant $C > 0$ that depends on α_0, β_0 and λ_L. Since the function $\omega_\delta(t,x)$ is increasing in t and the eigenfunction $\varphi_L(x)$ is uniformly bounded from below on $B(0,1)$, as long as $L > 2$, we deduce that there exist $\gamma > 0$ and $C > 0$ so that

$$\omega_\delta(t,x) \geq \min\left(\gamma, C\delta e^{(\beta_0 - \lambda_L)(t - t_0 - 1/2)}\right), \quad \text{for } t > t_0 + 1/2 \text{ and } |x - x_0| \leq 1. \tag{4.10.18}$$

Step 2. Using the sub-solution. We now use the sub-solution $\omega_\delta(t,x)$ constructed above to bound the function $v(t,x)$ from below. Suppose that $v(t_0,x)$ satisfies assumption (4.10.2) with some $\delta_0 \in (0,1)$. Let us fix $L > 0$ so that (4.10.7) holds. As $v(t,x)$ is a super-solution of the heat equation:

$$v_t \geq \kappa\Delta v, \tag{4.10.19}$$

it follows that there exists a constant $k_L > 0$ so that the assumption (4.10.2) implies that

$$v(t_0 + 1/2, x) \geq k_L\delta_0, \quad \text{for all } |x - x_0| \leq L. \tag{4.10.20}$$

We choose $\delta = \min(k_L\delta_0, \alpha_0)$. Then, at the time $t = t_0 + 1/2$ we have

$$v(t_0 + 1/2, x) \geq \omega_\delta(t_0 + 1/2, x), \quad \text{for all } x \in \mathbb{R}^n. \tag{4.10.21}$$

It follows from (4.9.11) and (4.10.10) that

$$v(t,x) \geq \omega_\delta(t,x), \quad \text{for all } t > t_0 + 1/2 \text{ and } x \in \mathbb{R}^n. \tag{4.10.22}$$

Going back to (4.10.18), we deduce that

$$v(t,x) \geq \omega_\delta(t,x) \geq \min\left(\gamma, C\delta e^{(\beta_0 - \lambda_L)(t - t_0 - 1/2)}\right), \quad \text{for } t > t_0 + 1/2 \text{ and } |x - x_0| \leq 1. \tag{4.10.23}$$

As $0 < \delta_0 < 1$, we have

$$\delta = \min(k_L\delta_0, \alpha_0) \geq \delta_0 \min(k_L, \alpha_0) = C\delta_0. \tag{4.10.24}$$

Using this in (4.10.23) gives

$$v(t,x) \geq \min\left(\gamma, C\delta_0 e^{(\beta_0-\lambda_L)(t-t_0-1/2)}\right), \quad \text{for } t > t_0+1/2 \text{ and } |x-x_0| \leq 1, \tag{4.10.25}$$

which is (4.10.3).

A consequence of (4.10.25) is that there is $C > 0$ so that at the time

$$t_1 = t_0 + \frac{1}{\beta_0 - \lambda_L}|\log \delta_0| + C \tag{4.10.26}$$

we have

$$v(t_1, x) \geq \gamma, \quad \text{for all } |x - x_0| \leq 1. \tag{4.10.27}$$

Now, the bound (4.10.4) follows from (4.10.27) and the classical results of Aronson and Weinberger on spreading for solutions to the Fisher-KPP type equations for solutions that are initially strictly positive on a unit ball. More precisely, let $\tilde{v}(t,x)$ be the solution to the initial value problem

$$\begin{aligned} &\tilde{v}_t = \kappa\Delta\tilde{v} + (d_0 - d_1\tilde{v})g(\tilde{v}), \quad t \geq t_1, \\ &\tilde{v}(t_1, x) = \gamma\mathbb{1}(|x - x_0| \leq 1), \end{aligned} \tag{4.10.28}$$

and let c_* be the minimal speed for a traveling wave solution for (4.10.10). Then, Theorems 3.1 and 5.3 of [18] imply that for any $0 < c < c_*$ we have

$$\liminf_{t\to+\infty} \inf_{|y-x_0|<c(t-t_1)} \tilde{v}(t,y) \geq \frac{d_0}{2d_1}. \tag{4.10.29}$$

Thus, there exists a time T_0 that depends only on γ so that

$$\tilde{v}(t,y) \geq \frac{d_0}{4d_1}, \quad \text{for all } t > t_1 + T_0 \text{ and } |y - x_0| \leq c(t - t_1). \tag{4.10.30}$$

It follows from the comparison principle, (4.9.11), (4.10.27), (4.10.28) and (4.10.30) that

$$v(t,y) \geq \tilde{v}(t,y) \geq \frac{d_0}{4d_1}, \quad \text{for all } t > t_1 + T_0 \text{ and } |y - x_0| \leq c(t - t_1), \tag{4.10.31}$$

which is (4.10.4). In particular, we can let the speed c_0 in (3.4.50) be smaller than the minimal speed for a traveling wave solution for (4.10.10). □

4.11 A forward in time bound on the fuel concentration

Proposition 4.10.1 has several important corollaries that will be crucial for us. First, we have a forward-in-time upper bound on the fuel.

Corollary 4.11.1. *Assume that the assumptions of Proposition 4.10.1 hold. There exist $c_2 > 0$ and $\gamma_2 > 0$ so that the following holds. Suppose that at some time $t_0 > 0$ the inequality (4.10.2) holds with some $\delta_0 \in (0,1)$:*

$$\int_{|x-x_0|\leq 1} v(t_0, x)dx \geq \delta_0, \tag{4.11.1}$$

Then, we have

$$u(t,x) \leq Ke^{-\gamma_2(t-t_1)}, \quad \text{for all } t \geq t_1 = t_0 + |\log\delta_0| + C \text{ and } |x - x_0| \leq c_2(t - t_1). \tag{4.11.2}$$

Proof. Informally, the reason for this estimate is that Proposition 4.10.1 implies that $v(t,x) \geq \gamma_0$ in the region $|x - x_0| \leq c_0(t - t_1)$ that is much bigger than the domain $|x - x_0| \leq c_2(t - t_1)$ which appears in (4.11.2), by a factor growing linearly in $t - t_1$, if we take, say, $c_2 < c_0/2$. This gives the exponential decay on the fuel $u(t,x)$.

To make this precise, let γ_0 be a lower bound for $v(t,x)$ in the cone

$$\mathcal{C}_{t_1,x_0} := \{|x - x_0| \leq c_0|t - t_1|\},$$

as in (3.4.50), and let

$$q_0 = \inf_{v \geq \gamma_0} g(v). \tag{4.11.3}$$

Then, in the cone $\mathcal{C}_{t_1,x_0}$ the function $u(t,x)$ satisfies

$$u_t - \nu\Delta u + q_0 u \leq 0. \tag{4.11.4}$$

Let us look for a super-solution to (4.11.4) in the form

$$\overline{u}(t,x) = Me^{-\alpha(t-t_1)} \cosh\ (\beta|x - x_0|), \tag{4.11.5}$$

with the positive constants M, α and β to be chosen. Note that we have, in the polar coordinates

$$\overline{u}_t - \nu\Delta\overline{u} + q_0\overline{u} = \overline{u}_t - \nu\overline{u}_{rr} - \nu\frac{n-1}{r}\overline{u}_r + q_0\overline{u} = \Big(- \alpha - \nu\beta^2 - (n-1)\nu\beta\frac{\tanh\ r}{r} + q_0\Big)\overline{u}. \tag{4.11.6}$$

As $\tanh r$ is concave for $r > 0$, we have

$$\frac{\tanh\ r}{r} \leq 1, \quad \text{for all } r > 0.$$

Using this in (4.11.6) gives

$$\overline{u}_t - \nu\Delta\overline{u} + q_0\overline{u} \geq (q_0 - \alpha - \nu\beta^2 - (n-1)\nu\beta)\overline{u}. \tag{4.11.7}$$

Therefore, we may choose $\alpha > 0$ and $\beta > 0$ suitably small so that the right side in (4.11.7) is positive. In addition, we may choose α and β so that we have the inclusion

$$\tilde{\mathcal{C}}_{t_1,x_0} := \{|x - x_0| \leq \frac{\beta}{\alpha}|t - t_1|\} \subset \mathcal{C}_{t_1,x_0} = \{|x - x_0| \leq c_0|t - t_1|\}.$$

With this choice of $\alpha > 0$ and $\beta > 0$, we know that $u(t,x)$ is a sub-solution and $\overline{u}(t,x)$ is a super-solution to (4.11.4) in the cone $\tilde{\mathcal{C}}_{t_1,x_0}$. Moreover, as $u(t,x) \leq K$ for all $t \geq 0$ and $x \in \mathbb{R}^n$, we can choose $M > 0$ sufficiently large, so that $u(t,x) \leq \overline{u}(t,x)$ for all (t,x) on the boundary of $\tilde{\mathcal{C}}_{t_1,x_0}$. It follows that then $u(t,x) \leq \overline{u}(t,x)$ in the whole cone $\tilde{\mathcal{C}}_{t_1,x_0}$. This gives the exponential decay of u in any cone of the form $\{|x - x_0| \leq c|t - t_1|\}$ with $c < \alpha/\beta$. □

4.12 Backward in time bound on the fuel temperature averages

Next, we state a backward-in-time bound that shows that if the average of the temperature is small in a unit ball then the averages over unit balls have to be small in a growing region at earlier times.

Corollary 4.12.1. *Assume that the assumptions of Proposition 4.10.1 hold, and let c_0 be as in the conclusion of that proposition. There exist the constants $T_0 > 0$, $C_1 > 0$, $\varepsilon_0 > 0$ and $\gamma_3 > 0$ so that the following holds. Assume that at some time $t_0 > T_0$ we have, for some $x_0 \in \mathbb{R}^n$:*

$$\int_{|x-x_0|\leq 1} v(t_0,x)dx \leq \varepsilon_0. \tag{4.12.1}$$

Then, we have

$$\int_{|y-x|\leq 1} v(t,y)dy \leq C\varepsilon_0 e^{-\gamma_3(t_0-t)}, \quad \textit{for all } t \leq t_0 \textit{ and } |x-x_0| \leq \frac{c_0}{2}(t_0-t) - C_1. \tag{4.12.2}$$

Proof. Let γ_0 be as in Proposition 4.10.1 and fix some $\varepsilon_0 \in (0, \gamma_0/10)$. Assume that (4.12.2) is violated at some time $t' < t_0$ and $x' \in \mathbb{R}^n$:

$$\int_{|y-x'|\leq 1} v(t',y)dy \geq C\varepsilon_0 e^{-\gamma_3(t_0-t)}. \tag{4.12.3}$$

It follows from Proposition 4.10.1 that then

$$v(t,x) \geq \gamma_0, \text{ for all } t \geq t_1' := t' + \frac{1}{\lambda_0}|\log\varepsilon_0| + \frac{\gamma_3}{\lambda_0}(t_0-t') + C \text{ and } |x-x'| \leq c_0(t-t_1'). \tag{4.12.4}$$

If we choose $\gamma_3 < \lambda_0$ then we have

$$t_1' := t' + \frac{1}{\lambda_0}|\log\varepsilon_0| + \frac{\gamma_3}{\lambda_0}(t_0-t') + C < t_0, \tag{4.12.5}$$

provided that $T_0 > 0$ is large enough. We deduce from (4.12.1) that we must have

$$|x_0 - x'| > c_0(t_0 - t_1' - C) = c_0\Big(t_0 - t' - \frac{1}{\lambda_0}|\log\varepsilon_0| - \frac{\gamma_3}{\lambda_0}(t_0-t') - C\Big) \geq \frac{c_0}{2}(t_0-t') - C. \tag{4.12.6}$$

Now, (4.12.2) follows. □

4.13 A bound on the averages

The last and key step is to prove a bound on the space-time averages of the function $v(t,x)$. As we have seen in Corollary 4.8.1, such bound would provide a global in time point-wise bound on $v(t,x)$ as well.

Proposition 4.13.1. *There exists a constant $C > 0$ such that for any parabolic cylinder $Q_1(t_0, x_0)$, with $x_0 \in \mathbb{R}^n$ and $t_0 > 1$, we have*

$$|(v)_{Q_1}| \leq C. \tag{4.13.1}$$

The rest of this section contains the proof of Proposition 4.13.1.

4.13.1 Reduction to a Hölder in time bound

Another elementary calculation gives the following representation for the function $v(t,x)$:

$$v(t,x) = \overline{v}(t,x) + F(t,x) + (\kappa - \nu)\frac{1}{\kappa}(\partial_t J - F). \tag{4.13.2}$$

Here, $\overline{v}(t,x)$ is the solution to the heat equation:

$$\overline{v}_t = \kappa\Delta\overline{v}, \quad \overline{v}(0,x) = v_0(x), \tag{4.13.3}$$

the function $F(t,x)$ is the solution to (note that this equation has diffusivity ν and not κ)

$$\begin{aligned} &F_t = \nu\Delta F + ug(v), \\ &F(0,x) = 0, \end{aligned} \tag{4.13.4}$$

and is uniformly bounded because the function

$$u(t,x) = \overline{u}(t,x) - F \tag{4.13.5}$$

is non-negative, while $J(t,x)$ satisfies

$$J_t - \kappa \Delta J = F, \tag{4.13.6}$$

with the initial condition $J(0,x) = 0$. Therefore, for any parabolic cube Q we have

$$(v)_Q \leq C(1 + (\partial_t J)_Q). \tag{4.13.7}$$

The advantage of this representation can be seen from writing

$$\begin{aligned} (\partial_t J)_{Q_R} &= \frac{1}{|Q|} \int_{|x-x_0|<R} \int_{|t-t_0|<R^2} \partial_t J(t,x) dt dx \\ &= \frac{1}{R^{n+2}|Q_1|} \int_{|x-x_0|<R} \big(J(t_0 + R^2, x) - J(t_0 - R^2, x) \big)\, dx. \end{aligned} \tag{4.13.8}$$

Thus the boundedness of the averages of $v(t,x)$ and the conclusion of Proposition 4.13.1 would follow from the following Hölder in time bound on $J(t,x)$.

Lemma 4.13.2. *For any $\alpha \in (0,1)$, there exists a constant $C_\alpha > 0$ so that*

$$|J(t_2,x) - J(t_1,x)| \leq C_\alpha |t_2 - t_1|^\alpha, \quad \text{for all } t_1 + 1 \geq t_2 \geq t_1 \geq 2 \text{ and } x \in \mathbb{R}^n. \tag{4.13.9}$$

A word of caution is that one can not deduce the uniform in time Hölder bound on $J(t,x)$ simply from the forced equation (4.13.6) by using the L^∞-boundedness of F. This would lead to a bound that grows as $\log t$ as $t \to +\infty$. Thus, one need to exploit some information about $F(t,x)$ other than the L^∞-bound. Interestingly, it is in the proof of this lemma that the assumption $g(v) \geq c_1 v$ for $0 \leq v \leq 1$ on the nonlinearity plays the key role even though this looks like a generic parabolic regularity result.

4.14 A bound on the spatial variations of the fuel in a "parabola"

After peeling various easily bounded terms off, and remembering from (4.13.5) that

$$F(t,x) = \overline{u}(t,x) - u(t,x), \tag{4.14.1}$$

the key term to estimate is

$$A(t_1,t_2,x) = \int_{t_1/2}^{t_1 - T} \int_{|z| \leq (t_1 - s)^q} \int_{t_1}^{t_2} h(z) \big(u(s, x - z\sqrt{4\kappa(\tau - s)}) - u(s,x) \big) \frac{d\tau ds dz}{\tau - s}, \tag{4.14.2}$$

with some $q > 1/2$, T large but fixed, and an integrable function

$$h(z) = \Big(-\frac{n}{2} + \frac{|z|^2}{4} \Big) e^{-|z|^2/4}. \tag{4.14.3}$$

The reason for what seems to be an inevitable $\log t_1$ growth in the Herrero-Lacey-Velasquez argument is the term $(\tau - s)^{-1}$ that decays too slowly in $(\tau - s)$. One needs to gain an extra power of $\tau - s$ in the denominator and for that it is not sufficient to use the uniform in time L^∞-bound on $u(s,x)$.

On the other hand, it is not clear how generic parabolic regularity would give such additional decay that can only come from the difference

$$u(s, x - z\sqrt{4\kappa(\tau - s)}) - u(s, x) \tag{4.14.4}$$

in the integrand. To estimate $A(t_1, t_2, x)$, we need to show that the variations of the fuel concentration $u(s, y)$ are small for most times s in the time interval $t_1/2 \leq s \leq t_1$ inside the domain of integration in (4.14.2) that we will call a "parabola". We also define, for all $s \leq t_1 - T$ the region

$$P_s = \left\{(s', y') : \ 0 \leq s' \leq s, \ |y' - x| \leq \sqrt{4\kappa}(t_1 - s')^{1/2+q}\right\} \tag{4.14.5}$$

for the part of the parabola at times earlier than s that we will refer to as sub-parabola.

An intuitive reason for the smallness of oscillations inside the parabola is that until the front arrives into this region, the fuel is not consumed while its initial variations are quickly uniformized by the diffusion. Once the front arrives, it spreads through the parabola at a linear speed, consuming the fuel, because of tthe hair trigger effect. The fuel variations during this period can not be controlled but this time interval is relatively short, also because of the hair trigger effect. After the front has passed, there is no fuel left in the parabola, hence its variations are small again.

According to the above intuition, given a point $x \in \mathbb{R}^n$ and a time t_1 sufficiently large we need to consider three cases: (1) the front "has reached the parabola not long before t_1", (2) the front "has reached the parabola long before t_1", (3) the front "has not reached the parabola by the time t_1".

To quantify the above informal notions, let us first assume that there is a time $s' \in [t_1/2, t_1 - T]$ such that there exists a point x' inside the parabola, so that

$$|x - x'| \leq \sqrt{4\kappa}(t_1 - s')^{1/2+q}, \tag{4.14.6}$$

and the spatial average of the temperature around x' is not small, in the sense that

$$\int_{|y-x'|\leq 1} v(s', y)dy \geq \frac{1}{(t_1 - s')^m}, \tag{4.14.7}$$

with $m > 0$ to be chosen. This is the regime when the front has reached the parabola before the time t_1. Then, we let s_0 be the smallest such time in the interval $[t_1/2, t_1 - T]$ and x_0 be the corresponding spatial point:

$$|x - x_0| \leq \sqrt{4\kappa}(t_1 - s_0)^{1/2+q}, \tag{4.14.8}$$

and

$$\int_{|y-x_0|\leq 1} v(s_0, y)dy \geq \frac{1}{(t_1 - s_0)^m}. \tag{4.14.9}$$

We need to distinguish two sub-cases: if $t_1/2 < s_0 \leq t_1 - T$, we will say that the front has reached the parabola "not long before t_1", and if $s_0 = t_1/2$ we say that the front has reached the parabola "long before t_1". If no time s' as defined above exists in the time interval $[t_1/2, t_1 - T]$, we will say that the front has not reached the parabola by the time t_1.

The front has reached the parabola not long before the time t_1

We will only describe the most interesting case when $t_1/2 < s_0 \leq t_1 - T$. Informally, in that case we may think of s_0 as the first time reaction starts in the region around the point x. In that situation, we have equality in (4.14.9):

$$\int_{|y-x_0|\leq 1} v(s_0, y)dy = \frac{1}{(t_1 - s_0)^m}. \tag{4.14.10}$$

Corollary 4.11.1 implies that then the fuel concentration $u(s,y)$ obeys an upper bound

$$u(s,y) \le Ke^{-\gamma_2(s-s_1)}, \quad \text{for all } s \ge s_1 = s_0 + m\log(t_1 - s_0) + C \text{ and } |y - x_0| \le c_2(s - s_1). \tag{4.14.11}$$

Note that, as we have $s_0 < t_1 - T$, we know that $s_1 < t_1 - T/2$, as long as we choose T sufficiently large, depending on the constants m and $C > 0$ in (4.14.11).

Next we take the time $s_2 \ge s_1$ such that

$$c_2(s_2 - s_1) = 10\sqrt{4\kappa}(t_1 - s_0)^{1/2+q}. \tag{4.14.12}$$

Once again, because $s_0 < t_1 - T$, we have $s_2 < t_1 - T/2$, as long as we choose T sufficiently large, depending on m and $C > 0$. Physically, this is the time it takes the reaction to sweep across the region of integration in the definition of $A(t_1, t_2, x)$.

Consider a time $s \in [s_2, t_1]$ and $y \in \mathbb{R}^n$ such that

$$|y - x| \le \sqrt{4\kappa}(t_1 - s)^{1/2+q}. \tag{4.14.13}$$

Then, due to (4.14.8), (4.14.12) and (4.14.13), we have

$$|y - x_0| \le \sqrt{4\kappa}(t_1 - s)^{1/2+q} + \sqrt{4\kappa}(t_1 - s_0)^{1/2+q} \le c_2(s_2 - s_1) \le c_2(s - s_1). \tag{4.14.14}$$

We deduce from (4.14.11) and (4.14.14) that the fuel concentration inside the parabola is, indeed, very small for $s > s_2$:

$$u(s,y) \le Ke^{-\gamma_2(s-s_1)}, \quad \text{for all } s \in [s_2, t_1] \text{ and } |y - x| \le \sqrt{4\kappa}(t_1 - s)^{1/2+q}. \tag{4.14.15}$$

According to the above scenario, we estimate $A(t_1, t_2, x)$ as

$$|A| \le A_e + A_m + A_\ell, \tag{4.14.16}$$

with

$$A_e(t_1,t_2,x) = \int_{t_1/2}^{s_0} \int_{|z| \le (t_1-s)^q} \int_{t_1}^{t_2} |h(z)| \big| u(s, x - z\sqrt{4\kappa(\tau - s)}) - u(s,x) \big| \frac{d\tau ds dz}{\tau - s}, \tag{4.14.17}$$

$$A_m(t_1,t_2,x) = \int_{s_0}^{s_2} \int_{|z| \le (t_1-s)^q} \int_{t_1}^{t_2} |h(z)| \big| u(s, x - z\sqrt{4\kappa(\tau - s)}) - u(s,x) \big| \frac{d\tau ds dz}{\tau - s}, \tag{4.14.18}$$

and

$$A_\ell(t_1,t_2,x) = \int_{s_2}^{t_1 - T/2} \int_{|z| \le (t_1-s)^q} \int_{t_1}^{t_2} |h(z)| \big| u(s, x - z\sqrt{4\kappa(\tau - s)}) - u(s,x) \big| \frac{d\tau ds dz}{\tau - s}. \tag{4.14.19}$$

We have extended the domain of integration in s in the term A_ℓ to the upper limit $s = t_1 - T/2$ simply to avoid considering separately the cases $s_2 < t_1 - T$ and $s_2 > t_1 - T$, while we do know that $s_2 < t_1 - T/2$.

The "late" term A_ℓ can be bounded using the fact that $u(t,x)$ is small in the parabola at the late stage:

$$\begin{aligned} |A_\ell(t_1,t_2,x)| &\le C \int_{s_2}^{t_1 - T/2} \int_{t_1}^{t_2} e^{-\gamma_2(s-s_1)} \frac{d\tau ds dz}{\tau - s} \le C \int_{s_1}^{t_1} e^{-\gamma_2(s-s_1)} \big(\log(t_2 - s) - \log(t_1 - s) \big) ds \\ &= C \int_{s_1}^{t_1 - T/2} e^{-\gamma_2(s-s_1)} \log\Big(\frac{t_2 - s}{t_1 - s}\Big) ds = C \int_{s_1}^{t_1 - T/2} e^{-\gamma_2(s-s_1)} \log\Big(1 + \frac{t_2 - t_1}{t_1 - s}\Big) ds. \end{aligned} \tag{4.14.20}$$

As $0 \le t_2 - t_1 \le 1$ and $t_1 - s \ge T/2$, we can estimate the integral above as

$$|A_\ell(t_1, t_2, x)| \le C \int_{s_1}^{t_1 - T/2} e^{-\gamma_2(s-s_1)} \frac{t_2 - t_1}{t_1 - s} ds \le \frac{C}{T} |t_2 - t_1|. \tag{4.14.21}$$

For the "middle" term A_m, the key point is that it can not last too long because reaction has to spread at a constant speed. More precisely, it follows from the definitions of the times s_1 and s_2 in (4.14.11) and (4.14.12) that for any $\varepsilon > 0$, if we choose $T > 0$ sufficiently large, we have

$$1 \ge \frac{t_1 - s_2}{t_1 - s_0} = \frac{t_1 - s_0 - (s_2 - s_1 + s_1 - s_0)}{t_1 - s_0} \ge 1 - \varepsilon. \tag{4.14.22}$$

Therefore, we have

$$\begin{aligned} |A_m(t_1, t_2, x)| &\le \int_{s_0}^{s_2} \int_{t_1}^{t_2} \frac{d\tau ds}{\tau - s} = \int_{s_0}^{s_2} [\log(t_2 - s) - \log(t_1 - s)] \\ &= \int_{t_1 - s_2}^{t_1 - s_0} \big(\log(t_2 - t_1 + s) - \log s \big) ds \le C \int_{t_1 - s_2}^{t_1 - s_0} \frac{t_2 - t_1}{s} ds \\ &= C(t_2 - t_1) \log \Big(\frac{t_1 - s_0}{t_1 - s_2} \Big) \le C(t_2 - t_1). \end{aligned} \tag{4.14.23}$$

The last step is to estimate the early contribution A_e. To do this, we will estimate the Hölder constant of the fuel concentration $u(t, x)$ inside the parabola for times s in the interval $t_1/2 < s < s_0$. The idea is as follows. We already know that the temperature $v(s, y)$ is small in all of the sub-parabola P_{s_0}. The parabolic regularity estimates will also imply that the gradient $\nabla v(s, y)$ is small in the sub-parabola P_{s_0}. Therefore, the fuel concentration is a solution to a parabolic equation of the form

$$u_t = \Delta u - V(t, x) u, \tag{4.14.24}$$

with the potential $V(t, x) = g(v(t, x))$ that is very small and positive and also has a very small gradient in any sub-parabola P_s, with $0 \le s \le s_0$. In addition, if we take a time $s < s_0$, and a point $y \in P_{s_0}$ that is not too close to the boundary of the parabola, then we will be able to fit a parabolic cylinder $Q_R(s, y)$ of a radius $R \sim (t_1 - s)^{1/2+q'}$, with some $0 < q' < q$, inside the parabola P_s. Hence, the fuel concentration $u(t, x)$ will basically solve the heat equation inside a very large parabolic cylinder $Q_R(s, y)$. This will show that $|\nabla u(s, y)| \sim O(R^{-1})$, and such bound will suffice for us to estimate A_e.

5 Lecture 4: Diffusion of knowledge models in macroeconomics

This is a joint work with H. Berestycki, A. Novikov and J.-M. Roquejoffre.

5.1 The Lucas-Moll model

Our goal today is to discuss the behavior of the solutions to the Lucas-Moll system that consists of an evolution equation for the cumulative probability distribution function of the agents

$$\frac{\partial F}{\partial t} - \kappa \frac{\partial^2 F}{\partial x^2} = F(t, x) \int_{-\infty}^{x} \alpha(s^*(t, y)) \big(- F_y(t, y) \big) dy, \tag{5.1.1}$$

together with an HJB equation for the value function $V(t,x)$

$$\rho V(t,x) = \frac{\partial V(t,x)}{\partial t} + \kappa \frac{\partial^2 V(t,x)}{\partial x^2} + \max_{s\in[0,1]} \Big[(1-s)e^x + \alpha(s) \int_x^\infty [V(t,y) - V(t,x)](-F_y(t,y))dy \Big]. \tag{5.1.2}$$

The strategy $s^*(t,x)$ that appears in (5.1.1) is the maximizer in (5.1.2).

This system was introduced by Robert Lucas and Ben Moll and, independently, Jesse Perla and Chris Tonetti introduced a discrete version of that model, both in 2014.

A key feature of the system (5.1.1)-(5.1.2) is that, on one hand, the density equation (5.1.1) is run forward in time: we need to prescribe an initial condition $F(0,x) = F_0(x)$. On the other hand, the HJB equation (5.1.2) is run backward in time: we prescribe the terminal value $V(T,x) = V_T(x)$. This is a common feature for mean-field games problems that makes the study of their long time behavior and even uniqueness of the solutions extremely challenging. This is also exactly what happens for the Euler-Lagrange trajectories in various problems in the control theory and, indeed, these are often exactly the same kinds of problems.

Most of the lecture will be concerned with the derivation of the system and then we will describe an interesting regime of what we have called "the lottery society".

5.2 The basic models of diffusion of knowledge

The diffusion of knowledge type dynamics in macroeconomics attempts to provide a basic mechanism that leads to the growth of an economy over long time periods. They are vastly simplified and should probably not be taken too literally. However, they equally likely capture some of the basic phenomena, and also seem to be challenging and interesting mathematically. I should also mention that very recently Mark Staley did a comparison of some of the predictions of a very basic knowledge diffusion model to real data and claims a good agreement with Fisher-KPP convergence to a traveling wave. This may indicate that the models are applicable to some extent. There are very few rigorous results currently available on these models, and the goal of this talk is mostly to focus on the models themselves, highlighting some of their mathematical aspects and the open questions.

To the best of my knowledge, the Pickwickians were the first to think of the importance of the diffusion of knowledge both and learning itself and of its connection to production: "That while this Association is deeply sensible of the advantages which must accrue to the cause of science, from the production to which they have just adverted – no less than from the unwearied researches of Samuel Pickwick, Esq., G.C.M.P.C., in Hornsey, Highgate, Brixton, and Camberwell – they cannot but entertain a lively sense of the inestimable benefits which must inevitably result from carrying the speculations of that learned man into a wider field, from extending his travels, and, consequently, enlarging his sphere of observation, to the advancement of knowledge, and the diffusion of learning." (Charles Dickens, "The Pickwick Papers", Chapter 1.)

5.3 The dynamics of agents

We now describe the basic underlying dynamics in the diffusion of learning models. Consider a population of $N \gg 1$ of agents (which we can think of as firms in terms of economics, or as individual workers, as in the Staley analysis of data), such that each agent A_k is assigned a certain knowledge $z_k(t) \geq 0$ that varies in time. Of course, the notion that knowledge is one-dimensional is very simplistic. In reality, the expertise of a company is a highly multi-dimensional object but we will completely ignore this. To partially justify this approach, we will usually think of knowledge as simply the overall productivity of a company, which can make its one-dimensionality more realistic. In any case, the one-dimensional case needs to be understood first, and this will be the focus of the

talk today. The informal definition of productivity is that the value of the goods that an agent with a current productivity $z(t) \geq 0$ would produce during a time interval $[t, t+\Delta t]$ is $z(t)\Delta t$, if the agent focuses solely on production during that time interval.

If an agent's goal is to maximize the value of its total production over time, with a proper discounting of the future production, then it may be of a benefit not just to solely produce while remaining at a fixed knowledge value but also to spend some time increasing its productivity via learning even if at the cost of temporarily reducing the current production volume, and thus the value of the produced goods. This is one of the economic concepts behind the existence of universities.

The two basic modeling issues in the above dynamics are
(1) how to model the acquisition of knowledge,
(2) how to balance the future benefit of the acquired knowledge and its current cost.

To model the balance between knowledge acquisition and production, we will assume that at each moment of time an agent can either produce or learn. To quantify this, we will denote by $s_k(t) \in [0,1]$ the fraction of time an agent with a knowledge $z_k(t) > 0$ spends learning on a time interval $[t, t+\Delta t]$, so that $(1 - s_k(t))$ is the fraction of time it spends producing on this time interval. We will refer to $s_k(t)$ as the learning strategy of the corresponding agent.

5.3.1 Acquisition of knowledge

We first describe how the knowledge is acquired and postpone the production aspect of this dynamics until later. As we will see, as long as the learning strategies $s_k(t)$, $k = 1, \dots, N$, are fixed, the learning acquisition dynamics is not far from the Brunet-Derrida type models in branching processes.

Agents in the economy learn by searching for other agents, with a higher production knowledge, and then possibly acquiring their knowledge if they do meet them. The outcome of the search is random in two aspects: (i) a search can either be successful or not, with the probability of a successful search that we describe below, and (ii) if a search is successful, the "discovered" agent $z_j(t)$ is chosen uniformly among all other agents. If $z_j(t) > z_k(t)$ then the agent $z_k(t)$ jumps to the location $z_j(t)$ of that other agent. Otherwise, $z_k(t)$ stays put. The searching agent does not control which other agent it would find in the search.

We assume that the probability of a successful search depends solely on the fraction $s_k(t)$ of time that an agent with the current knowledge $z_k(t)$ allocates to learning. More precisely, the probability that a search during a time interval $(t, t+\Delta t)$, by an agent with the current knowledge $z_k(t) \geq 0$, is successful, is $\alpha(s_k(t))\Delta t$. Here, $\alpha(s) : [0,1] \to \mathbb{R}^+$ is a given increasing function such that $\alpha(0) = 0$. The "law of diminishing returns" implies that we should choose $\alpha(s)$ to be a concave function. In addition, we will often assume that

$$\alpha'(0) = +\infty. \tag{5.3.1}$$

This encourages the "very elite" agents to still spend a positive fraction of time searching. A typical example to keep in mind is $\alpha(s) = \alpha_1\sqrt{s}$ with some $\alpha_1 > 0$. We will assume that for most of this lecture.

Alternatively, one can think that each individual agent $z_k(t)$ has an exponential clock attached to it that runs at the rate $\alpha(s_k(t))$. If the clock rings then the particle chooses another agent $z_j(t)$, uniformly among all agents, and jumps to the location $z_j(t)$ if $z_j(t) > z_k(t)$.

An important feature in this model is that the strategy $s_k(t)$ typically depends not only on the current knowledge $z_k(t)$ of a given agent who is performing the search but also on the positions of the other agents. For example, one possible strategy is

$$s_k(t) = \frac{1}{N}\#\{z_m(t) \geq z_k(t),\ 1 \leq m \leq N\}, \tag{5.3.2}$$

which says that agents with a smaller knowledge search more than the advanced agents. A smoother version of that choice is

$$s_k(t) = \frac{1}{N}\sum_{k=1}^{N} \zeta(z_m(t) - z_k(t)), \tag{5.3.3}$$

with a smooth increasing function $\zeta(z)$ such that $\zeta(z) = 0$ for all $z \le 0$ and $\zeta(z) = 1$ for $z \ge L_1$, with some L_1 fixed.

Another interesting for us case is

$$s_k(t) = \min\Big[1, \frac{1}{N}\sum_{m:z_m(t)\ge z_k(t)} \Big(\frac{z_m(t)}{z_k(t)} - 1\Big)\Big], \tag{5.3.4}$$

or, more generally,

$$s_k(t) = \min\Big[1, \frac{1}{N}\sum_{m:z_m(t)\ge z_k(t)} \Big(\frac{\varphi(z_m(t))}{\varphi(z_k(t))} - 1\Big)\Big], \tag{5.3.5}$$

with a given positive increasing function $\varphi(z)$. With all of the above strategies, the agents that are "lagging behind" search more than the "advanced agents".

As we have mentioned, given that a search is successful, the agent B that the agent A meets is simply chosen uniformly among all other agents. If the productivity z_A of agent A is lower than the productivity z_B of agent B, then agent A updates its production knowledge to that of agent B, so that $z_A(t^+) = z_B$. If $z_A \ge z_B$, then the productivity of A does not change, and $z_A(t^+) = z_A(t^-)$.

A crucial modeling assumption here is that the choice of the "knowledge giver" agent B does not depend on the initial gap between the productivities of A and B. In particular, the "instantaneous" jump in knowledge can be arbitrarily large if $z_B(t^-) \gg z_A(t^-)$. This can be restricted in many ways but we will not focus on this issue. Let us just comment that, even if we relax this assumption, in order to keep the agents spreading at a positive speed, we will need to allow some very large jumps of productivity, with a reasonably large probability. In other words, the tail distribution of the jumps needs to be sufficiently heavy. Otherwise, if the interactions are local, then the system will only exhibit diffusive spreading like $O(\sqrt{t})$, while in the real world economies grow at a constant rate in time.

Let us also comment that, as is common in economics, we will consider the dynamics in logarithmic variable $x = \log z$.

5.3.2 The macroscopic dynamics

Let us emphasize that even though the probabilistic descriptions of the models are discussed for a finite $N \gg 1$, all our results are obtained in a "mean-field" macroscopic description, already in the limit $N \to +\infty$. We will not discuss and have not addressed the passage to the mean-field limit at all. This, in itself, is a nice open problem.

Let us now describe heuristically the macroscopic dynamics in the limit $N \to +\infty$. We let $\psi(t, x)$ be the probability density of the agents at a time $t \ge 0$ obtained in the limit $N \to +\infty$ (recall that $x = \log z$ is the logarithmic variable). Then, the probability that the agent B, who A encounters, has knowledge in the interval $(x', x' + \Delta x')$ is $\psi(t, x')\Delta x'$. We will assume that the agents use a common search strategy $s_c(t, x)$ that depends only on the log-productivity $x \ge 0$ and time $t \ge 0$, but not on an individual agent. Once $s_c(t, x)$ is prescribed, the evolution of the density of the agents is governed by an integral equation

$$\frac{\partial\psi(t,x)}{\partial t} = \int_{-\infty}^{x} \alpha(s_c(t,y))\psi(t,y)\psi(t,x)dy - \int_{x}^{\infty} \alpha(s_c(t,x))\psi(t,x)\psi(t,y)dy. \tag{5.3.6}$$

The first term in the right side is due to the agents at positions $y \leq x$ who execute a successful search and end up at knowledge x, and the second one comes from the agents with knowledge x who do a successful search and move up to a level $y \geq x$.

5.3.3 Modeling the production

Let us now describe the cost of learning and the production process even though it will not be immediately important for us, until we introduce the value function. If $s(t,x) \in [0,1]$ is the fraction of time an agent with a knowledge $z = e^x \geq 0$ spends learning on a time interval $[t, t+\Delta t]$, then $(1-s(t,x))$ is the fraction of time it spends producing on this time interval. Its total production between the times t and $t+\Delta t$ is then not $e^x \Delta t$ but

$$[1 - s(t,x)]e^x \Delta t. \tag{5.3.7}$$

In other words, in this model, the sole cost of learning is in the complete suspension of the production during the learning process. Otherwise, we would need to reduce the above expression by the extra cost of learning. This can be done, in principle, but we will not discuss this aspect today.

5.3.4 Innovation and diffusion of knowledge

Another important component of the dynamics is that, independent of the learning from other agents, each agent also attempts to innovate internally, a process that has nothing to do with the distribution of the other agents. Innovation can lead both to an improvement and a regression, and is often modeled by a Brownian motion. Let us comment that Mark Staley's data analysis shows that, indeed, the mean drift associated to that diffusion is zero. Once again, for simplicity, we will disregard the cost of innovation, though some of it may be attributed to the "regression due to innovation" inherent in the Brownian motion.

Note that this diffusion via innovation is very different from the Pickwickian diffusion of knowledge: here, diffusion is an internal process for each given agent. On the other hand, in the Pickwickian model, diffusion of one agent is crucial for the learning of the other agents: there, an agent performs a Brownian motion (note that Brownian motion is an excellent model for the travels of Mr. Pickwick) and transmits knowledge to agents it encounters on the way. The Pickwickian model is also an interesting interacting particle system but is very different from the diffusion of knowledge models we consider here.

The presence of innovation modifies the macroscopic evolution equation (5.3.6) for the common agent density $\psi(t,x)$ to

$$\frac{\partial \psi(t,x)}{\partial t} = \kappa \frac{\partial^2 \psi(t,x)}{\partial x^2} + \int_{-\infty}^{x} \alpha(s_c(t,y))\psi(t,x)\psi(t,y)dy - \int_{x}^{\infty} \alpha(s_c(t,x))\psi(t,x)\psi(t,y)dy, \tag{5.3.8}$$

with a prescribed initial condition $\psi_c(0,x) = \psi_0(x)$.

5.3.5 Evolution of the cumulative distribution function

Equation (5.3.8) has the form of a nonlinear Boltzmann equation from the kinetic theory with an extra diffusion term. A very different point of view is based on the evolution of the cumulative distribution function

$$F(t,x) = \int_{x}^{\infty} \psi(t,z)dz. \tag{5.3.9}$$

A simple computation shows that $F(t,x)$ satisfies a non-local equation

$$\frac{\partial F(t,x)}{\partial t} = \kappa \frac{\partial^2 F(t,x)}{\partial x^2} + F(t,x) \int_{-\infty}^{x} \alpha(s_c(t,z))(-F_z(t,z))dz, \tag{5.3.10}$$

with the limits at infinity

$$\lim_{x\to-\infty} F(t,x) = 1, \quad \lim_{x\to+\infty} F(t,x) = 0, \quad \text{for each } t \geq 0 \text{ fixed.} \tag{5.3.11}$$

Let us recall that if, as the result of a successful search, an agent A with a knowledge x_A discovers an agent B with a location x_B, then the agent A jumps to the location x_B only if $x_B > x_A$. This is what creates the positive integral term in the right side of (5.3.10) – the whole system is spreading to the right.

A remarkable observation is that in the special case $\alpha(s) = \alpha_1 = \text{const}$, this equation reduces exactly to the classical Fisher-KPP equation

$$\frac{\partial F}{\partial t} = \kappa \frac{\partial^2 F}{\partial x^2} + \alpha_1 F(1-F). \tag{5.3.12}$$

This is the case when the exponential clocks of all agents are identical and do not depend on their positions or time. In particular, the only interaction between the agents is through them jumping on top of each other but the search strategy of an agent does not depend on the location of the other agents.

We should briefly mention that the front-like nature and linear in time spreading of the solutions to the Fisher-KPP equation have been used in macroeconomics to model the growth of an economy since at least the 1980's.

The Fisher-KPP equation (5.3.12) arises in the not most natural from the economics perspective special case of (5.3.10) when $\alpha(s)$ is a constant. However, it may still give us some idea as to what may happen in the general case. Let us recall how it is related to branching Brownian motion, with a branching rate $\alpha_1 > 0$ and diffusion coefficient $\kappa > 0$. If $M(t) = \max(X_1(t), X_2(t), \ldots, X_{N_t}(t))$ is the running maximum of such BBM, then the probability distribution function

$$F(t,x) = \mathbb{P}(M(t) > x), \tag{5.3.13}$$

satisfies (5.3.12) with the initial condition $F(0,x) = \mathbb{1}(x \geq 0)$. Thus, when $\alpha(s)$ is a constant, the cumulative distribution of the agents in the macroscopic limit of the knowledge diffusion model behaves exactly as the law of the maximum of the binary BBM.

5.3.6 A digression on a very special non-local case

We are not aware of a natural version of this interpretation for the non-local equation (5.3.10) when $\alpha(s) \neq \text{const}$, without something like a passage to the limit of infinitely many interacting particles. It is possible that such connection exists, via voting models. At least, a non-local Fisher-KPP equation of the form

$$\frac{\partial u(t,x)}{\partial t} = \frac{1}{2}\Delta u(t,x) + \zeta[F(u(t,\gamma^{-1}x)) - u(t,x)], \tag{5.3.14}$$

with a multiplicative convolution term does have such interpretation in terms of a majority voting model for a purely local branching BBM, as shown in our recent joint work with Alex Dunlap. To get (5.3.14), consider the infinite system of PDEs

$$\begin{aligned} \partial_t u_\ell(t,x) &= \frac{1}{2}\gamma^{2\ell}\Delta u_\ell(t,x) + \zeta[F(u_{\ell+1}(t,x)) - u_\ell(t,x)], \quad \ell \in \mathbb{Z},\ t > 0,\ x \in \mathbb{R}; \\ u_\ell(0,x) &= \operatorname{sgn}(x). \end{aligned} \tag{5.3.15}$$

indexed by the generation ℓ. We note by rescaling the system of PDEs (5.3.15), that

$$u_\ell(t,x) = u_0(t,\gamma^{-\ell}x) \qquad \text{for all } \ell \in \mathbb{Z},\ t \geq 0, \text{ and } x \in \mathbb{R}. \tag{5.3.16}$$

Writing now $u = u_0$, we see that u satisfies the single *nonlocal* PDE

$$\begin{aligned} \partial_t u(t,x) &= \frac{1}{2}\Delta u(t,x) + \zeta[F(u(t,\gamma^{-1}x)) - u(t,x)], \quad t > 0,\ x \in \mathbb{R}; \\ u(0,x) &= \operatorname{sgn}(x). \end{aligned} \tag{5.3.17}$$

Note that when the mobility growth factor $\gamma = 1$, the PDE (5.3.17) becomes local. Let us also mention that if were looking for a similar probabilistic description for the solutions to (5.3.17), with an initial condition different from $\operatorname{sgn}(x)$, then the initial conditions for u_ℓ in (5.3.15) would be ℓ-dependent. It is straightforward to adapt the voting model to that situation. We use the $\operatorname{sgn}(x)$-based voting rule for the sake of simplicity.

5.3.7 Similarity to N-BBM

When the search strategy of each agent does depend on the locations of the other agents, the learning process described above has some similarity with a branching process known as N-BBM. This is a version of a branching Brownian motion, introduced by Brunet, Derrida, Mueller and Munier in 2006-07. In that model, N particles perform a binary branching Brownian motion. After one of the particles branches, so that the system momentarily has $N+1$ particles, the particle with the smallest value of a prescribed scoring function is removed. Thus, the overall number $N \gg 1$ of particles stays constant in the dynamics. In one dimension a natural scoring function is simply $s(x) = x$. In higher dimensions, the cases $s(x) = \|x\|$ and $s(x) = 1/\|x\|$ have been considered.

In the present context, the N-BBM corresponds to a version of the learning dynamics where only the agent with the smallest knowledge can learn, so that $s(x_k(t)) = 0$ unless $x_k(t) = \min_{1\leq j\leq N} x_j(t)$. As a consequence, when its search is successful, the smallest particle does jump to the discovered agent with probability one. Since the discovered agent is chosen uniformly among all agents, this is equivalent to choosing uniformly a particle that branches and removing the smallest one, which is the N-BBM process.

More generally, one can think of the learning process we described above as the following modification of N-BBM. When one of the N particles, say a particle at a location $x_k(t)$, branches, instead of removing from the system the particle with the smallest value of the scoring function, we do the following. First, another particle $x_j(t)$ is chosen uniformly among all N particles. If $s(x_j(t)) > s(x_k(t))$, then we place the newborn particle at the location $x_j(t)$ and remove from the system its parent. This leaves two particles at the location $x_j(t)$ and none at $x_k(t)$. On the other hand, if $s(x_j(t)) < s(x_k(t))$ then the newborn particle is removed from the system, and we have one particle at $x_j(t)$ and one particle at $x_k(t)$, as if the branching event did not happen. The total number of particles stays fixed in both cases.

There are relatively few rigorous results about N-BBM. The mean-field limit as $N \to +\infty$, an analog of (5.3.8), for a jump-process version of N-BBM, was proved by Daniel Remenick and Rick Durrett (2011) for a random walk version, and then for a heavy-tailed version of the jumps by Jean Berard and Pascal Maillard (2014). The hydrodynamic limit for N-BBM itself was analyzed in a series of papers by De Masi, Ferrari, Presutti and Soprano-Loto. These papers used the scoring function $s(x) = x$ in $d = 1$ (leftmost particles are always removed). For the N-BBM process the

macroscopic limit is a free boundary problem

$$\begin{aligned} &\partial_t \phi = \kappa \partial_x^2 \phi + \phi, \quad x > \Gamma(t), \\ &\int_{\Gamma(t)}^{\infty} \phi(t,x) dx = 1. \end{aligned} \tag{5.3.18}$$

The well-posedness of (5.3.18) was proved by Julien Berestycki, Eric Brunet, Jim Nolen and Sarah Penington (2019). The crucial observation of this paper is that the cumulative distribution function

$$F(t,x) = \int_x^{\infty} \phi(t,y) dy, \tag{5.3.19}$$

of the limit of N-BBM as $N \to +\infty$ is also the limit as $n \to +\infty$ of the solutions to

$$\partial_t F_n = \partial_x^2 F_n + F_n - F_n^n, \quad x \in \mathbb{R}. \tag{5.3.20}$$

As a side remark, it would be very interesting to understand if there is a direct connection between the N-BBM and the voting model corresponding to F_n as both of them have the same limit, in the former case as $N \to +\infty$, and in the latter as $n \to +\infty$.

Multidimensional N-BBM. One of the most interesting properties of N-BBM was shown by Nathanael Berestycki and Lee Zhao (2018). They prove that in dimensions $d > 1$ the N-BBM particles with the scoring function $\|x\|$ move, as a clump, with a deterministic speed but in a random direction chosen uniformly on the unit sphere. One would expect a similar phenomenon in more general learning processes in $d > 1$ but that seems to be a difficult question especially when the learning strategy is chosen via an optimization problem, which is what we will discuss below.

J. Berestycki, Brunet, Nolen and Penington (2021,2022) also studied a kind of N-BBM in $d > 1$ with the scoring function $s(x) = 1/\|x\|$, a system known as Brownian bees, and $N \to +\infty$ is referred to as the infinite swarm limit. There, the particle with the largest norm is removed after a branching event. This leads to the particles staying in a ball of a fixed radius $R_\infty < +\infty$ as $t \to +\infty$, so that the system does not spread.

5.3.8 Possible prescriptions of a strategy

Let us comment on what would constitute a sensible choice of the search strategy $s_c(t,x)$. The very unproductive agents should be learning almost all the time, or even all the time. On the other hand, the very advanced agents would spend very little time learning because for them the chances of meeting a more knowledgeable agent are very small. Internal innovation gives them a higher chance of a technological advance than an outside search. Hence, the search strategy $s_c(t,x)$ should have the limits

$$s_c(t,-\infty) = 1, \quad s_c(t,+\infty) = 0. \tag{5.3.21}$$

To capture the above intuition, one natural choice is to postulate that the common search strategy is given by

$$s_c(t,x) = F(t,x). \tag{5.3.22}$$

This corresponds to the rank-based search strategy we have seen in (5.3.2):

$$s_k(t) = \frac{1}{N} \#\{z_m(t) \geq z_k(t),\ 1 \leq m \leq N\}, \tag{5.3.23}$$

for a finite $N \gg 1$. In that case, the non-local Fisher-KPP equation (5.3.10)

$$\frac{\partial F(t,x)}{\partial t} = \kappa \frac{\partial^2 F(t,x)}{\partial x^2} + F(t,x) \int_{-\infty}^{x} \alpha(s_c(t,z))(-F_z(t,z)) dz, \tag{5.3.24}$$

becomes local:

$$\frac{\partial F}{\partial t} = \kappa \frac{\partial^2 F}{\partial x^2} + F\big(Q(1) - Q(F)\big), \tag{5.3.25}$$

with the nonlinearity

$$Q(u) = \int_0^u \alpha(u)du. \tag{5.3.26}$$

Equation (5.3.25) is a reaction-diffusion equation of the Fisher-KPP type and its solutions converge to the minimal speed traveling wave as $t \to +\infty$, moving with the speed

$$c_* = 2\sqrt{Q(1)} = 2\Big(\int_0^1 \alpha(u)du\Big)^{1/2}. \tag{5.3.27}$$

However, (5.3.22) is the only known choice of the common search strategy $s_c(t,x)$ when such result on spreading is readily available. Even for the slightly non-local search strategies

$$s_c(t,x) = \int_{\mathbb{R}} F(t, x-y)\zeta(y)dy, \tag{5.3.28}$$

with a nice rapidly decaying kernel $\zeta(x) \geq 0$, with $\|\zeta\|_{L^1} = 1$, the long time convergence of the solutions to (5.3.10) to a traveling wave is not known. This choice of strategy corresponds to (5.3.3) for a finite $N \gg 1$.

Linearization dependent on the solution. One serious difficulty for the choice (5.3.28) and other strategies different from (5.3.22) is that the linearization in the tail of an equation of the form (5.3.24) is

$$\frac{\partial u}{\partial t} = \kappa \frac{\partial^2 u}{\partial x^2} + R(t)u, \tag{5.3.29}$$

with the reaction rate

$$R(t) = \int_{-\infty}^{\infty} \alpha(s_c(t,z))(-F_z(t,z))dz \tag{5.3.30}$$

that still depends on the solution $F(t,x)$ of the full nonlinear problem (5.3.10), unlike for the Fisher-KPP type problem (5.3.25) when the linearization has the form

$$\frac{\partial F}{\partial t} = \kappa \frac{\partial^2 F}{\partial x^2} + Q(1)F, \tag{5.3.31}$$

with the coefficient $Q(1)$ that does not depend on the solution.

On the other hand, when the linearization has the form (5.3.29)-(5.3.30), one expects that the solution converges to a traveling wave $U(x-ct)$ and the strategy converges to the traveling wave strategy

$$s_c(t,x) \to S(x-ct), \quad S(x) = \int_{\mathbb{R}} U(x-y)\zeta(y)dy. \tag{5.3.32}$$

Then, the coefficient $R(t)$ defined by (5.3.30) should have the limit

$$R_\infty = \lim_{t\to+\infty} R(t) = \int_{-\infty}^{\infty} \alpha(S(z))(-U_z(z))dz, \tag{5.3.33}$$

and the solution should propagate at the speed

$$c = 2\sqrt{R_\infty}. \tag{5.3.34}$$

None of that is known at the moment.

Prescribed environments. Alternatively to (5.3.21), one can consider the strategy $s_c(t,x)$ as an environment that can be more or less favorable for jumps. Then, one can take the probability $\alpha(s_c(x))$ as, for instance, a periodic function in x, or a space-time stationary function of t and x, leading to an equation of the form

$$\frac{\partial F(t,x)}{\partial t} = \kappa \frac{\partial^2 F(t,x)}{\partial x^2} + F(t,x) \int_{-\infty}^{x} \alpha(z)(-F_z(t,z))dz. \tag{5.3.35}$$

Then, one expects the particles to spread but also concentrate in the places where $\alpha(x)$ is small as the only possibility to leave such traps is via diffusion.

5.4 The value function and its evolution

So far, we have described the dynamics of the density of agents assuming that the common search strategy $s_c(t,x)$ is given. In macroeconomics models, the search strategies of interest will come from optimization problems that we now describe.

5.4.1 An individual agent in a soup of agents

Let us first imagine that "nearly all" agents use a common search strategy $s_c(t,x)$ and their cumulative distribution function satisfies the non-local Fisher-KPP equation (5.3.10), but a "black market" individual agent follows an individual strategy $s_i(t,x)$ that may differ from $s_c(t,x)$. Then, the cumulative distribution function $G(\tau,y;t,x)$ for the individual agent satisfies a linearized Fisher-KPP type equation

$$\frac{\partial G(\tau,y;t,x)}{\partial \tau} = \kappa \frac{\partial^2 G(\tau,y;t,x)}{\partial y^2} + F(t,y;s_c) \int_{-\infty}^{y} \alpha(s_i(\tau,z))(-G_z(\tau,z;t,x))dz, \tag{5.4.1}$$

for $\tau > t$, with the initial condition at the time $\tau = t$

$$\begin{aligned} &G(\tau = t, y; t, x; s_c, s_i) = 1 \text{ for } y < x, \\ &G(\tau = t, y; t, x; s_c, s_i) = 0 \text{ for } y > x. \end{aligned} \tag{5.4.2}$$

A standard example of such behavior is the evolution of the density of traffic on a highway. The common agents would listen to Waze while the individual agents will look for a better opportunity.

5.4.2 The standard Hamilton-Jacobi control theory setup

Let us recall that in the standard finite-dimensional Hamilton-Jacobi theory, the setup is as follows. A process $X(\tau;t,x;s)$ satisfies an ODE

$$\frac{dX(\tau;t,x;s)}{d\tau} = V(X(\tau;t,x;s);s(\tau)), \quad \tau > t, \quad X(\tau = t) = x. \tag{5.4.3}$$

Here, $s(\tau)$ is a control that is to be chosen. Given the the trajectory $X(\tau)$, we can define its value as

$$Z(t,x;s) = \int_t^T e^{-\rho(\tau - t)} L(X(\tau;t,x;s), s(\tau))d\tau + V_T(X(T)). \tag{5.4.4}$$

Here, $L(x,s)$ is a prescribed Lagrangian. The goal is to choose the control so as to optimize the value function

$$V(t,x) = \sup_{s \in \mathcal{A}} Z(t,x;s), \tag{5.4.5}$$

where $\mathcal{A}$ is some set of admissible controls. The value function satisfies the Hamilton-Jacobi equation

$$\begin{aligned} &\frac{\partial V}{\partial t} + H(x, \nabla V) - \rho V = 0, \quad t < T, \\ &V(T, x) = V_T(x). \end{aligned} \tag{5.4.6}$$

Here, the Hamiltonian $H(x, p)$ is the generalized Legendre transform of the Lagrangian:

$$H(x, p) = \sup_s [L(x, s) + p \cdot V(x, s)]. \tag{5.4.7}$$

You recover the standard Legendre transform when $V(x, s) = s$.

5.4.3 The best individual strategy for a given common strategy

We now describe the choice of the best individual strategy in terms of an optimization problem. Given a common strategy $s_c(t, x)$ and an individual strategy $s_i(t, x)$, we can already find (i) the common density of agents $\psi_c(t, x)$, from the solution to the non-local Fisher-KPP equation (5.3.10), and (ii) the individual transition probability density

$$q(\tau, y; t, x) = -G_y(\tau, y; t, x; s_c, s_i), \tag{5.4.8}$$

from the solution to (5.4.1). With these in hand, we define the average individual production, between the starting time t and a terminal time $T > t$, corresponding to these strategies, as (we set the terminal value $V_T = 0$ for the moment)

$$Z(t, x; s_c, s_i) = \int_t^T \int e^{-\rho(\tau - t)} e^y [1 - s_i(\tau, y)] q(\tau, y; t, x) dy d\tau. \tag{5.4.9}$$

Note that it has exactly the form (5.4.4), if we think of the cumulative distribution function $G(\tau, \cdot)$ as an element of the infinite-dimensional state space. Its evolution is governed by the non-local partial differential equation (5.4.1), which serves as an analog of (5.4.3), and $s_i(t, x)$ is the control for the infinite-dimensional trajectory of the function $G(\tau, \cdot)$. The Lagrangian here is

$$L(G, s_i) = \int e^y [1 - s_i(y)](-G_y(y)) dy. \tag{5.4.10}$$

Let us fix a given common strategy $s_c(t, x)$ followed by the "vast majority" of agents and look for an optimal strategy $s_i(t, x)$ for an individual agent. The agent's goal is to maximize its total production $Z(t, x)$ given by (5.4.9). In other words, we define the individual value function corresponding to a collective strategy s_c as

$$W(t, x; s_c) = \sup_{s_i \in \mathcal{A}_T} Z(t, x; s_c, s_i). \tag{5.4.11}$$

Here, $\mathcal{A}_T$ is the set of all admissible strategies on the time interval $[0, T]$. This function satisfies the Hamilton-Jacobi-Bellman equation

$$\begin{aligned} \rho W(t, x; s_c) = &\frac{\partial W(t, x; s_c)}{\partial t} + \kappa \frac{\partial^2 W(t, x; s_c)}{\partial x^2} \\ &+ \max_{s \in [0,1]} \left(e^x (1 - s) + \alpha(s) \int_x^\infty \psi(t, z; s_c) [W(t, z; s_c) - W(t, x; s_c)] dz \right), \end{aligned} \tag{5.4.12}$$

with $W(T, x; s_c) = V_T(x)$. Given a common strategy $s_c \in \mathcal{A}_T$, we may now define the best individual strategy as the maximizer in (5.4.12):

$$s_i^*(t, x; s_c) = \operatorname{argmax}_{s \in [0,1]} \left(e^x (1 - s) + \alpha(s) \int_x^\infty \psi(t, z; s_c) [W(t, z; s_c) - W(t, x; s_c)] dz \right). \tag{5.4.13}$$

This gives us a map

$$\mathcal{F}_T : \ \mathcal{A}_T \to \mathcal{A}_T, \quad \mathcal{F}[s_c](t,x) = s_i^*(t,x;s_c). \tag{5.4.14}$$

The map $\mathcal{F}_T$ depends on the terminal condition $V_T(x)$ and on the initial condition $\psi_0(x)$.

5.4.4 The Nash equilibrium strategy

If the optimal individual strategy $s_i(t,x)$ found in the above optimization procedure does not coincide with the common strategy $s_c(t,x)$, then it is natural to expect that all agents will "go to the black market" and start shifting to the best individual strategy $s_i(t,x) \neq s_c(t,x)$. Thus, $s_c(t,x)$ will no longer be the strategy used by the "vast majority" of the agents. This, in turn, will modify the best individual strategy $s_i(t,x)$, and so on. One expects that this adjustment would eventually result in the system being in a Nash equilibrium – this is a state where, given the common strategy $s_c(t,x)$, the best individual strategy is also $s_c(t,x)$. That is, a Nash equilibrium is a fixed point of the map $\mathcal{F}_T$ defined by (5.4.14):

$$\mathcal{F}[s_{Nash}] = s_{Nash}. \tag{5.4.15}$$

This, finally, brings us to the Lucas-Moll system that consists of an evolution equation for the cumulative probability distribution function of the agents

$$\frac{\partial F}{\partial t} - \kappa \frac{\partial^2 F}{\partial x^2} = F(t,x) \int_{-\infty}^{x} \alpha(s^*(t,y))\big(- F_y(t,y)\big) dy, \tag{5.4.16}$$

together with an HJB equation for the value function $V(t,x)$

$$\rho V(t,x) = \frac{\partial V(t,x)}{\partial t} + \kappa \frac{\partial^2 V(t,x)}{\partial x^2} + \max_{s\in[0,1]} \Big[(1-s)e^x + \alpha(s) \int_x^{\infty} [V(t,y) - V(t,x)](-F_y(t,y)) dy \Big]. \tag{5.4.17}$$

The strategy $s^*(t,x)$ that appears in (5.4.16) is the maximizer in (5.4.17).

A key feature of the system (5.4.16)-(5.4.17) is that the density equation (5.4.16) is run forward in time: we need to prescribe an initial condition $F(0,x) = F_0(x)$. On the other hand, the HJB equation (5.4.17) is run backward in time: we prescribe the terminal value $V(T,x) = V_T(x)$. This is a common feature for mean-field games problems that makes the study of their long time behavior and even uniqueness of the solutions extremely challenging.

5.5 Balanced growth paths: a candidate for the long time limit

In their paper, Lucas and Moll not only propose the Lucas-Moll system (5.4.16)-(5.4.17) but also study numerically its balanced growth paths (BGP) solutions. These are special solutions to (5.4.16)-(5.4.17) valid for all time, corresponding to a constant growth rate for the economy in equilibrium. They have the form

$$F(t,x) = F_*(x-ct), \quad V(t,x) = e^{ct}Q(x-ct), \quad s^*(t,x) = s^*(x-ct). \tag{5.5.1}$$

In other words, the density of agents and the optimal strategy have the form of a traveling wave moving with a speed $c \in \mathbb{R}$, while the value function is exponentially growing in time.

As we have mentioned, in the very special case when $\alpha(s)$ is independent of s, the cumulative distribution function of the agents is decoupled from (5.4.17) and satisfies the Fisher-KPP equation. This equation admits traveling wave solutions that are stable. The assumption $\alpha(s) = \text{const}$ means that the success of the search does not depend on the fraction of the time spent searching and is not realistic. However, the structure of the full coupled problem without this assumption, for a general

search function $\alpha(s)$, still inherits some Fisher-KPP features that allow us to use a strategy originating in the construction of traveling waves for reaction-diffusion equations, albeit with non-trivial modifications coming from the required estimates for the HJB equation, to analyze the existence of balanced growth paths for the Lucas-Moll system. We have the following result on the existence of balanced growth paths. For simplicity and some explicitness, from now on we assume that

$$\alpha(s) = \alpha_1 \sqrt{s}, \tag{5.5.2}$$

but it is not at all necessary, as long as $\alpha(s)$ is concave and $\alpha''(0) = -\infty$. A parallel result by Alessio Porretta and Luca Rossi (2022) uses completely different methods and does not use the assumption (5.5.2).

Theorem 5.5.1. *[Papanicoloau-R-Velcheva'21,Porretta-Rossi'22]*
(1) If $\rho > \kappa$ and $\alpha_1 > \kappa$, then there exists $c \in (0, 2\sqrt{\kappa\alpha_1})$ so that the system (5.4.16)-(5.4.17) has a solution of the form (5.5.1), such that $F_*(x)$ is monotonically decreasing, $Q(x)$ is monotonically increasing, and

$$\begin{aligned} &\lim_{x\to-\infty} F_*(x) = 1, \ \lim_{x\to+\infty} F_*(x) = 0, \\ &\lim_{x\to-\infty} Q(x) \quad \textit{exists and is positive,} \\ &\lim_{x\to+\infty} (Q(x)e^{-x}) = \frac{1}{\rho-\kappa}. \end{aligned} \tag{5.5.3}$$

In addition, there exists $\eta_\ell \in \mathbb{R}$ so that $s^*(x) = 1$ for all $x < \eta_\ell$ and $s^*(x) < 1$ for all $x > \eta_\ell$.
(2) Any traveling wave solution to (5.4.16)-(5.4.17) satisfies $2\kappa < c < \kappa + \alpha_1$.

The precise range of speeds is not known but we have the following restriction on the range of possible speeds making it different from the Fisher-KPP situation.

Theorem 5.5.2. *(Porretta-Rossi'22) There exists a traveling wave for all $c \in [2\sqrt{\kappa\alpha_1}, \kappa + \alpha_1)$.*

We expect that there exists an interval of speeds $[c_{min}, c_{max})$ so that (5.4.16)-(5.4.17) has traveling wave solutions for all $c \in [c_{min}, c_{max})$ but that question is still open: note the gap between the speeds in Theorems 5.5.1 and 5.5.2.

5.6 Basic properties of the solutions to the Lucas-Moll system

In this section, we discuss some of the basic properties of the solutions to the Lucas-Moll system.

Remark 5.6.1. From now on, we will always assume that $\alpha(s) = \alpha_1 s^k$ with $1/2 \le k < 1$.

We now discuss some recent results for the time-dependent problem, beyond the balanced growth path solutions. The optimal strategy $s^*(t,x)$ that appears in (5.4.16)-(5.4.17) is

$$s^*(t,x) = \arg\max_{s\in[0,1]} \Big[(1-s) + \alpha(s) I(t,x)\Big], \tag{5.6.1}$$

with

$$I(t,x) := e^{-x} \int_x^\infty [V(t,y) - V(t,x)](-F_y)(t,y)dy = e^{-x} \int_x^\infty V_y(t,y)F(t,y)dy. \tag{5.6.2}$$

As $I(t,x)$ is decreasing and $I(t,x) \to +\infty$ as $x \to -\infty$, while $I(t,x) \to 0$ as $x \to +\infty$, there exists a unique point $\eta_\ell(t) \in \mathbb{R}$ (the learning front location) so that

$$I(t, \eta_\ell(t)) = \frac{2}{\alpha_1}, \tag{5.6.3}$$

Let us also define the function $s_m(I)$ as follows:

$$s_m(I) = 1, \quad \text{for} \quad I > \frac{1}{\alpha'(1)}, \tag{5.6.4}$$

while $s_m(I)$ is determined implicitly by

$$\alpha'(s_m(I)) = \frac{1}{I}, \quad \text{for} \quad I \leq \frac{1}{\alpha'(1)}. \tag{5.6.5}$$

Then, the maximizer in (5.6.1) is given by

$$s^*(t,x) = s_m(I(t,x)). \tag{5.6.6}$$

An elementary computation shows that if $\alpha(s) = \alpha_1\sqrt{s}$ then

$$\begin{aligned} s^*(t,x) &= \frac{\alpha_1^2 I^2(t,x)}{4} \quad \text{for } x > \eta_\ell(t), \\ s^*(t,x) &= 1, \quad \text{for } x \leq \eta_\ell(t). \end{aligned} \tag{5.6.7}$$

It is convenient to introduce the function

$$w(t,x) = (\rho - \kappa)V_x(t,x)e^{-x}. \tag{5.6.8}$$

Note that $I(t,x)$ can be written in terms of $w(t,x)$ as

$$I(t,x) = e^{-x}\int e^y w(t,y)F(t,y)dy. \tag{5.6.9}$$

We will mostly work with the derivative formulation of the Lucas-Moll system:

$$\begin{aligned} &\frac{\partial F(t,x)}{\partial t} = \kappa\frac{\partial^2 F(t,x)}{\partial x^2} + F(t,x)\int_{-\infty}^{x} \alpha(s_m(I(t,y))(-F_y)(t,y)dy, \\ &\frac{\partial w(t,x)}{\partial t} + \kappa\frac{\partial^2 w(t,x)}{\partial x^2} + 2\kappa w_x(t,x) + (\rho-\kappa)\big[1 - s_m(I(t,x)) - w(t,x)\big] \\ &\qquad - \alpha(s_m(I(t,x)))w(t,x)F(t,x) = 0, \\ &I(t,x) = \frac{1}{\rho-\kappa}e^{-x}\int_x^\infty e^y w(t,y)F(t,y)dy. \end{aligned} \tag{5.6.10}$$

The function $F(t,x)$ satisfies the boundary conditions

$$F(t,-\infty) = 1, \quad F(t,+\infty) = 0. \tag{5.6.11}$$

while $w(t,x)$ satisfies the boundary conditions

$$w(t,-\infty) = 0, \quad w(t,+\infty) = 1. \tag{5.6.12}$$

Our main interest in this section is in the properties of the solutions that do not depend on the terminal condition $w_T(x)$ in any significant way.

5.6.1 Monotonicity of the solutions

Proposition 5.6.2. *We have*

$$\begin{aligned} &F_x(t,x) < 0, \quad \text{for all } 0 < t < T \text{ and all } x \in \mathbb{R}, \\ &w(t,x) > 0, \quad \text{for all } 0 < t < T \text{ and all } x \in \mathbb{R}, \\ &w_x(t,x) > 0, \quad \text{for all } 0 < t < T \text{ and all } x \in \mathbb{R}, \end{aligned} \tag{5.6.13}$$

Moreover, the optimal strategy $s^*(t,x) = s_m(I(t,x))$ is decreasing in x:

$$s^*_x(t,x) \leq 0, \quad \text{for all } 0 < t < T \text{ and all } x \in \mathbb{R}. \tag{5.6.14}$$

5.6.2 Bounds to the right of the learning front

We have the following decay estimates to the right of the learning front $\eta_\ell(t)$.

Lemma 5.6.3. *We have the following bounds:*

$$I(t,x) \leq \frac{1}{\alpha'(1)} e^{-(x-\eta_\ell(t))}, \quad \text{for all } x > \eta_\ell(t), \tag{5.6.15}$$

as well as

$$\alpha(s_m(I(t,x))) \leq \alpha(1) e^{-(x-\eta_\ell(t))}, \text{ for all } x > \eta_\ell(t), \tag{5.6.16}$$

and

$$s_m(I(t,x)) \leq e^{-2(x-\eta_\ell(t))}, \quad \text{for all } x > \eta_\ell(t). \tag{5.6.17}$$

Proof. Note that for $x > \eta_\ell(t)$ we have

$$\begin{aligned} I(t,x) &= e^{-x} \int_x^\infty e^y w(t,y) F(t,y) dy \leq e^{-x} \int_{\eta_\ell(t)}^\infty e^y w(t,y) F(t,y) dy = e^{-x} e^{\eta_\ell(t)} I(t, \eta_\ell(t)) \\ &= \frac{1}{\alpha'(1)} e^{-(x-\eta_\ell(t))}. \end{aligned} \tag{5.6.18}$$

Remark 5.6.4. These estimates essentially say that a reasonable way to think of this problem is that $s^*(t,x) = 1$ for $x < \eta_\ell(t)$ and $s^*(t,x) \approx 0$ for $x > \eta_\ell(t)$.

5.6.3 Tightness of the learning front

Here, we establish a lower bound for $w(t,x)$ to the right of the learning front $\eta_\ell(t)$. In other words, the learning front has a width that does not grow in time.

Proposition 5.6.5. *Suppose that $0 \leq \alpha_1 \leq A$, with some $A > 0$. For any $\gamma \in (0,1)$ there exist $T_\gamma > 0$, $L_{A,\gamma} > 0$ and $t_\gamma > 0$ such that for all $T \geq T_\gamma$ we have*

$$w(t,x) \geq \gamma, \quad \text{for all } x > \eta_\ell(t) + L_{A,\gamma} \text{ and } 0 \leq t \leq T - t_\gamma. \tag{5.6.19}$$

These constants do not depend on the terminal condition $w_T(x)$.

Remark 5.6.6. This proposition says that a sensible way to think of the function $w(t,x)$ is that $w(t,x) \approx 0$ for $x < \eta_\ell(t)$ and $w(t,x) = 1$ for $x > \eta_\ell(t)$.

5.6.4 The intrinsic learning front location

Let us now describe an approximation to the learning front location in terms of the agent distribution function $F(t,x)$ alone, motivated by the tightness of the learning front and the approximation mentioned in Remark 5.6.6. Consider the "intrinsic" pay-off functional

$$J(t,x) = J[F](t,x) = \frac{1}{\rho-\kappa} e^{-x} \int_x^\infty e^y F(t,y)dy. \tag{5.6.20}$$

This definition is "intrinsic" in the sense that it involves only the agents cumulative distribution function and not the value function. In other words, an agent can compute it purely from observing the positions of the other agents, or a large sample of other agents. As $J(t,x)$ is monotonically decreasing, there exists a unique *intrinsic learning front location* $e_\ell(t)$ such that

$$J(t, e_\ell(t)) = \frac{1}{\alpha'(1)}. \tag{5.6.21}$$

Once again, this location can be computed just from the positions of the agents, so an individual agent can find it without solving a "complicated" mean-field game problem.

The following proposition shows that $e_\ell(t)$ is a good approximation to the learning front $\eta_\ell(t)$. We will refer to $e_\ell(t)$ as the intrinsic learning front.

Proposition 5.6.7. *Suppose that $0 \le \alpha_1 \le A$, with some $A > 0$. There exist $L_{\ell,A} > 0$ and $t_A > 0$ that do not depend on T,so that*

$$e_\ell(t) - L_{\ell,A} \le \eta_\ell(t) \le e_\ell(t), \quad \text{for all } 0 \le t \le T - t_M. \tag{5.6.22}$$

Proof. Comparing to the expression for $I(t,x)$ in (5.6.10) and recalling that $0 \le w(t,x) \le 1$ for all $x \in \mathbb{R}$ and $0 \le t \le T$, we immediately see that

$$I(t,x) \le J(t,x), \quad \text{for all } 0 \le t \le T \text{ and } x \in \mathbb{R}. \tag{5.6.23}$$

As a consequence, we have the upper bound

$$\eta_\ell(t) \le e_\ell(t), \quad \text{for all } 0 \le t \le T. \tag{5.6.24}$$

On the other hand, for any $\gamma \in (0,1)$, we also have, by Proposition 5.6.5

$$\begin{aligned} I(t,x) &= \frac{1}{\rho-\kappa} e^{-x} \int_x^\infty e^y w(t,y) F(t,y)dy \ge \frac{\gamma}{\rho-\kappa} e^{-x} \int_x^\infty e^y F(t,y)dy \\ &= \gamma J(t,x), \quad \text{for all } 0 \le t \le T - t_\gamma \text{ and } x \ge \eta_\ell(t) + L_{A,\gamma}. \end{aligned} \tag{5.6.25}$$

Let us fix some $\gamma \in (0,1)$. Recalling the upper bound (5.6.15) on $I(t,x)$ to the right of the learning front, we obtain from (5.6.25) that

$$J(t,x) \le \frac{1}{\gamma\alpha'(1)} e^{-(x-\eta_\ell(t))}, \quad \text{for all } 0 \le t \le T - t_\gamma \text{ and } x \ge \eta_\ell(t) + L_{A,\gamma}. \tag{5.6.26}$$

Thus, if $e_\ell(t) \ge \eta_\ell(t) + L_{A,\gamma}$, then

$$J(t, e_\ell(t)) \le \frac{1}{\gamma\alpha'(1)} e^{-(e_\ell(t)-\eta_\ell(t))}. \tag{5.6.27}$$

Comparing to (5.6.21), we deduce that there exists $L_{\ell,A}$ so that

$$e_\ell(t) \le \eta_\ell(t) + L_{\ell,A}, \quad \text{for all } 0 \le t \le T - t_\gamma, \tag{5.6.28}$$

finishing the proof. □

One advantage of the intrinsic learning front location $e_\ell(t)$ is that it always moves to the right with a positive speed.

Proposition 5.6.8. *The function $J(t,x)$ defined by (5.6.20) is monotonically increasing in time and*

$$J(t,x) \geq e^{\kappa(t-s)} J(s,x), \quad \text{for all } 0 < s < t. \tag{5.6.29}$$

Moreover, the point $e_\ell(t)$ always moves to the right:

$$\dot{e_\ell}(t) \geq \kappa. \tag{5.6.30}$$

Proof. The function $F(t,x)$ is a super-solution to the heat equation

$$F_t \geq \kappa F_{xx}, \tag{5.6.31}$$

and is non-increasing. Let us multiply (5.6.31) by e^x and integrate from x to $+\infty$. This gives

$$\begin{aligned}(\rho-\kappa)e^x \frac{\partial J(t,x)}{\partial t} &\geq \kappa \int_x^\infty e^y F_{yy}(t,y)dy = -\kappa \int_x^\infty e^y F_y(t,y)dy - \kappa e^x F_x(t,x) \\ &= \kappa \int_x^\infty e^y F(t,y)dy + \kappa e^x F(t,x) - \kappa e^x F_x(t,x) \geq (\rho-\kappa)\kappa e^x J(t,x).\end{aligned} \tag{5.6.32}$$

This implies (5.6.29). It also follows that

$$\frac{\partial J(t,e_\ell(t))}{\partial t} \geq \kappa J(t,e_\ell(t)) + \frac{\kappa}{\rho-\kappa} F(t,e_\ell(t)) = \frac{\kappa}{\alpha'(1)} + \frac{\kappa}{\rho-\kappa} F(t,e_\ell(t)). \tag{5.6.33}$$

In addition, as

$$-J_x(t,x) = J(t,x) + \frac{1}{\rho-\kappa} F(t,x), \tag{5.6.34}$$

we know that

$$-J_x(t,e_\ell(t)) = J(t,e_\ell(t)) + \frac{\kappa}{\rho-\kappa} F(t,e_\ell(t)) = \frac{1}{\alpha'(1)} + \frac{\kappa}{\rho-\kappa} F(t,e_\ell(t)). \tag{5.6.35}$$

It follows from (5.6.21), (5.6.33) and (5.6.35) that

$$\dot{e_\ell}(t) = -\frac{J_t(t,e_\ell(t))}{J_x(t,e_\ell(t)} \geq \kappa, \tag{5.6.36}$$

finishing the proof □

A corollary of Propositions 5.6.7 and 5.6.8 is the following lower bound on the learning front location.

Corollary 5.6.9. *There exists a time T_0 and two constants $M_A > 0$ and $t_A > 0$ so that for all $T \geq T_0$ we have*

$$\eta_\ell(t) \geq \kappa t - M_A, \quad \text{for all } 0 \leq t \leq T - t_A. \tag{5.6.37}$$

5.6.5 The median front relative to the learning front

One may ask how the median front location $\eta_m(t)$ defined by

$$F(t,\eta_m(t)) = \frac{1}{2}. \tag{5.6.38}$$

compares to the intrinsic learning front location $e_\ell(t)$ and the learning front location $\eta_\ell(t)$. We have, for any $x > e_\ell(t)$:

$$\begin{aligned}\frac{1}{\alpha'(1)} = J(t, e_\ell(t)) &= \frac{e^{-e_\ell(t)}}{\rho - \kappa}\int_{e_\ell(t)}^{\infty} e^y F(t,y)dy \geq \frac{e^{-e_\ell(t)}}{\rho-\kappa}F(t,x)\int_{e_\ell(t)}^{x} e^y dy \\ &= F(t,x)\frac{1}{\rho-\kappa}\Big(e^{x-e_\ell(t)} - 1\Big).\end{aligned} \tag{5.6.39}$$

As $F(t, \eta_m(t)) = 1/2$, we deduce from (5.6.39) and Proposition 5.6.7 the following,

Proposition 5.6.10. *There exist T_0, $L_{A,\ell} > 0$, and $t_A > 0$ so that for all $T \geq T_0$ we have*

$$\eta_m(t) \leq \eta_\ell(t) + L_{A,\ell}, \qquad \text{for all } 0 \leq t \leq T - t_A. \tag{5.6.40}$$

Thus, the median front can not be far to the right of the learning front.

5.7 The state lottery society

The lottery society is the regime when the median front is located far behind the learning front:

$$\eta_m(t) \ll \eta_\ell(t), \tag{5.7.1}$$

and

$$\eta_\ell(t) - \eta_m(t) \to +\infty, \text{ as } t \to +\infty. \tag{5.7.2}$$

In particular, it follows that

$$s_m(t,x) = 1 \text{ unless } F(t,x) \ll 1. \tag{5.7.3}$$

In other words, in the lottery regime the overwhelming majority of the agents are only searching and are not producing anything at all. Moreover, the small fraction of those who do produce is diminishing in time because the learning and median fronts diverge, as seen in (5.7.2).

We will be more precise below but for now let us informally analyze when one can expect the lottery society to form. If (5.7.3) holds, then $F(t,x)$ solves, approximately, the Fisher-KPP equation

$$F_t \approx \kappa F_{xx} + \alpha_1 F(1-F). \tag{5.7.4}$$

Therefore, the median front has the asymptotics

$$\eta_m(t) = c_* t + o(t), \quad c_* = 2\sqrt{\kappa\alpha_1}. \tag{5.7.5}$$

Neglecting the algebraic and logarithmic terms in t and x, the solution to (5.7.4) has the exponential asymptotics

$$F(t, x + c_* t) \sim e^{-\lambda_* x} e^{-|x|^2/(4\kappa t)}, \ x \gg c_* t, \tag{5.7.6}$$

with

$$\lambda_* = \sqrt{\frac{\alpha_1}{\kappa}} = \frac{c_*}{2\kappa}. \tag{5.7.7}$$

For reasons that will become clear very soon, we expect the lottery society to exist as long as the exponential decay rate satisfies

$$\lambda_* < 1, \tag{5.7.8}$$

that is, as long as

$$c_* < 2\kappa, \tag{5.7.9}$$

or, equivalently,

$$\alpha_1 < \kappa. \tag{5.7.10}$$

Let us now informally compute the asymptotics of the learning front location when the exponential decay rate does satisfy (5.7.8). In that regime, the intrinsic pay-off functional $J(t,x)$ from (5.6.20) has the asymptotics, for $x \gg \eta_m(t) \approx c_* t$:

$$\begin{aligned} J(t,x) &= \frac{1}{\rho-\kappa} e^{-x} \int_x^\infty e^y F(t,y) dy \approx \frac{1}{\rho-\kappa} e^{-x} \int_x^\infty e^y e^{-\lambda_*(y-c_*t)} e^{-|y-c_*t|^2/(4\kappa t)} dy \\ &= \frac{1}{\rho-\kappa} e^{-x} \int_x^\infty e^y e^{-\lambda_*(y-c_*t)} e^{-|y-c_*t|^2/(4\kappa t)} dy \\ &= \frac{1}{\rho-\kappa} e^{-x+c_*t} \int_{x-c_*t}^\infty e^{(1-\lambda_*)y} e^{-|y|^2/(4\kappa t)} dy. \end{aligned} \tag{5.7.11}$$

Assuming that (5.7.8) holds, once again, dropping algebraic factors in t and constants, gives

$$\begin{aligned} J(t,x) &\sim e^{-x+c_*t} \int_{(x-c_*t)/\sqrt{4\kappa t}}^\infty e^{(1-\lambda_*)\sqrt{4\kappa t}y} e^{-|y|^2} dy \\ &= e^{-x+c_*t+(1-\lambda_*)^2\kappa t} \int_{(x-c_*t)/\sqrt{4\kappa t}}^\infty e^{-|y-(1-\lambda_*)\sqrt{\kappa t}|^2} dy \\ &= e^{-x+c_*t+(1-\lambda_*)^2\kappa t} \int_{\Gamma(t)}^\infty e^{-|y|^2} dy, \end{aligned} \tag{5.7.12}$$

with

$$\Gamma(t) = \frac{(x-c_*t)}{\sqrt{4\kappa t}} - (1-\lambda_*)\sqrt{\kappa t}. \tag{5.7.13}$$

Therefore, we have $J(t,x) \gg 1$ if we have both

$$-x + c_* t + (1-\lambda_*)^2 \kappa t \gg 1, \tag{5.7.14}$$

and $\Gamma(t) \leq 0$, which is

$$x - c_* t < 2(1-\lambda_*)\kappa t \tag{5.7.15}$$

Note that (5.7.15) is a consequence of (5.7.14) when $\lambda_* < 1$. Thus, one expects that if (5.7.8) holds then we have

$$\eta_\ell(t) \sim v_* t, \tag{5.7.16}$$

with

$$v_* = c_* + (1-\lambda_*)^2\kappa = 2\sqrt{\kappa\alpha_1} + (\sqrt{\kappa} - \sqrt{\alpha_1})^2 = \kappa + \alpha_1. \tag{5.7.17}$$

According to (5.7.10), we should expect that the transition out of the lottery society happens exactly when $\alpha_1 = \kappa$ and $v_* = 2\kappa$. This exactly matches the lower bound for the existence of balanced growth paths in [**?**].

We now restate our main result on existence of the lottery society.

Theorem 5.7.1. *Consider that $\alpha_1 < \kappa$. Suppose the initial condition $F_0(x)$ is monotonically decreasing, satisfies $F_0(x) = 1$ for all $x \leq -L_0$, and $F_0(x) = 0$ for all $x \geq L_0$. Suppose the terminal condition $w_T(x)$ is increasing and satisfies $w_T(-\infty) = 0$, $w_T(+\infty) = 1$. Then, we have the following asymptotics:*

$$\eta_m(t) = 2\sqrt{\kappa\alpha_1}t + o(t), \tag{5.7.18}$$

and

$$\eta_\ell(t) = (\kappa + \alpha_1)t + o(t). \tag{5.7.19}$$

As the Lucas-Moll system is only formulated on a finite time interval $0 \le t \le T$, the $o(t)$ terms in Theorem 5.7.1 should be understood as follows: we say that $g(t) = o(t)$ if for any $\varepsilon > 0$ there exist $t_0^\varepsilon > 0$ and $t_1^\varepsilon > 0$ so that for all $T > t_1^\varepsilon + t_0^\varepsilon$ we have

$$|g(t)| < \varepsilon, \quad \text{for all } 0 < t_0^\varepsilon < t < T - t_1^\varepsilon. \tag{5.7.20}$$

We deduce that

$$\eta_\ell(t) - \eta_m(t) \ge \beta t + o(t), \tag{5.7.21}$$

with

$$\beta = (\kappa + \alpha_1) - 2\sqrt{\kappa\alpha_1} > 0. \tag{5.7.22}$$

Thus, paradoxically, when learning has a very small chance of success, everyone is only doing the learning, based on an extremely small chance of becoming extremely knowledgeable. This is the state lottery society. In particular, the fraction of the agents who produce anything goes to zero as $t \to +\infty$, Note that, according to Theorem 5.5.1 this is exactly the regime where the balanced growth paths do not exist.

5.8 The effective learning regime

The approximation of the learning front $\eta_\ell(t)$ by the "intrinsic" location $e_\ell(t)$ motivates the following "intrinsic" toy model, as an approximation to (5.1.1)-(5.1.2):

$$\begin{aligned} u_t &= \kappa u_{xx} + u(t,x) \int_{-\infty}^{x} \beta(J(t,y))(-u_y)dy, \\ J(t,x) &= e^{-x} \int_x^\infty e^y u(t,y)dy, \end{aligned} \tag{5.8.1}$$

with

$$\begin{aligned} \beta(J) &= \frac{\beta_0^2}{2} J, \quad \text{for } J \le \frac{2}{\beta_0}, \\ \beta(J) &= \beta_0, \quad \text{for } J > 2/\beta_0. \end{aligned} \tag{5.8.2}$$

The advantage is that this model is posed purely forward in time and does not involve the value function.

Going back to the interacting agents learning model for a finite $N \gg 1$, one may guess that the toy model (5.8.1)-(5.8.2) is the mean-field limit for the case when the exponential clock rates of the individual agents are given by a version of (5.3.4) (but now written in the logarithmic variables $x_k(t)$)

$$\alpha_k(t) = \beta_0 \min\Big[1, \frac{1}{N} \sum_{m: x_m(t) \ge x_k(t)} \Big(e^{x_m(t) - x_k(t)} - 1\Big)\Big]. \tag{5.8.3}$$

A careful study of this limit is open at the moment.

We have the following result in the "very effective learning" regime for the intrinsic toy model.

Theorem 5.8.1. *For any $\delta > 0$ there exists B_δ so that for all $\beta > B_\delta$ we have*

$$2\sqrt{\kappa} - \delta \le \liminf_{t \to +\infty} \frac{\eta_m(t)}{\sqrt{\beta_0 t}} \le \limsup_{t \to +\infty} \frac{\eta_m(t)}{\sqrt{\beta_0 t}} \le 2\sqrt{\kappa}. \tag{5.8.4}$$

This says that when the learning is very effective, in the sense that $\beta_0 \gg 1$, then the solution spreads at a speed that approaches the Fisher-KPP speed $\bar{c}_* = 2\sqrt{\kappa\beta_0}$. The key step in the proof is the following estimate at the learning front location.

Proposition 5.8.2. *For any $m \in (0, 1/2)$ there exist $c_m > 0$, $T_0 > 0$ and $B_0 > 0$ such that*

$$u(t, e_\ell(t; u)) \leq \frac{c_m}{\beta_0^m}, \quad \text{for all } \beta_0 > B_0, \text{ and } t \geq \frac{T_0}{\beta_0}. \tag{5.8.5}$$

In other words, when learning is very effective, the fraction of the agents who produce nothing but only learn is very close to 1. This is similar to the state lottery society. However, an important difference with Theorem 5.8.1 for the "very effective learning" regime in the toy model is that in the state lottery society the fraction of the agents who produce anything goes to zero as $t \to +\infty$, while in the regime of Theorem 5.8.1 is stays small but fixed as $t \to +\infty$, with $\beta_0 \gg 1$ fixed.

The following seems to be true but still being checked.

Theorem 5.8.3. *The claim of Theorem 5.8.1 is also true for the full Lucas-Moll system.*

References

[1] E. Aïdékon, J. Berestycki, E. Brunet and Z. Shi, Branching Brownian motion seen from its tip, Probab. Theory Related Fields, **157**, 2013, 405–451.

[2] E. Aidekon and Z. Shi, The Seneta-Heyde scaling for the branching random walk, Ann. Probab., **42**, 2014, 959–993.

[3] J. An, C. Henderson and L. Ryzhik, Front location determines convergence rate to traveling waves, Preprint, 2023.

[4] J. An, C. Henderson and L. Ryzhik, Quantitative steepness, semi-FKPP reactions, and pushmi-pullyu fronts, to appear in Arch. Rat. Mech. Anal., 2023.

[5] J. An, C. Henderson and L. Ryzhik, Voting models and semilinear parabolic equations, to appear in Nonlinearity, 2023.

[6] J. An, C. Henderson and L. Ryzhik, Pushed, pulled and pushmi-pullyu fronts of the Burgers-FKPP equation, to appear in Jour. Eur. Math. Soc., 2023

[7] E. Aidékon, J. Berestycki, É. Brunet, Z. Shi, Branching Brownian motion seen from its tip, Probab. Theory Relat. Fields **157**, 2013, 405–451.

[8] S. Angenent, The zero set of a solution of a parabolic equation. J. Reine Angew. Math. **390**, 79–96, 1988.

[9] L.-P. Arguin, A. Bovier, and N. Kistler, Poissonian statistics in the extremal process of branching Brownian motion. Ann. Appl. Probab. **22**, 1693–1711, 2012.

[10] L.-P. Arguin, A. Bovier, and N. Kistler, The extremal process of branching Brownian motion. Probab. Theory Relat. Fields **157**, 535–574, 2013.

[11] L.-P. Arguin, A. Bovier, and N. Kistler, An ergodic theorem for the frontier of branching Brownian motion, Electron. J. Probab., **18**, 2013, no. 53, 25.

[12] L.-P. Arguin, D. Belius, and P. Bourgade, Maximum of the characteristic polynomial of random unitary matrices. Comm. Math. Phys., **349**, 703–751, 2017.

[13] L.-P. Arguin, D. Belius, P. Bourgade, M. Radziwill, and K. Soundararajan, Maximum of the Riemann zeta function on a short interval of the critical line, Comm. Pure Appl. Math. **72**, 500–535, 2019.

[14] L.-P. Arguin, P. Bourgade and M. Radziwill, Tc. I. Preprint, arXiv:2007.00988, 2020.

[15] L.-P. Arguin, G. Dubach and L. Hartung, Maxima of a random model of the Riemann zeta function over intervals of varying length, Preprint arXiv: 2103.04817..

[16] L.-P. Arguin, L. Hartung, N. Kistler, High points of a random model of the Riemann-zeta function and Gaussian multiplicative chaos, Stoch. Proc. Appl. **151**, 2022, 174–190.

[17] L.-P. Arguin, F. Ouimet and M. Radziwill, Moments of the Riemann zeta function on short intervals of the critical line, Ann. Probab. **49**, 3106–3141, 2021.

[18] D. Aronson and H. Weinberger, Multidimensional nonlinear diffusion arising in population genetics, Adv. Math. **30**, 1978, 33–76.

[19] M. Avery and L. Garénaux, Spectral stability of the critical front in the extended Fisher-KPP equation, Preprint arXiv:2009.01506v1, 2020.

[20] M. Avery, M. Holzer, and A. Scheel, Pushed-to-pulled front transitions: continuation, speed scalings, and hidden monotonicity, Preprint, arXiv:2206.09989, 2022.

[21] M. Avery and A. Scheel, Asymptotic stability of critical pulled fronts via resolvent expansions near the essential spectrum, SIAM J. Math. Anal. **53**, 2206–2242, 2021.

[22] M. Avery and A. Scheel, Universal selection of pulled fronts, Preprint arXiv:2012.06443, 2020.

[23] M. Bachmann, Limit theorems for the minimal position in a branching random walk with independent log-concave displacements, Adv. Appl. Probabi., **32**, 2000, 159–176.

[24] J. Berestycki, Topics on Branching Brownian motion, Lecture notes, 2014, https://www.stats.ox.ac.uk/~berestyc/Articles/EBP18_v2.pdf

[25] J. Berestycki, N. Berestycki and J. Schweinsberg, The genealogy of branching Brownian motion with absorption, Ann. Probab. **41**, 2013, 527–618.

[26] J. Berestycki, É. Brunet, and B. Derrida. Exact solution and precise asymptotics of a Fisher–KPP type front, Jour. Phys. A: Math. Theor. **51**, 035204, 2018.

[27] J. Berestycki, É. Brunet, and B. Derrida. A new approach to computing the asymptotics of the position of Fisher-KPP fronts, EPL (Europhysics Letters) **122**, 10001, 2018.

[28] J. Berestycki, Y. Kim, E. Lubetzky, B. Mallein, and O. Zeitouni, The extremal point process of branching Brownian motion in $\mathbb{R}^d$, Ann. Probab., **52**, 2024, 955–982.

[29] E. Bouin and C. Henderson. The Bramson delay in a Fisher-KPP equation with log-singular non-linearity, Nonlinear Anal. **213**, 112508, 2020.

[30] J. Bourgain, Decoupling, exponential sums and the Riemann zeta function, Jour. AMS, **30**, 2017, 205–224.

[31] A. Bovier, From spin glasses to branching Brownian motion – and back?, in "Random Walks, Random Fields, and Disordered Systems" (Proceedings of the 2013 Prague Summer School on Mathematical Statistical Physics), M. Biskup, J. Cerny, R. Kotecky, Eds., Lecture Notes in Mathematics 2144, Springer, 2015.

[32] M. D. Bramson, Maximal displacement of branching Brownian motion, Comm. Pure Appl. Math. **31**, 531–581, 1978.

[33] M. D. Bramson, *Convergence of solutions of the Kolmogorov equation to travelling waves*, Mem. Amer. Math. Soc. **44**, 1983.

[34] E. Brunet and B. Derrida. Statistics at the tip of a branching random walk and the delay of traveling waves. Eur. Phys. Lett. **87**, 60010, 2009.

[35] E. Brunet and B. Derrida. A branching random walk seen from the tip, Jour. Stat. Phys. **143**, 420–446, 2011.

[36] E. Carlen and M. Loss, Sharp constant in Nash's inequality, Int. Math. Res. Not., 1993, no. 7, 213–215.

[37] E. Carlen and M. Loss, Optimal smoothing and decay estimates for viscously damped conservation laws, with applications to the 2-D Navier-Stokes equation, Duke Math. Jour. **31**, 135–157, 1996.

[38] R. Chhaibi, T. Madaule, and J. Najnudel. On the maximum of the CβE field, Duke Math. J., **167**, 2243–2345, 2018.

[39] P. Constantin, Generalized relative entropies and stochastic representation, Int. Math. Res. Not. 2006, Art. ID 39487, 9 pp.

[40] A. Cortines, L. Hartung and O. Louidor, The Structure of Extreme Level Sets in Branching Brownian Motion, Ann. Probab. **47**, 2019, 2257–2302.

[41] A. De Masi, P. A. Ferrari, and J. L. Lebowitz. Reaction-diffusion equations for interacting particle systems. J. Stat. Phys., **44**, 589–644, 1986.

[42] B. Derrida, Cross-overs of Bramson's shift at the transition between pulled and pushed fronts, Jour. Stat. Phys. **190**, 2023, Paper No. 67, 14 pp.

[43] U. Ebert and W. van Saarloos, Front propagation into unstable states: universal algebraic convergence towards uniformly pulled fronts, Physica D **146**, 1-99, 2000.

[44] A. Etheridge, N. Freeman and S. Penington, Branching Brownian motion, mean curvature flow and the motion of hybrid zones, Electr. Jour. Probab. **22**, 2017, 1–40.

[45] A. Etheridge and S. Penington, Genealogies in bistable waves, arXiv:2009.03841, 2020.

[46] P.C. Fife and J.B. McLeod, The approach of solutions of nonlinear diffusion equations to travelling front solutions, Arch. Rat. Mech. Anal. **65**, 335–361, 1977.

[47] R. A. Fisher, The wave of advance of advantageous genes, Ann. Eugen. **7**, 355–369, 1937.

[48] G. Flath, A simpler path to ergodic theorems for the frontier of branching Brownian motion, Preprint, 2025,

[49] Y. V. Fyodorov, G. A. Hiary, and J. P. Keating, Freezing transition, characteristic polynomials of random matrices, and the Riemann zeta function, Phys. Rev. Lett., **108**:170601, 2012.

[50] Y. V. Fyodorov and J. P. Keating, Freezing transitions and extreme values: random matrix theory, and disordered landscapes, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., **372** (2007), 20120503, 32pp., 2014.

[51] X. Fu, Q. Griette, and P. Magal, A cell-cell repulsion model on a hyperbolic Keller-Segel equation. J. Math. Biol. **80**, 2257–2300, 2020.

[52] X. Fu, Q. Griette, and P. Magal, Sharp discontinuous traveling waves in a hyperbolic Keller-Segel equation. Math. Models Methods Appl. Sci. **31**, 861–905, 2021.

[53] H. Fujita, On the blowing up of solutions of the Cauchy problem for $u_t = \Delta u + u^{1+\varepsilon}$, J. Fac. Sci. Univ. Tokyo, **13**, 1966, 109–124.

[54] T. Gallay and E. Risler, A variational proof of global stability for bistable travelling waves, Differential Integral Equations **20**, 2007, 901–926.

[55] J. Garnier, T. Giletti, F. Hamel, and L. Roques. Inside dynamics of pulled and pushed fronts. Jour. Math. Pures Appl., **98**, 428–449, 2012.

[56] T. Giletti, Monostable pulled fronts and logarithmic drifts, Preprint arXiv:2105.12611, 2021.

[57] T. Giletti and H. Matano, Existence and uniqueness of propagating terraces, Comm. Contemp. Math. **22**, 1950055, 38 pp., 2020.

[58] C. Graham, Precise asymptotics for Fisher–KPP fronts, Nonlinearity **32**, 1967–1998, 2019.

[59] C. Graham, L. Mytnik and L. Ryzhik, Persistent anisotropy in multi-dimensional branching Brownian motion, Preprint, 2026.

[60] K. Hadeler and F. Rothe, Travelling fronts in nonlinear diffusion equations, J. Math. Biol. **2**, 251–263, 1975.

[61] F. Hamel and R. Monneau, Solutions of semilinear elliptic equations in $\mathbb{R}^N$, Comm. PDEs, **25**, 2000, 769–819.

[62] F. Hamel, R. Monneau and J.-M. Roquejoffre, Existence and qualitative properties of multidimensional conical bistable fronts, DCDS, **13**, 2005, 1069–1096.

[63] F. Hamel and N. Nadirashvili, Travelling fronts and entire solutions of the Fisher-KPP equation in $\mathbb{R}^N$, Arch. Rat. Mech., **157**, 2001, 91–163.

[64] F. Hamel, J. Nolen, J.-M. Roquejoffre and L. Ryzhik, A short proof of the logarithmic Bramson correction in Fisher-KPP equations, Netw. Heterog. Media **8**, 275–289, 2013.

[65] C. Henderson, Population stabilization in branching Brownian motion with absorption and drift, Commun. Math. Sci. **14**, 973–985, 2016.

[66] C. Henderson, Slow and fast minimal speed traveling waves of the FKPP equation with chemotaxis, arXiv preprint arXiv:2102.06065, 2021.

[67] P. Howard, Pointwise estimates and stability for degenerate viscous shock waves J. reine angew. Math. **545**, 19–65, 2002.

[68] Y. Hu, The almost sure limits of the minimal position and the additive martingale in a branching random walk, J. Theoret. Probab., **28**, 2015, 467–487.

[69] Y. Hu and Z. Shi, Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees, Ann. Probab., **37**, 742–789.

[70] Y. Kim, E. Lubetzky and O. Zeitouni, The maximum of branching Brownian motion in $\mathbb{R}^d$, Ann. Appl. Probab., **33**, 2023, 1315–1368.

[71] Y. Kim and O. Zeitouni, The shape of the front of multidimensional branching Brownian motion, Probab. Theory Related Fields, **193**, 2025, 1121-1160.

[72] A.N. Kolmogorov, I.G. Petrovskii, and N.S. Piskunov, Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. Bull. Univ. Moskow, Ser. Internat., Sec. A **1**, 1–25, 1937.

[73] S. M. Krone and C. Neuhauser, Ancestral processes with selection. Theor. Pop. Biol., **51**, 210–237, 1997.

[74] K.-S. Lau, On the nonlinear diffusion equation of Kolmogorov, Petrovskii and Piskunov, J. Diff. Eqs. **59**, 44–70, 1985.

[75] H. Lofting, *The Story of Doctor Dolittle*, New York: Frederick A. Stoke Co., 1920.

[76] J. Leach and E. Hanaç, On the evolution of travelling wave solutions of the Burgers-Fisher equation, Quart. Appl. Math. **74**, 337–359, 2016.

[77] J.A. Leach and D.J. Needham, *Matched Asymptotic Expansions in Reaction-Diffusion Theory*, Springer-Verlag London, 2004.

[78] M.T. Keating and J.T. Bonner, Negative chemotaxis in cellular slime molds, J. Bacteriology **130**, 144–147, 1977.

[79] X. Kriechbaum, L. Ryzhik and O. Zeitouni, Voting models and tightness for a family of recursion equations, Jour. d'Anal. Math., **157**, 2025, 559–616.

[80] J. La, J.-M. Roquejoffre and L. Ryzhik, A uniform bound for solutions to a thermo-diffusive system, Arch. Rat. Mech. Anal., **248**, Paper 96, 2024.

[81] S. P. Lalley and T. Sellke A conditional limit theorem for the frontier of a branching Brownian motion Ann. Probab., **15**, 1987, 1052–1061.

[82] G. Lambert and E. Paquette, The law of large numbers for the maximum of almost Gaussian log-correlated fields coming from random matrices, Probab. Theory Related Fields **173**, 157–209, 2019.

[83] M. Lucia, C. Muratov and M. Novaga, Linear vs. nonlinear selection for the propagation speed of the solutions of scalar reaction-diffusion equations invading an unstable equilibrium, Comm. Pure. Appl. Math. **57**, 616–636, 2004.

[84] R. Lyons, A simple path to Biggins' martingale convergence for branching random walk, in *Classical and modern branching processes (Minneapolis, MN, 1994)*, IMA Vol. Math. Appl., **84**, 1997, 217–221.

[85] T. Madaule, The tail distribution of the derivative martingale and the global minimum of the branching random walk, arXiv:1606.03211, 2016.

[86] P. Maillard, The number of absorbed individuals in branching Brownian motion with a barrier, Ann. Inst. Henri Poincaré Probab. Stat., **49**, 2013, 428–455.

[87] P. Maillard and M. Pain, 1-stable fluctuations in branching Brownian motion at critical temperature I: the derivative martingale, Ann. Prob., **47**, 2019, 2953–3002.

[88] B. Mallein, Maximal displacement of d-dimensional branching Brownian motion, Electron. Commun. Probab., **20**, 2015, no. 76, 12.

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

[90] P. Michel, S. Mischler,and B. Perthame, General entropy equations for structured population models and scattering, C. R. Math. Acad. Sci. Paris **338**, 697–702, 2004.

[91] C. Muratov and M.Novaga, Global exponential convergence to variational traveling waves in cylinders, SIAM J. Math. Anal. **44**, 2012, 293–315.

[92] C. Muratov and M.Novaga, Front propagation in infinite cylinders. I. A variational approach, Commun. Math. Sci. **6**, 2008, 799–826.

[93] J. D Murray. *Mathematical biology: I. An introduction*, Third ed., Vol. 17, Interdisciplinary Applied Mathematics Series, Springer New York, 2002.

[94] C. Mueller, L. Mytnik and L. Ryzhik, The speed of a random front for stochastic reaction-diffusion equations with strong noise, Comm. Math. Phys. **384**, 2021, 699–732.

[95] L. Mytnik, J.-M. Roquejoffre and L. Ryzhik, Fisher-KPP equation with small data and the extremal process of branching Brownian motion, Adv. Math., **386**, 2022, Paper No. 108106.

[96] J. Najnudel, On the extreme values of the Riemann zeta function on random intervals of the critical line, Probab. Theory Related Fields, **172**, 387–452, 2018.

[97] C. Neuhauser and S. M. Krone, Genealogies of samples in models with selection. Genetics, **145**, 519–534, 1997.

[98] J. Nolen, J.-M. Roquejoffre and L. Ryzhik, Convergence to a single wave in the Fisher–KPP equation, Chin. Ann. Math. Ser. B **38**, 629–646, 2017.

[99] J. Nolen, J.-M. Roquejoffre and L. Ryzhik, Refined long-time asymptotics for Fisher–KPP fronts, Comm. Contemp. Math., 2018, 1850072.

[100] Z. O'Dowd, Branching Brownian motion and partial differential equations, Thesis, Oxford University, 2019.

[101] E. Paquette and O. Zeitouni, The maximum of the CUE field, Int. Math. Res. Not. IMRN, pp. 5028–5119, 2018.

[102] E. Risler, Global convergence toward traveling fronts in nonlinear parabolic systems with a gradient structure, Ann. Inst. H. Poincaré Anal. Non Linéarie **25**, 2008, 381–424.

[103] M. Roberts, A simple path to asymptotics for the frontier of a branching Brownian motion, Ann. Prob. **41**, 3518–3541, 2013.

[104] J.-M. Roquejoffre, Eventual monotonicity and convergence to travelling fronts for the solutions of parabolic equations in cylinders, Ann. Inst. Henri Poincaré, **14**, 499-552, 1997.

[105] F. Rothe, Convergence to pushed fronts, Rocky Mount. Jour. Math. **11**, 617–633, 1981.

[106] D.H Sattinger, On the stability of waves of nonlinear parabolic systems. Adv. Math., **22**, 312–355, 1976.

[107] D.H. Sattinger, Weighted norms for the stability of traveling waves, Jour. Diff. Eqs. **25**, 130–144, 1977.

[108] R. Stasiński, J. Berestycki, and B. Mallein, Derivative martingale of the branching Brownian motion in dimension $d \geq 1$, Ann. Inst. Henri Poincaré Probab. Stat., **57**, 2021, 1786–1810.

[109] A. Sznitman, Topics in propagation of chaos, École d'Été de Probabilités de Saint-Flour XIX–1989, 165–251. Lecture Notes in Math., 1464. Springer-Verlag, Berlin, 1991

[110] G. Talenti A weighted version of a rearrangement inequality. Ann. Univ. Ferrara Sez. VII (N.S.), **43**, 121–133, 1997.

[111] W. van Saarlos, Front propagation into unstable states, Physics Reports, **386**, 29-222, 2003.

[112] K. Uchiyama, The behavior of solutions of some nonlinear diffusion equations for large time, J. Math. Kyoto Univ. **18**, 453–508, 1978.

[113] N. Vladimirova, G. Weirs, and L. Ryzhik, Flame capturing with an advection–reaction–diffusion model. Comb. Theory Model., **10**, 727–74, 2006.

[114] M. Zaki, N. Andrew, and R.H. Insall, Entamoeba histolytica cell movement: a central role for selfgenerated chemokines and chemorepellents, Proc. Nat. Acad. Sci. **103**, 18751–18756, 2006.