UNIVERSITY LECTURE SERIES VOLUME 68

50 Years of First-Passage Percolation

Antonio Auffinger
Michael Damron
Jack Hanson

American Mathematical Society

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068

UNIVERSITY LECTURE SERIES VOLUME 68

50 Years of First-Passage Percolation

Antonio Auffinger
Michael Damron
Jack Hanson

American Mathematical Society
Providence, Rhode Island

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

EDITORIAL COMMITTEE

Jordan S. Ellenberg　　Robert Guralnick
William P. Minicozzi II (Chair)　　Tatiana Toro

2010 *Mathematics Subject Classification*. Primary 60K35, 82B43.

For additional information and updates on this book, visit
www.ams.org/bookpages/ulect-68

Library of Congress Cataloging-in-Publication Data

Names: Auffinger, Antonio, 1983– author. | Damron, Michael, 1981– author. | Hanson, Jack, 1985– author.
Title: 50 years of first-passage percolation / Antonio Auffinger, Michael Damron, Jack Hanson.
Other titles: Fifty years of first-passage percolation
Description: Providence, Rhode Island : American Mathematical Society, [2017] | Series: University lecture series ; volume 68 | Includes bibliographical references.
Identifiers: LCCN 2017029669 | ISBN 9781470441838 (alk. paper)
Subjects: LCSH: Random walks (Mathematics) | Limit theorems (Probability theory) | Probabilities. | AMS: Probability theory and stochastic processes – Special processes – Interacting random processes; statistical mechanics type models; percolation theory. msc | Statistical mechanics, structure of matter – Equilibrium statistical mechanics – Percolation. msc
Classification: LCC QA274.73 .A94 2017 | DDC 519.2/3–dc23
LC record available at https://lccn.loc.gov/2017029669

Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy select pages for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Permissions to reuse portions of AMS publication content are handled by Copyright Clearance Center's RightsLink® service. For more information, please visit: `http://www.ams.org/rightslink`.

Send requests for translation rights and licensed reprints to `reprint-permission@ams.org`.

Excluded from these provisions is material for which the author holds copyright. In such cases, requests for permission to reuse or reprint material should be addressed directly to the author(s). Copyright ownership is indicated on the copyright page, or on the lower right-hand corner of the first page of each article within proceedings volumes.

© 2017 by the authors. All rights reserved.
Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1　　22 21 20 19 18 17

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Contents

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Preface

We celebrate the fiftieth anniversary of one the most classical models in probability theory. In this book, we describe the main results of first-passage percolation, paying special attention to the recent burst of advances. The purpose of this book is twofold. We first give self-contained proofs of seminal results obtained in the '80s and '90s on limit shapes and geodesics, while covering the state of the art of these questions. Second, aside from these classical results, we discuss recent perspectives and directions including (1) the connection between Busemann functions and geodesics, (2) proofs of sublinear variance of the passage time, and (3) the role of growth and competition models. We also provide a collection of old and new open questions, in the hope that they will be solved before the one hundredth anniversary.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/01

CHAPTER 1

Introduction

1.1. The model of first-passage percolation and its history

First-passage percolation (FPP) was introduced by Hammersley and Welsh [**108**] in 1965 as a model of fluid flow through a random medium. It has been a stage of research for probabilists since its origin, but despite all efforts through the past decades, most of the predictions about its important statistics remain to be understood. Most of the beauty of the model lies in its simple definition (as a random metric space) and that several of its fascinating conjectures do not require much effort to be stated. During these 50 years, FPP has attracted the attention of theoretical physicists, biologists, and computer scientists, and also gave birth to some now fundamental mathematical tools, like the sub-additive ergodic theorem. Here, we will focus on the model defined on the lattice $\mathbb{Z}^d$ with independent and identically distributed (i.i.d.) edge-weights; some variants will be discussed in Chapter 7.

The model is defined as follows. We place a non-negative random variable τ_e, called the passage time of the edge e, at each nearest-neighbor edge in $\mathbb{Z}^d$. The collection (τ_e) is assumed to be i.i.d. with common distribution function F and corresponding probability measure ν. The random variable τ_e is interpreted as the time or the cost needed to traverse the edge e.

A path Γ is a finite or infinite sequence of edges $e(1), e(2), \ldots$, in $\mathbb{Z}^d$ such that for each $n \geq 1$, $e(n)$ and $e(n+1)$ share at least one endpoint. For any finite path Γ we define the passage time of Γ to be

$$T(\Gamma) = \sum_{e \in \Gamma} \tau_e.$$

Given two points $x, y \in \mathbb{R}^d$ one then sets

$$T(x,y) = \inf_{\Gamma} T(\Gamma), \tag{1.1}$$

where the infimum is over all finite paths Γ that contain both x' and y', and x' is the unique vertex in $\mathbb{Z}^d$ such that $x \in x' + [0,1)^d$ (similarly for y'). The random variable $T(x,y)$ will be called the passage time between points x and y. In the original interpretation of the model, $T(x,y)$ represents the time that a fluid with source at x takes to reach the location y.

For each $t \geq 0$ let

$$B(t) = \{y \in \mathbb{R}^d \;:\; T(0,y) \leq t\}.$$

In the case that $F(0) = 0$, the pair $(\mathbb{Z}^d, T(\cdot,\cdot))$ is almost surely (a.s.) a metric space and $B(t) \cap \mathbb{Z}^d$ is the (random) ball of radius t centered at the origin. (See

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Figure 1.) The ultimate goal of first-passage percolation is to understand the large-scale properties of this metric. Some of the main questions include the following. We write $|\cdot|$ for the ℓ^2 norm in $\mathbb{R}^d$.

(1) What is the typical distance between two points that are far from each other in the lattice? Or in other words, what can we say about $T(x, y)$ as $|x - y| \to \infty$? Does it converge, when possibly rescaled and recentered? If so, what is the rate of convergence?
(2) What does a ball of large radius look like? Is there a scaling limit and fluctuation theory for the set $B(t)$?
(3) What is the geometry of geodesics (time-minimizing paths) between two distant points? How different are they from straight lines?
(4) What role does the distribution of the passage times play in describing the metric?

In this book, we will discuss progress on these and related questions. The purpose is twofold. First, we hope that this book will serve as a quick guide for readers who are not necessarily experts in the field. We will try to provide not only the main results, but also the main techniques and a large collection of open problems. Second, the field has had a burst of activity in the past five years and the most complete survey is more than a decade old. We hope that this book will fill this gap, or at least share some of the beautiful mathematical ideas and constructions that arise through FPP and which have enchanted many throughout these years.

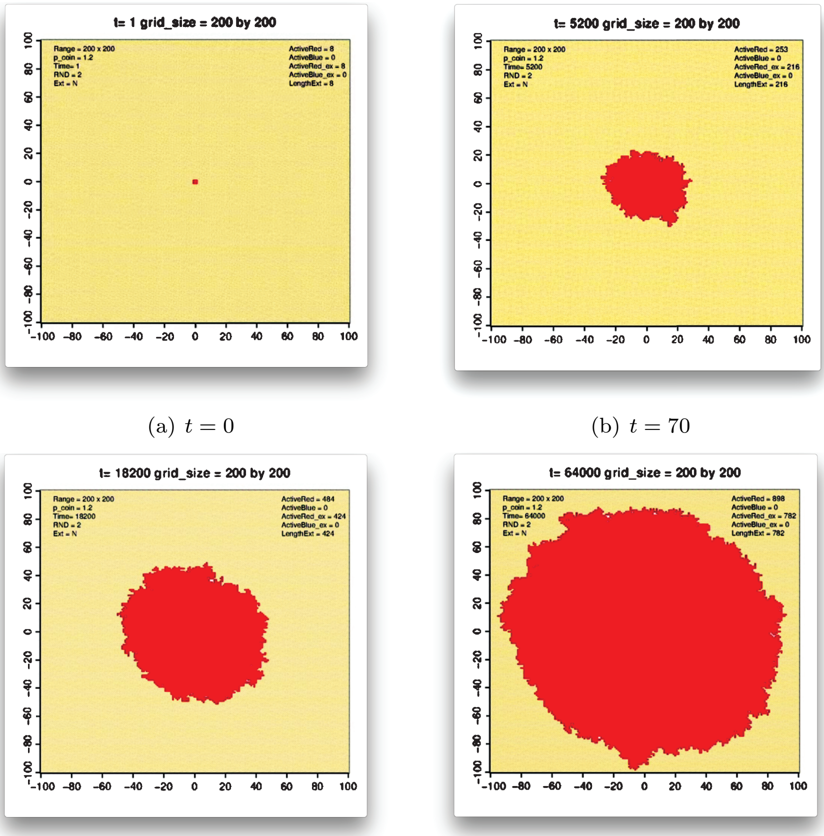
Let us return to questions 1 to 4. The original paper of Hammersley and Welsh [**108**] considered question 1 for a class of passage times in $\mathbb{Z}^2$. If we write e_1 for the first coordinate vector, they showed that $T(0, ne_1)$ grows linearly in n. Their result was extended in the famous work of Kingman [**44**, **128**, **129**]. It was also the building block for the classical "shape theorem" of Richardson [**155**], improved by Cox and Durrett [**62**] and Kesten [**125**], that gives the analogue of the law of large numbers for the random ball $B(t)$. The shape theorem roughly says that $B(t)$ grows linearly in t and, when properly normalized, it converges to a deterministic subset $\mathcal{B}$ of $\mathbb{R}^d$, called the limit shape. The set $\mathcal{B} = \mathcal{B}_\nu$ is not universal and depends on the distribution ν of the passage times. Chapter 2 is devoted to explaining the shape theorem and certain properties of the limit shape $\mathcal{B}$.

In Chapter 3, we discuss the variance and the order of fluctuations of the passage time T. In two dimensions, it is expected that under certain assumptions on ν the fluctuations are governed by the predictions of physicists, including Kardar, Parisi, and Zhang [**76**, **120**, **121**, **132**]. In higher dimensions, the picture is less clear and some of the predictions disagree. After stating what is conjectured, we focus our attention on presenting proofs of more recent results, including sublinear variance of $T(0, x)$ in x valid under minimal assumptions on the passage time.

Chapter 4 is devoted to the study of geodesics. We discuss the existence and properties of finite geodesics between any two points, then move to the study of geodesic rays. We present results on coalescence, directional properties of geodesic rays, and a proof sketch of the absence of geodesic lines (or bigeodesics) in the upper half-plane. We also present the important connection between geodesic lines and ground states of the two-dimensional Ising ferromagnet.

In Chapter 5 we describe the role of Busemann functions in the model. We explain a beautiful argument by Hoffman for the existence of two or more geodesic

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms



(a) $t = 0$

(b) $t = 70$

(c) $t = 135$

(d) $t = 250$

FIGURE 1. Simulation of the ball $B(t)$ at $t = 0, t = 70, t = 135$ and $t = 250$. The passage times τ_e have exponential distribution with mean one. Simulations by Si Tang.

rays. We then focus on Busemann-type limits and their relation to limiting geodesic graphs. Chapter 6 introduces the vast relation between FPP, growth processes, and infection models. We focus on questions of coexistence of multiple species and the limiting interface.

Chapter 7 is our attempt to show the reader what this book is not about. In the literature, there are thousands of pages of related (and equally fascinating) questions and models, similar to or inspired by FPP. We collect a few of these examples and point out the appropriate references. In particular, we briefly discuss FPP with non-independent weights, FPP on different graphs, the maximum flow problem, and exactly solvable models for last-passage percolation. Chapter 8 recalls open questions spread throughout this manuscript, put in one place for easy reference.

This book is intended to serve as an introduction to the field for researchers, as a reference, and also as a textbook for a graduate course. No previous knowledge

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

of percolation theory is assumed. A first-year graduate course in probability theory should suffice.

1.2. Acknowledgments

We thank the American Institute of Mathematics and its staff for helping us organize a workshop on this subject. A.A. thanks the hospitality of the mathematics department at Indiana University, and its visiting professor program, during which part of this book was completed. He also thanks Elizabeth Housworth, who introduced him to the Japanese eraser and chalk dust cleaner. M.D. thanks the mathematics department at Northwestern University for hospitality. A. A. and J. H. thank the School of Mathematics at Georgia Tech for hospitality. We thank Phil Sosoe for several fruitful discussions on this topic, especially on concentration estimates. We thank Si Tang for making the simulations used in Figure 1 available to us, Sven Erick Alm and Maria Deijfen for simulations and Figures 2, 4, and 5, and Xuan Wang for discussions on nonrandom fluctuations. We also thank Daniel Ahlberg, Rick Durrett, Chris Hoffman, Wai-Kit Lam, Timo Seppäläinen, Vladas Sidoravicius, Phil Sosoe, Si Tang, Chen Xu, three anonymous reviewers, and countless others for spotting typos and giving comments. Special thanks to Vladas Sidoravicius who encouraged us to write a book on first-passage percolation. Last, we are indebted to our families for love and support.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/02

CHAPTER 2

The time constant and the limit shape

2.1. Subadditivity and the time constant

Let $e_1, \dots, e_d$ be the standard coordinate vectors of $\mathbb{Z}^d$. The first-order behavior of the growth of the passage time $T(0, ne_1)$ is described by the following theorem.

THEOREM 2.1 (Theorem 2.18 in [**123**]). *Assume that*

$$\mathbb{E} \min\{t_1, \dots, t_{2d}\} < \infty \tag{2.1}$$

where the t_i's are i.i.d copies of τ_e. Then there exists a constant $\mu(e_1) \in [0, \infty)$ (called the time constant) such that

$$\lim_{n\to\infty} \frac{T(0, ne_1)}{n} = \mu(e_1) \quad \textit{a.s. and in } L^1.$$

The proof of Theorem 2.1 is a classic application of the subadditive ergodic theorem that we now state. The version that we write here is due to Liggett [**142**] and suffices for our purposes. Several versions with different hypotheses, including the one with Kingman's original assumptions, can be found in [**142**].

THEOREM 2.2 (Subadditive ergodic theorem [**142**]). *Let $(X_{m,n})_{0\le m<n}$ be a family of random variables that satisfies the following conditions.*

(*a*) $X_{0,n} \le X_{0,m} + X_{m,n}$ *for all* $0 < m < n$.
(*b*) *The distributions of the sequences $(X_{m,m+k})_{k\ge1}$ and $(X_{m+1,m+k+1})_{k\ge1}$ are the same for all $m \ge 0$.*
(*c*) *For each $k \ge 1$, the sequence $(X_{nk,(n+1)k})_{n\ge1}$ is stationary.*
(*d*) $\mathbb{E}X_{0,1} < \infty$ *and* $\mathbb{E}X_{0,n} > -cn$ *for some finite constant c.*

Then

$$\lim_{n\to\infty} \frac{X_{0,n}}{n} \textit{ exists a.s. and in } L^1. \tag{2.2}$$

Furthermore, if the stationary sequences in (*c*) *are also ergodic, then the limit in* (2.2) *is constant a.s. and equal to*

$$\lim_{n\to\infty} \frac{\mathbb{E}X_{0,n}}{n} = \inf_n \frac{1}{n}\mathbb{E}X_{0,n}.$$

One can find different proofs of the subadditive ergodic theorem in the literature, and some of them are in standard books of probability theory (for instance [**79**, Section 6.6]). We will discuss the original proof given by Kingman in Section 2.6. We now show how Theorem 2.1 is a straightforward consequence of Theorem 2.2.

PROOF OF THEOREM 2.1. We apply the subadditive ergodic theorem to

$$X_{m,n} = T(me_1, ne_1).$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

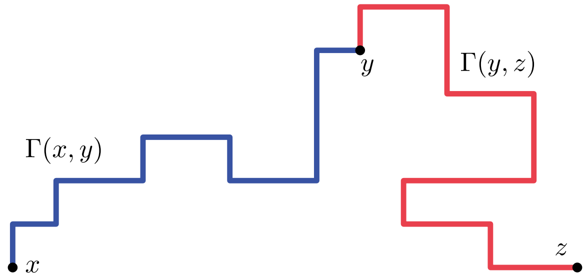


FIGURE 2. The minimizing paths $\Gamma(x, y)$ and $\Gamma(y, z)$ and their concatenation. As a minimizing path from x to z does not need to pass through y, we have $T(x, z) \leq T(x, y) + T(y, z)$.

It is not difficult to verify conditions (a) through (d). To check (a), which is just the triangle inequality, note that a path from 0 to ne_1 need not contain me_1, while a concatenation of paths from 0 to me_1 and from me_1 to ne_1 yields a path from 0 to ne_1. Therefore, as in Figure 2,

$$T(0, ne_1) \leq T(0, me_1) + T(me_1, ne_1).$$

Items (b), (c), and ergodicity follow directly from the fact that the environment is i.i.d.; thus it is invariant under horizontal shifts of $\mathbb{Z}^d$. $\mathbb{E}X_{0,n} > -cn$ holds, as passage times are non-negative. It remains to show that $\mathbb{E}T(0, e_1) < \infty$. This follows from assumption (2.1), as there are $2d$ disjoint deterministic paths $\Gamma_1, \ldots, \Gamma_{2d}$ in $\mathbb{Z}^d$ joining 0 to e_1. Indeed

$$T(0, e_1) \leq \min\{T(\Gamma_1), \ldots, T(\Gamma_{2d})\}.$$

Order them in such a way that Γ_1 is the path with the largest number (say B) of edges. Then

$$\mathbb{P}(T(0, e_1) > s) \leq \prod_{i=1}^{2d} \mathbb{P}(T(\Gamma_i) > s) \leq \mathbb{P}(T(\Gamma_1) > s)^{2d},$$

and

$$\mathbb{P}(T(\Gamma_1) > s) \leq B\mathbb{P}(\tau_e > s/B).$$

The second display comes from the fact that if $T(\Gamma_1) > s$ then at least one of the edges in Γ_1 must have a passage time larger than s/B. Combining the previous two inequalities and setting $Y = \min\{t_1, \ldots, t_{2d}\}$, we obtain

$$\mathbb{P}(T(0, e_1) > s) \leq B^{2d}\mathbb{P}(\tau_e > s/B)^{2d} = B^{2d}\mathbb{P}(Y > s/B), \tag{2.3}$$

which implies that $\mathbb{E}T(0, e_1) < \infty$ and proves the desired result. □

The following lemma comes directly from (2.3) and the observation that any path from 0 to x must contain at least one edge incident to 0.

LEMMA 2.3. *Let $k \geq 1$. Then $\mathbb{E}\min[t_1, \ldots, t_{2d}]^k < \infty$ if and only if $\mathbb{E}T(0, x)^k < \infty$ for all $x \in \mathbb{Z}^d$.*

The convergence in probability of the normalized passage time $T(0, ne_1)/n$ was first proved in two dimensions in the original paper of Hammersley and Welsh [**108**] under the assumption of finite mean of the random variable τ_e. Also under the

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

assumption of $\mathbb{E}\tau_e < \infty$, this convergence in probability was strengthened to a.s. and L^1 convergence by Kingman [**128**], using his subadditive ergodic theorem.

The condition (2.1) is necessary to have almost sure or L^1 convergence for $T(0, ne_1)/n$. Indeed, if (2.1) does not hold, denoting by $\tau_1^{(n)}, \ldots, \tau_{2d}^{(n)}$ the edge weights of edges incident to the vertex $2ne_1$, then for any $c > 0$, the events

$$A_n := \{\min\{\tau_1^{(n)}, \ldots, \tau_{2d}^{(n)}\} > cn\}$$

are independent and satisfy

$$\sum_n \mathbb{P}(A_n) = \infty.$$

Thus, an application of the Borel-Cantelli lemma shows that with probability one $\limsup_n T(0, ne_1)/n > c$.

Under no assumption for the τ_e's (except that they are nonnegative and i.i.d.), in particular not assuming (2.1), Kesten [**123**, Page 137], Cox-Durrett [**62**] and Wierman [**176**] establish the existence of a constant $\bar{\mu}(e_1)$ such that

$$\lim_{n\to\infty} T(0, ne_1)/n = \bar{\mu}(e_1) \text{ in probability.}$$

Clearly, if (2.1) holds, then $\bar{\mu}(e_1) = \mu(e_1)$.

We now gather more information on the time constant $\mu(e_1)$. By considering the direct path from 0 to ne_1 and using the law of the large numbers, we see that $\mu(e_1)$ satisfies

$$0 \leq \mu(e_1) \leq \mathbb{E}\tau_e.$$

That strict inequality does not always hold is seen by taking $\tau_e = 1$ a.s.. However, if the distribution F of the passage times has at least two points in its support, one can prove the following result.

THEOREM 2.4 (Hammersley-Welsh [**108**]). *If F is not a trivial distribution, then $\mu(e_1) < \mathbb{E}\tau_e$.*

PROOF. Choose $a < b$ such that $0 < F(a) \leq F(b) < 1$. Pick $n_0 > 2a/(b-a)$. Let $e(1), \ldots, e(n_0)$ be the n_0 edges following the straight line segment from 0 to n_0e_1 and let $f(1), \ldots, f(n_0+2)$ be the n_0+2 edges beginning with the edge from 0 to e_2, following the straight line segment from e_2 to $e_2 + n_0e_1$, and ending with the edge from $e_2 + n_0e_1$ to n_0e_1. On the positive probability event

$$\{\tau_{e(i)} \geq b, \tau_{f(j)} \leq a \text{ for all } i, j\}$$

one has $T(0, n_0e_1) < \tau_{e(1)} + \ldots + \tau_{e(n_0)}$. Thus, $\mathbb{E}T(0, n_0e_1) < n_0\mathbb{E}\tau_e$ and

$$\mu(e_1) = \inf_{n\geq 1} \frac{\mathbb{E}T(0, ne_1)}{n} \leq \frac{\mathbb{E}T(0, n_0e_1)}{n_0} < \mathbb{E}\tau_e.$$

□

We now look at lower bounds for the time constant. If $F(0) > 0$; that is, if we have edges that are cost-free to cross, one may wonder if μ is equal to 0 and the growth of the time constant is in fact not linear. This issue is handled in the next theorem. Let $p_c = p_c(d)$ be the critical probability for Bernoulli bond percolation

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

in $\mathbb{Z}^d$. It is defined as

$$p_c := \sup\Big\{p \in [0,1] : \mathbb{P}\big(\exists \text{ infinite self-avoiding path of edges } e \text{ with } \tau_e = 0\big) = 0\Big\}, \tag{2.4}$$

in a model in which edge-weights τ_e have $\mathbb{P}(\tau_e = 0) = p$. The significance of $F(0) < p_c$ is that it ensures that a.s. there is no self-avoiding infinite path of zero-weight edges. It is known that $p_c \in (0,1)$ for all $d \geq 2$. See the standard text [**102**] for more on Bernoulli percolation.

THEOREM 2.5 (Kesten [**123**], Theorem 6.1). *For FPP on $\mathbb{Z}^d$, $\mu(e_1) > 0$ if and only if $F(0) < p_c$.*

Returning to the issue of convergence to the time constant, one can extend Theorem 2.1 with a similar proof to arbitrary directions, replacing e_1 with any x having rational coordinates. This establishes existence of a homogeneous function $\mu : \mathbb{Q}^d \to \mathbb{R}$ such that, for any $x \in \mathbb{Q}^d$,

$$\lim_{n\to\infty} \frac{T(0,nx)}{n} = \mu(x) \quad \text{a.s. and in } L^1. \tag{2.5}$$

The reader should see (2.5) as an analogue of a law of large numbers. In Chapter 3, we will discuss the fluctuations of $T(0,nx)$ around $n\mu(x)$, and we will argue that, in general, a central limit theorem with Gaussian fluctuations does not hold. It is not difficult to establish the following properties of μ for $x, y \in \mathbb{Q}^d$ and $c \in \mathbb{Q}$:

(1) $\mu(x+y) \leq \mu(x) + \mu(y)$.
(2) $\mu(cx) = |c|\mu(x)$.
(3) μ is invariant under symmetries of $\mathbb{Z}^d$ that fix the origin.
(4) μ is Lipschitz on bounded subsets of $\mathbb{Q}^d$, so it has a unique continuous extension to $\mathbb{R}^d$.

Furthermore, these properties imply the following:

THEOREM 2.6. *$\mu(e_1) > 0$ if and only if $\mu(x) > 0$ for all $x \in \mathbb{R}^d \setminus \{0\}$.*

Subadditivity is a useful tool for showing the existence of the time constant; however, it gives no explicit algebraic expression for μ. The determination of $\mu(e_1)$ as a function of F is a fundamental and difficult problem in FPP.

QUESTION 2.1.1. Find a nontrivial distribution for which one can explicitly determine $\mu(e_1)$.

It is not clear what "determine $\mu(e_1)$" means. In FPP, we do not know any nontrivial distribution that allows exact computations, and maybe there are none. In Chapter 7, we discuss similar solvable models where explicit computations are possible. A related question is the following.

(2.6) *Given F and $\tilde{F}$, how different are their time constants $\mu(e_1)$ and $\tilde{\mu}(e_1)$?*

In general, we lack strong information about how the time constant (and the limiting shape, defined in Section 2.3) changes under small perturbations of the edge-weight distribution. If one could derive results in this direction, perhaps the establishment of various conjectures about the limit shape (for example, curvature)

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

could be made easier, or reduced to finding some special class of distributions for which the properties are explicitly derivable. The best current results on stability, dating back over thirty years, say simply that the time constant is a continuous function of the edge-weight distribution.

THEOREM 2.7 (Cox-Kesten [**63**], Kesten [**123**]). *The time constant $\mu(e_1)$ is continuous under weak convergence of i.i.d. distributions. That is, if (F_n) is a sequence of distribution functions for the edge weight τ_e with $F_n \Rightarrow F$, and if $(\mu_n(e_1))$, $\mu(e_1)$ denote the respective time constants, then*

$$\lim_n \mu_n(e_1) = \mu(e_1).$$

One improvement of Theorem 2.7 is the recent work of Garet, Marchand, Procaccia, and Théret [**95**], which establishes an analogous continuity result in the case that the edge-weights are allowed to assume the value $\tau_e = +\infty$. On the other hand, when comparing distributions F, $\tilde{F}$ which obey certain stochastic orderings, much more can be said [**63**, Theorem 3] (see also [**108**, Section 6.4] and [**61**]). When $\tilde{F}(t) \le F(t)$ for all $t \in \mathbb{R}$, one can directly couple the passage times to obtain $\mu(e_1) \le \tilde{\mu}(e_1)$.

The question (2.6) was considered by several authors [**31**, **123**, **143**]. An answer came with the work of van den Berg and Kesten [**31**], who proved the strict inequality $\mu(e_1) < \tilde{\mu}(e_1)$ if F is strictly more variable than $\tilde{F}$ (defined below). Here we follow an extension of the van den Berg - Kesten comparison theorem provided by Marchand [**143**].

DEFINITION 2.8. Let F and $\tilde{F}$ be two distributions on $\mathbb{R}$. We say that F is more variable than $\tilde{F}$ if for every concave increasing function $\phi : \mathbb{R} \to \mathbb{R}$, one has

$$\int \phi \, dF \le \int \phi \, d\tilde{F},$$

provided the two integrals converge absolutely. If in addition $F \neq \tilde{F}$, then we say that F is strictly more variable than $\tilde{F}$.

EXAMPLE 2.9. If $\tilde{F}$ stochastically dominates F; that is, $\tilde{F}(x) \le F(x)$ for all $x \in \mathbb{R}$, then F is more variable than $\tilde{F}$.

EXAMPLE 2.10. For any non-negative random variable X and any constant M the distribution of $\min(X, M)$ is more variable than the distribution of X.

EXAMPLE 2.11. Let a, ϵ be real numbers. For $0 < p < 1$ and ρ a probability measure, consider the probability measures $\tilde{\nu} = p\delta_a + (1-p)\rho$ and $\nu = (p/2)\delta_{a-\epsilon} + (p/2)\delta_{a+\epsilon} + (1-p)\rho$. Then the distribution function F_ν corresponding to ν is strictly more variable than $F_{\tilde{\nu}}$.

THEOREM 2.12 (van den Berg-Kesten [**31**], Marchand [**143**]). *Assume that $d = 2$ and suppose that $F(0) < p_c$. If F is strictly more variable than $\tilde{F}$ then $\mu(e_1) < \tilde{\mu}(e_1)$.*

REMARK 2.13. Using Theorem 2.12, one can even exhibit pairs $(F, \tilde{F})$ such that $\mu(e_1) < \tilde{\mu}(e_1)$ but $\int x \, \mathrm{d}F(x) > \int x \, \mathrm{d}\tilde{F}(x)$, as the following example from [**31**, Remark 2.22] illustrates. Consider F to be uniform on the interval $[1/2 + \epsilon, 5/2 + \epsilon]$ with $\epsilon > 0$ small and $\tilde{F}$ to be uniform on the set $\{1, 2\}$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

For $x \neq e_1$ a version of $\mu(x) < \tilde{\mu}(x)$ can also be found in [**31**,**143**]. Theorem 2.12 also comes with the following remark. Suppose that the distribution of τ_e has unbounded support. One might think that there exists a threshold M such that an optimal path between 0 and ne_1, n large, never takes order n number of edges with weights above M. However, the theorem above implies that, if one truncates the passage time at level M, the time constant of the truncated model is strictly less than the original one. Thus, it is more efficient for paths to use a certain positive proportion of edges with very large passage time than to avoid them.

This remark leads to the following question, which seems to be open. In the case of passage times with exponential tails it was considered by Nakajima [**146**].

QUESTION 2.1.2. Suppose that the support of the distribution of τ_e is unbounded. Let

$$X_n = \max\{\tau_e : e \text{ is in a geodesic from } 0 \text{ to } ne_1\}.$$

How does X_n scale with n?

The assumption $F(0) < p_c$ in Theorem 2.12 cannot be removed because of Theorem 2.5. Now, let I be the infimum of the support of F. In higher dimensions, a version of Theorem 2.12 above is known [**31**] if $F(I) < \vec{p}_c$, where $\vec{p}_c$ is the critical probability for directed edge percolation. The only missing case at this point is described in the next question.

QUESTION 2.1.3. Extend Theorem 2.12 to the case $d > 2$, $I > 0$ and $\vec{p}_c \leq F(I)$.

REMARK 2.14. Marchand's original approach did not work in higher dimensions, as she used large deviation estimates for supercritical oriented percolation available only in dimension two at the time of her paper (see the estimate in [**143**, Page 1014]). New estimates were obtained for $d > 2$ in [**96**, Proposition 2] recently which may overcome this obstacle.

A question in the spirit of the above is whether the expected passage time is monotone. By the convergence to the time constant, we know that the expected value of $T(0, ne_1)$ asymptotically grows at least linearly in n, but is it always strictly non-decreasing? Hammersley and Welsh [**108**] conjectured that for general edge-weights, monotonicity should hold:

$$\mathbb{E}T(0,(n+1)e_1) > \mathbb{E}T(0,ne_1) \quad \text{for all } n \geq 0. \tag{2.7}$$

A version of (2.7) was proven by Gouéré in [**101**] with an explicit lower bound for the difference between the two expectations, under the assumption that $\mathbb{P}(a < \tau_e < b) = 1$. Here a and b are positive constants with $b \leq 2a$. (Other slightly less restrictive conditions are given in [**101**, p. 566-567].) See also [**13**], where (2.7) is proved on the upper half-plane. On the other hand, a counterexample to a statement similar to (2.7) was proved in [**28**] in the special case of Bernoulli $0, 1$ edge-weights on $\mathbb{Z}^2$, where $\mathbb{P}(\tau_e = 1)$ is sufficiently small.

2.2. The time constant through a homogenization problem

Another way to interpret the time constant was recently provided by Krishnan [**131**] in FPP and by Georgiou, Rassoul-Agha, and Seppäläinen [**97**] in last-passage percolation. We briefly describe it here.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

The idea is to interpret the passage time as a solution of an optimal-control problem. Define

$$A := \{\pm e_1, \dots, \pm e_d\}.$$

We think of A as the collection of possible directions to exit a vertex. We now write $\tau(v, \alpha)$ to refer to the weight τ_e at $v \in \mathbb{Z}^d$ along the direction $\alpha \in A$. It is now possible to check that

$$T(0, x) = \inf_{\alpha \in A} \{T(0, x + \alpha) + \tau(x, \alpha)\}.$$

DEFINITION 2.15. For a function $f : \mathbb{Z}^d \to \mathbb{R}$, let

$$Df(x, \alpha) = f(x + \alpha) - f(x)$$

be its discrete derivative at x in the direction $\alpha \in A$.

We write Ω for the probability space and $\sigma_v : \Omega \to \Omega$ for the shift that translates the random variables τ_e by v. Define $\Omega_d = \mathbb{Z}^d \times \Omega$ and

$$\mathcal{F} := \left\{ f : \Omega_d \to \mathbb{R} \;\middle|\; \begin{array}{l} Df(x + v, \alpha)(\omega) = Df(x, \alpha)(\sigma_v \omega),\ \forall v, x \in \mathbb{Z}^d, \forall \omega \in \Omega \\ \mathbb{E}[Df(x, \alpha)] = 0\ \forall x \in \mathbb{Z}^d \text{ and } \alpha \in A \end{array} \right\}.$$

For $p \in \mathbb{R}^d$, $x \in \mathbb{Z}^d$ and $f \in \mathcal{F}$ define

$$H(f, p, x) = \sup_{\alpha \in A} \left\{ \frac{-Df(x, \alpha) - p \cdot \alpha}{\tau(x, \alpha)} \right\}. \tag{2.8}$$

THEOREM 2.16 (Krishnan [**131**], Theorem 2.3). *Assume that there exist m, M such that $0 < m < \tau_e < M$ a.s.. Then $\mu(x)$ solves the following Hamilton-Jacobi equation:*

$$\bar{H}(D\mu(x)) = 1,$$
$$\mu(0) = 0,$$

where $\bar{H}(p)$ is a convex, coercive, Lipschitz continuous function given by

$$\bar{H}(p) = \inf_{f \in \mathcal{F}} \operatorname*{ess\,sup}_{\omega \in \Omega} \sup_{x \in \mathbb{Z}^d} H(f, p, x)(\omega).$$

Furthermore, $\bar{H}(p)$ is the dual norm of $\mu(x)$ on $\mathbb{R}^d$, defined by

$$\bar{H}(p) = \sup_{\mu(x) = 1} p \cdot x.$$

Although the theorem above gives a different characterization for the time constant, it has not yet been used to get a better understanding of $\mu(e_1)$ as a function of F. An algorithm for finding a minimizer for the above variational formula is explained in [**131**, Section 6]. It would be useful to extend these ideas to give more geometric information on the limit shape.

2.3. The limit shape: Cox-Durrett shape theorem

For each unit vector $x \in \mathbb{R}^d$, we define the time constant $\mu(x)$ in direction x by (2.5) and item 4 below it. In this section, we will see how the function μ describes the first-order approximation of the random ball $B(t)$ as t goes to infinity. For this, we define the limit shape $\mathcal{B}$ as

$$\mathcal{B} = \{x \in \mathbb{R}^d : \mu(x) \leq 1\}. \tag{2.9}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

The main result of the section is the shape theorem, Theorem 2.17.

Let $\mathcal{M}$ be the set of Borel probability measures on $[0, \infty)$ satisfying

$$\mathbb{E} \min\{t_1^d, \ldots, t_{2d}^d\} < \infty, \tag{2.10}$$

where the t_i's, are independent copies of τ_e with

$$F(0) < p_c, \tag{2.11}$$

where p_c (from (2.4)) is the threshold for Bernoulli bond percolation on $\mathbb{Z}^d$. If S is a subset of $\mathbb{R}^d$ and $r \in \mathbb{R}$ we write $rS = \{rs \ : \ s \in S\}$.

THEOREM 2.17 (Cox and Durrett [**62**]). *For each $\nu \in \mathcal{M}$, there exists a deterministic, convex, compact set $\mathcal{B}$ in $\mathbb{R}^d$ such that for each $\epsilon > 0$,*

$$\mathbb{P}\left((1-\epsilon)\mathcal{B} \subset \frac{B(t)}{t} \subset (1+\epsilon)\mathcal{B} \textit{ for all large } t\right) = 1. \tag{2.12}$$

Furthermore, $\mathcal{B}$ has non-empty interior and has the symmetries of $\mathbb{Z}^d$ that fix the origin.

REMARK 2.18. If (2.11) does not hold, edges with zero passage time form large instantaneous “highways.” In this case, Theorem 2.5 says that the time constant is equal to zero. As for the limit shape, one has $\mathcal{B} = \mathbb{R}^d$ [**123**, Theorem 1.10]. Precisely, under assumption (2.10), we have $F(0) \geq p_c$ if and only if for every $M > 0$

$$\mathbb{P}\left(\{x : |x| \leq M\} \subset \frac{B(t)}{t} \text{ for all large } t\right) = 1.$$

REMARK 2.19. For $x \in \mathbb{Z}^d$, let $m(x)$ be the minimum of τ_e over all edges e incident to x. If (2.10) fails, then one can show that for any $C > 0$,

$$\sum_{x \in \mathbb{Z}^d} \mathbb{P}\Big(m(x) > C|x|\Big) = \infty.$$

Now note that $T(0,x) \geq m(x)$ and the random variables $\{m(x) : x \in 2\mathbb{Z}^d\}$ are independent. Then we can apply the Borel-Cantelli lemma to obtain

$$\frac{T(0,x)}{|x|} > C \text{ for infinitely many } x \in \mathbb{Z}^d, \text{ a.s..}$$

Thus, if (2.10) fails, (2.12) also does not hold.

REMARK 2.20. With Theorem 2.17, Cox and Durrett provided necessary and sufficient conditions for the shape theorem to hold in FPP. The first shape theorem, however, was proven in the seminal work of Richardson [**155**] in 1973. We will come back to discuss Richardson's class of models in Chapter 6. See Figure 3 for some illustrations of limit shapes.

REMARK 2.21. The continuity result for the time constant, Theorem 2.7, has an analogue for limit shapes. Specifically, if $F_n \Rightarrow F$, then the limit shapes $\mathcal{B}_n$ for the FPP models with edge-weight distributions F_n converge in Hausdorff distance to the limit shape $\mathcal{B}$ of the FPP model with edge-weight distribution F. While this preceding result is not strong enough to preserve curvature of the limiting shape under weak limits, it does guarantee a certain semicontinuity property of the set of its extreme points (see Section 2.8 for these definitions). In [**67**], this was used to establish the existence of limit shapes with arbitrarily many extreme points for

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

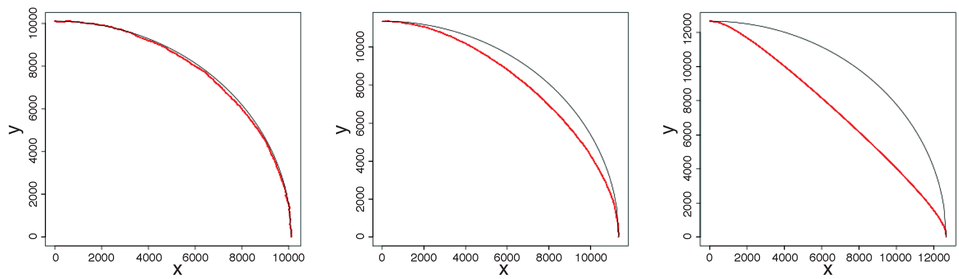


FIGURE 3. Simulation of the two-dimensional ball $B(t)$, intersected with the first quadrant, when $t = 15000$. The passage times (τ_e) are distributed according to an exponential random variable of mean 1 plus a constant C. Here, $C = 0$ (left), $C = 0.5$ (middle), and $C = 4$ (right). The outer curve is a circle of the same e_1-radius as $B(t)$. [These figures originally appeared in: "First Package Percolation on $\mathbb{Z}^2$: A Simulation Study" (S.-E. Alm and M. Deijfen), Journal of Statistical Physics **161** (2015), no. 3, pp. 657–698. ©2015 Springer Science + Business Media New York. Reprinted with permission of Springer.]

some nonatomic edge-weight distributions; improvements to Theorem 2.7 could be useful for similar constructions.

The idea of the proof of Theorem 2.17 is to first use subadditivity to demonstrate the linear growth of $B(t)$ in a fixed rational direction. This implies that, with probability one, we have the correct growth rate in a countable dense set of directions simultaneously. To obtain the full result from this, we need some bound which allows us to interpolate between these directions, ensuring that the convergence occurs along all rays uniformly with probability one.

There are different ways to implement this interpolation step. Here we sketch a method which first appeared in [**109**] (inspired by [**36**]) and has the advantage of applying (with modifications) to the stationary ergodic case of FPP. This proof method was further used for the Busemann shape theorem (see Lemma 5.12), which appeared in [**68**, Theorem 4.3].

LEMMA 2.22 (Difference estimate). *Let $\nu \in \mathcal{M}$. Then there exists a constant $\kappa < \infty$ such that, for any $x \in \mathbb{Z}^d$,*

$$\mathbb{P}\left(\sup_{\substack{z \in \mathbb{Z}^d \\ z \neq x}} \frac{T(x,z)}{\|x-z\|_1} < \kappa \right) > 0. \tag{2.13}$$

IDEA OF PROOF. This proof follows the line of a similar estimate in [**62**]. If $\mathbb{E}\tau_e^d < \infty$, the proof follows by noting that there exist $2d$ edge-disjoint paths $\{\Gamma_i\}$ between x and z, each of length order $\|x - z\|_1$. For the event

$$\{T(x,z) > \kappa \|x - z\|_1\}$$

to occur, each of these paths must have $T(\Gamma_i) > \kappa\|x-z\|_1$. Using standard estimates for i.i.d. sums, this probability is small for κ sufficiently large — in fact, we can get

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

an estimate which is summable in z. Using the Borel-Cantelli lemma (and adjusting the constant κ upwards if necessary) allows us to complete the proof.

In the general case, we consider a sparse lattice $C\mathbb{Z}^d$ such that between "neighboring" vertices of this lattice, there exist $2d$ disjoint paths lying inside cells of side length of order C. In particular, paths corresponding to well-separated pairs of "neighboring" vertices are disjoint, and their passage times are independent. The smallest passage time among these $2d$ paths is an upper bound for the passage time between "neighboring" vertices; we treat this as the passage time of a "renormalized" edge of the sparse lattice. Mimicking the argument of the preceding paragraph, then extending the bound to the rest of $\mathbb{Z}^d$, completes the proof. $\square$

Proof of Theorem 2.17. We will call an x in $\mathbb{Z}^d$ for which the event appearing in (2.13) occurs a "good" vertex. We can immediately leverage the information in Lemma 2.22 to show the following.

Claim 1. Let $\zeta \in \mathbb{Z}^d \setminus \{0\}$. For a given realization of edge-weights, denote by (n_k) the sequence of natural numbers such that $n_k\zeta$ is a good vertex. Then with probability one, the sequence (n_k) is infinite and $\lim_{k\to\infty}(n_{k+1}/n_k) = 1$.

To justify the claim, note that the ergodic theorem implies that the sequence (n_k) is infinite a.s.. Let B_m denote the event that $m\zeta$ is a good vertex. Then

$$\frac{k}{n_k} = \frac{1}{n_k}\sum_{i=1}^{n_k} \mathbf{1}_{B_i};$$

the right side converges to the (positive) probability in (2.13) by the ergodic theorem. Thus,

$$\frac{n_{k+1}}{n_k} = \left(\frac{n_{k+1}}{k+1}\right)\left(\frac{k}{n_k}\right)\left(\frac{k+1}{k}\right) \to 1$$

a.s.. This proves the claim.

Let Ξ_1 denote the event that

$$\lim_{n\to\infty} \frac{T(0,nq)}{n} = \mu(q)$$

for all q having rational coordinates; let Ξ_2 denote the event that for every $\zeta \in \mathbb{Z}^d$, the sequence (n_k) defined in Claim 1 is infinite and that the ratio of successive terms tends to one. From here, the proof of Theorem 2.17 proceeds by contradiction. Assume the shape theorem does not hold in the following equivalent form:

$$\limsup_{\|x\|_1\to\infty} \frac{|T(0,x)-\mu(x)|}{\|x\|_1} = 0 \text{ a.s..}$$

Then there exists $\delta > 0$ and a collection of edge-weight configurations D_δ with $\mathbb{P}(D_\delta) > 0$ such that, for every outcome in D_δ, there are infinitely many vertices $x \in \mathbb{Z}^d$ with

$$|T(0,x)-\mu(x)| > \delta\|x\|_1. \tag{2.14}$$

Since $\mathbb{P}(\Xi_1) = 1 = \mathbb{P}(\Xi_2)$, the event $D_\delta \cap \Xi_1 \cap \Xi_2$ contains some outcome ω; we claim that ω has contradictory properties. On ω, there must exist a sequence $(x_i) \subset \mathbb{Z}^d$ satisfying the condition in (2.14). We can assume that $x_i/\|x_i\|_1$ converges to some y with $\|y\|_1 = 1$ by compactness of the unit sphere. Let $\delta' > 0$ be arbitrary; we will fix its value at the end of the proof. We first choose some large N such that

$$\|x_n/\|x_n\|_1 - y\|_1 < \delta'$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

and such that

$$|\mu(x_n) - \|x_n\|_1 \mu(y)| < \delta \|x_n\|_1/2$$

for $n > N$. Then we have for $n > N$ (using our assumption (2.14):

$$|T(0, x_n) - \|x_n\|_1 \mu(y)| > \delta \|x_n\|_1/2. \tag{2.15}$$

Next, we set up a sequence of approximating good vertices. We find some $z \in \mathbb{R}^d$ with $\|z\|_1 = 1$ such that $\|z - y\|_1 < \delta'$, and the additional property that $z = x/M$ for some $x \in \mathbb{Z}^d$ and some positive integer M. This can be done because vectors with rational coordinates are dense in the unit sphere. On ω, there must exist a sequence (n_k) such that $n_k M z$ is a good vertex and such that n_{k+1}/n_k tends to one. For any n, there exists a value of k such that

$$n_{k+1} M \geq \|x_n\|_1 \geq n_k M;$$

denote this value by $k(n)$. Finally, fix $K > 0$ such that $n_{k+1} < (1 + \delta')n_k$ and

$$|T(0, n_k M z)/(n_k M) - \mu(z)| < \delta'$$

for all $k > K$. We now let $n > N$ be large enough that $k(n) > K$.

Before completing the calculation here, it is worth considering where the contradiction will arise. We have (essentially by assumption) that $T(0, ny) - n\mu(y)$ is at least of order n for infinitely many n. Since μ is a norm, $\mu(y)$ and $\mu(z)$ are arbitrarily close and since infinitely many of the $\{nz\}$ are good vertices, $T(0, ny)$ and $T(0, nz)$ are arbitrarily close on scale n. Thus $|T(0, nz) - n\mu(z)|$ is large – but this is counter to the properties assumed for ω.

To turn the above into a rigorous estimate, write k for $k(n)$ and expand

$$\begin{aligned} \left| \frac{T(0, x_n)}{\|x_n\|_1} - \mu(y) \right| \leq & \left| \frac{T(0, x_n) - T(0, n_k M z)}{\|x_n\|_1} \right| + \frac{T(0, n_k M z)}{n_k M} \left(1 - \frac{n_k M}{\|x_n\|_1} \right) \\ & + \left| \frac{T(0, n_k M z)}{n_k M} - \mu(z) \right| + |\mu(z) - \mu(y)|. \end{aligned}$$

There are four terms on the right side of the above, which we number from left to right and bound individually in terms of δ'.

Term 1. Since $n > N$ and $k > K$, one has $n_k M \leq \|x_n\|_1 \leq (1 + \delta')n_k M$, that $\|n_k M y - n_k M z\|_1 \leq \delta' n_k M$, and that $\|x_n/\|x_n\|_1 - y\|_1 < \delta'$. Therefore,

$$\|x_n - n_k M z\|_1 \leq 3\delta' \|x_n\|_1.$$

Using the fact that $n_k M z$ is a good vertex yields

$$|T(0, x_n) - T(0, n_k M z)| \leq \kappa \|x_n - n_k M z\|_1 \leq 3\kappa\delta' \|x_n\|_1.$$

Term 2. The relationship between $n_k M$ and $\|x_n\|_1$ given in the Term 1 estimates yield an upper bound for the second factor of Term 2. By the fact that $k > K$, we can bound the first factor. The overall bound is

$$[\mu(z) + \delta'] \left(1 - (1 + \delta')^{-1} \right).$$

Term 3. By the fact that k is chosen greater than K, this term is bounded above by δ'.

Term 4. If μ is identically zero, this term is trivially zero. If μ is not identically zero, it is a norm on $\mathbb{R}^d$ and is thus bounded by the $\| \cdot \|_1$ norm:

$$c_L \| \cdot \|_1 \leq \mu(\cdot) \leq c_U \| \cdot \|_1.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Since $\|z-y\|_1 < \delta'$, Term 4 is bounded above by δ' times a constant depending only on μ.

We have therefore bounded the left side of (2.15) by an expression of the form $f(\delta')\|x_n\|_1$, where f tends to zero as $\delta' \to 0$. Since δ' was arbitrary, we can choose it such that $f(\delta')\|x_n\|_1$ is smaller than the right side of (2.15). This contradiction proves the theorem.

□

2.4. Other shape theorems

In this section, we briefly discuss a few extensions of Theorem 2.17.

2.4.1. Shell passage times. As we saw in Remark 2.19, when (2.10) fails, we have with probability one

$$\limsup_{|v|\to\infty} \frac{T(0,v)}{|v|} = \infty.$$

Nevertheless, without any moment condition on τ_e, one can define a modified passage time $\hat{T}(u,v)$ such that the family of random variables $\hat{T}(u,v) - T(u,v)$ is tight and one has a shape theorem for the modified $\hat{T}$. This was first done by Cox-Durrett in dimension two and later extended by Kesten to all dimensions. Their construction goes as follows. Let $M \geq 0$ be large enough so that $F([0,M])$ is very close to 1. The collection of edges e such that $\tau_e \leq M$ is a highly super-critical percolation process, so if we denote by $\mathcal{C}_M$ its infinite cluster, each point $u \in \mathbb{Z}^d$ is a.s. surrounded by a small contour (or shell) $S(u) \subset \mathcal{C}_M$. Define $\hat{T}(u,v) = T(S(u), S(v))$ for $u, v \in \mathbb{Z}^d$. The times $\hat{T}(u,v)$ have finite moments of all orders because it is possible to travel from $S(u)$ to $S(v)$ using only edges e with $\tau_e \leq M$; thus their limit shape can be defined using the a.s. and L^1 limit given by the previous arguments. The details of this construction in $d > 2$ require certain topological properties of the exterior boundary of a subset of $\mathbb{Z}^d$. These properties were derived by Kesten and simplified in the work of Timár [**168**].

2.4.2. FPP in the supercritical percolation cluster. Another place where shape theorems have been proven is the setting in which we allow passage times to be infinite. This is equivalent to considering FPP on supercritical Bernoulli percolation clusters. When the edge weights are either 1 or ∞, the passage time is also known as the chemical distance.

In this setting, the benchmark is the work of Garet-Marchand [**92**], where analogues of Theorems 2.5 and 2.17 were proven under a moment condition $\mathbb{E}\tau_e^{\alpha(d)}\mathbf{1}_{\{\tau_e<\infty\}} < \infty$, where $\alpha(d) = 2(d^2+d-1)+\epsilon$; see hypothesis H_α on page 4 of [**92**]. Their results are also valid for stationary ergodic passage times, in the spirit of the work of Boivin [**36**].

In the i.i.d. case, in two independent works, Cerf-Théret [**47**] and Mourrat [**145**] recently removed all moment assumptions of [**92**], by proving a weak shape theorem. In his paper, Mourrat considers a model of a random walk in a random potential, but he discusses how his theorems easily extend to our setting (see [**145**, Section 11]). We describe their results below, as they are a nice compromise between the results of Garet-Marchand and shell passage times of Cox-Durrett and Kesten. Let $\mathcal{C}_\infty$ be the infinite cluster for the Bernoulli percolation. For any $x \in \mathbb{Z}^d$, let $x^* \in \mathbb{Z}^d$ be

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

the random point of $\mathcal{C}_\infty$ such that $\|x - x^*\|_1$ is minimal, with a deterministic rule to break any possible ties. Defining

$$T^*(x,y) = T(x^*, y^*),$$

one has existence of a time constant for the time T^*. For the statements, F is the distribution function of τ_e, supported on $[0,\infty]$.

THEOREM 2.23 (Cerf-Théret [**47**] Theorem 4; Mourrat [**145**], Theorem 1.2). *Suppose that* $\lim_{x\to\infty} F(x) > p_c$. *Then for any* $x \in \mathbb{Z}^d$, *there exists* $\mu^*(x)$ *such that*

$$\lim_n \frac{T^*(0,nx)}{n} = \mu^*(x) \quad \textit{in probability,}$$

and

$$\lim_n \frac{T(0,nx)}{n} = Z \qquad \textit{in law,}$$

where the distribution of Z *is given by* $\theta^2\delta_{\mu^*(x)} + (1-\theta^2)\delta_\infty$ *and* $\theta = \mathbb{P}(0 \in C_\infty)$.

For $t \geq 0$, put

$$B_t^* = \{z + u \;:\; z \in \mathbb{Z}^d, T^*(0,z) \leq t,\; u \in [-1/2, 1/2]^d\}.$$

There is a corresponding shape theorem for B_t^*.

THEOREM 2.24 (Cerf-Théret [**47**], Theorem 5; Mourrat [**145**], Theorem 1.2). *Suppose that* $\lim_{x\to\infty} F(x) > p_c$ *and* $F(0) < p_c$. *Then there exists a compact set* B^* *such that a.s.*

$$\lim_t \lambda^d \left(\frac{B_t^*}{t} \triangle B^* \right) = 0,$$

where λ^d *denotes the Lebesgue measure in* $\mathbb{R}^d$ *and* $A \triangle B$ *is the symmetric difference of sets* A *and* B.

2.5. The class of possible limit shapes

In this section, we address one of the questions presented in the introduction.

(2.16) *Which compact convex sets* C *arise as limit shapes?*

The question above is open in the i.i.d. setting. A conjecture is given by the following.

QUESTION 2.5.1. Show that if F is a continuous distribution then the limit shape is strictly convex.

Surprisingly, not even the following is known:

QUESTION 2.5.2. Show that the d-dimensional cube (ℓ^∞ ball) is not a possible limit shape for an FPP model with i.i.d. passage times.

REMARK 2.25. Interestingly, question (2.16) is solved by Häggström and Meester [**105**] in the case of translation-invariant (not necessarily i.i.d.) passage times (see Theorem 7.6). They establish that any non-empty compact, convex set C that is symmetric about the origin (that is, $x \in C \Rightarrow -x \in C$) is a limit shape for some FPP model with weights distributed according to a stationary (under translations of $\mathbb{Z}^d$) and ergodic measure. This is in sharp contrast to the i.i.d. case explained above.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

However, there is one class of weights where the limit shape is known in some directions. This collection was introduced by Durrett and Liggett [**81**] and further studied by Marchand [**143**], Zhang [**187**, **188**] and by Auffinger and Damron [**19**]. Its main feature is the presence of a flat edge for the limit shape, as we describe below. We will stick to dimension two in what follows.

Write *supp*(ν) for the support of ν, the distribution of τ_e. Let $\mathcal{M}_p$ be the set of measures ν that satisfy the following:

(1) $supp(\nu) \subseteq [1, \infty)$ and
(2) $\nu(\{1\}) = p \geq \vec{p}_c$,

where $\vec{p}_c$ is the critical parameter for oriented percolation on $\mathbb{Z}^2$ (see, e.g., [**80**]). In [**81**], it was shown that if $\nu \in \mathcal{M}_p$ then $\mathcal{B}$ has some flat edges. The precise location of these edges was found in [**143**]. To describe this, write $\mathcal{B}_1$ for the closed ℓ^1 unit ball:

$$\mathcal{B}_1 = \{(x, y) \in \mathbb{R}^2 \ : \ |x| + |y| \leq 1\}$$

and write *int* $\mathcal{B}_1$ for its interior. For $p > \vec{p}_c$, let α_p be the asymptotic speed of oriented percolation [**80**], define the points

$$M_p = \left(\frac{1}{2} - \frac{\alpha_p}{\sqrt{2}}, \frac{1}{2} + \frac{\alpha_p}{\sqrt{2}}\right) \text{ and } N_p = \left(\frac{1}{2} + \frac{\alpha_p}{\sqrt{2}}, \frac{1}{2} - \frac{\alpha_p}{\sqrt{2}}\right) \tag{2.17}$$

and let $[M_p, N_p]$ be the line segment in $\mathbb{R}^2$ with endpoints M_p and N_p. For symmetry reasons, the following theorem is stated only for the first quadrant.

THEOREM 2.26 (Durrett-Liggett [**81**], Marchand [**143**]). *Let* $\nu \in \mathcal{M}_p$.

(1) $\mathcal{B} \subset \mathcal{B}_1$.
(2) *If* $p < \vec{p}_c$ *then* $\mathcal{B} \subset$ *int* $\mathcal{B}_1$.
(3) *If* $p > \vec{p}_c$ *then* $\mathcal{B} \cap [0, \infty)^2 \cap \partial\mathcal{B}_1 = [M_p, N_p]$.
(4) *If* $p = \vec{p}_c$ *then* $\mathcal{B} \cap [0, \infty)^2 \cap \partial\mathcal{B}_1 = (1/2, 1/2)$.

The angles corresponding to points in the line segment $[M_p, N_p]$ are said to be in the percolation cone; see Figure 4 below.

Let $\beta_p := 1/2 + \alpha_p/\sqrt{2}$; that is, define β_p as the e_1-coordinate of N_p. Convexity and symmetry of the limit shape imply that $1/\mu(e_1) \geq \beta_p$. A non-trivial statement about the edge of the percolation cone came in 2002 when Marchand [**143**, Theorem 1.4] proved that this inequality is in fact strict:

$$1/\mu(e_1) > \beta_p.$$

In other words, Marchand's result says that the line that goes through N_p and is orthogonal to the e_1-axis is not a tangent line of $\partial\mathcal{B}$. The following theorem builds on Marchand's result and technique and states that at the edge of the percolation cone, one cannot have a corner.

THEOREM 2.27 (Auffinger-Damron [**19**]). *Let* $\nu \in \mathcal{M}_p$ *for* $p \in [\vec{p}_c, 1)$. *The boundary* $\partial\mathcal{B}$ *is differentiable at* N_p.

REMARK 2.28. Theorem 2.27 is stated for the single point N_p but, due to symmetry, it is valid for M_p and for the reflections of these two points about the coordinate axes.

The theorem above implies that any measure in $\mathcal{M}_p$ has a non-polygonal limit shape. The first example of a non-polygonal limit shape was discovered by Damron-Hochman [**67**]. In that paper, the authors also provide examples of (a) $\nu \in \mathcal{M}_p$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

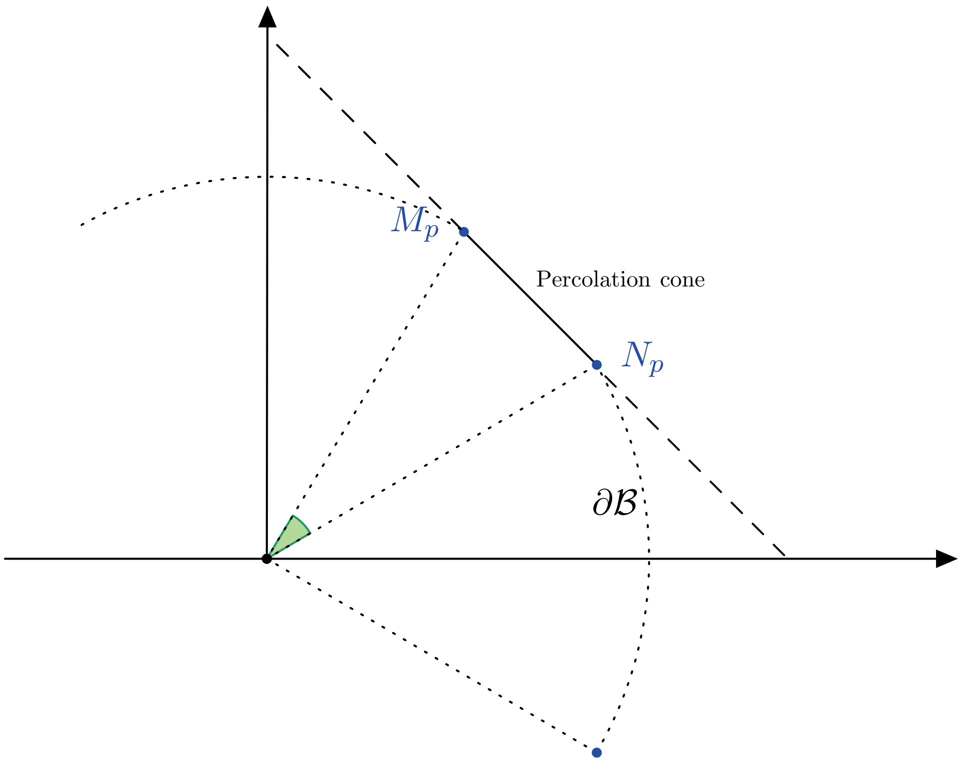


FIGURE 4. Illustration of a portion of the limit shape $\mathcal{B}$. If $\nu \in \mathcal{M}_p$, the shape has a flat edge between the points M_p and N_p. The limit shape is differentiable at M_p and N_p. Outside the percolation cone, the limit shape is unknown.

such that $\partial\mathcal{B}$ is infinitely differentiable at M_p and N_p and (b) $\nu \in \mathcal{M}_p$ (and also ν with continuous F) such that for a given $\epsilon > 0$, the extreme points of $\mathcal{B}$ are ϵ-dense in $\partial\mathcal{B}$.

The flat part of the percolation cone ends at M_p and N_p; however, that does not exclude the limit shape from having further flat spots. This is not expected though.

QUESTION 2.5.3. Show that for any measure $\nu \in \mathcal{M}_p$ the boundary of the limit shape is not flat outside the percolation cone.

A similar, but perhaps more ambitious question is to show:

QUESTION 2.5.4. Show that if the limit shape of a measure ν has a flat piece then $\nu \in \mathcal{M}_p$, for $p > p_c$ and the flat piece is delimited by the percolation cone.

More approachable may be the following open question:

QUESTION 2.5.5. Show that for any measure ν, in direction e_1 the boundary of the limit shape does not contain any segment parallel to the e_2-axis.

A solution to 2.5.5 immediately implies a solution to Question 2.5.2. One may reasonably guess that to produce a nontrivial ν satisfying the conditions of

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Question 2.5.5, it suffices to consider small perturbations of the trivial measure $\nu = \delta_1$ (every edge has nonrandom passage time equal to 1) as the limit shape in this case is just the ℓ^1 ball, a diamond. This observation was explored by Basdevant and coauthors by considering FPP in the highly supercritical bond percolation cluster. Precisely, let τ_e be equal to 1 with probability $p \in (0,1)$ and infinity with probability $1-p$.

THEOREM 2.29 (Basdevant et al. [**24**]). *For all $0 \leq \lambda \leq 1$, on the event that the origin is connected to infinity, a.s. we have*

$$\lim_{n\to\infty} \frac{T(0, n(1, \lambda(1-p))}{n} = 1 + (1-p)\frac{1+\lambda^2}{2} + O\big((1-p)^2\big).$$

The theorem above roughly says that when p is close to 1, the four corners of the ℓ^1 ball are replaced by curves that resemble parabolas.

2.6. The subadditive ergodic theorem revisited

Recall that the main tool to prove the existence of the time constant (Theorem 2.1) was the subadditive ergodic theorem (Theorem 2.2). The fact that Theorem 2.2 requires few assumptions makes it very powerful and widens its scope far beyond FPP. This strength also comes with two main drawbacks. First, it only provides the existence of the limit. The characterization as $\inf_{n\geq 1} \mathbb{E}X_{0,n}/n$ is not always helpful to obtain further information on the time constant. Second, it gives no information about, for example, rate of convergence, fluctuations, concentration, or large deviations for ergodic processes under the same assumptions.

The goal of this section is to dissect a major tool in Kingman's original proof of the subadditive ergodic theorem. Kingman's proof provides a nontrivial decomposition of the subadditive process. One term of this decomposition has the same properties as the Busemann function in FPP, an object that we will study in detail in Chapter 5. Later on, we will pursue this direction and provide extra assumptions that allow one to derive further information about the subadditive process $(X_{0,n})$. In this chapter, we stick with the first task.

An array $X = (X_{m,n})$ satisfying the assumptions of Theorem 2.2 but also $X_{0,n} = X_{0,m} + X_{m,n}$ is called an additive ergodic sequence. The proof of Kingman's theorem depends on the following decomposition:

THEOREM 2.30. *If $X_{m,n}$ is a subadditive ergodic sequence then there exist arrays $(Y_{m,n})$ and $(Z_{m,n})$ such that*

(*a*) *$Y_{m,n}$ is an additive ergodic sequence with $\mathbb{E}Y_{0,1} = \mu = \inf_{n\geq 1} \mathbb{E}X_{0,n}/n$.*
(*b*) *$Z_{m,n}$ is a nonnegative subadditive ergodic sequence with time constant equal to 0.*
(*c*) *$X_{m,n} = Y_{m,n} + Z_{m,n}$.*

The decomposition above is not necessarily unique, as the following example illustrates. Let $(y_k)_{k\geq 1}$ be a sequence of independent standard Gaussians and set $S_{m,n} = y_{m+1} + \ldots + y_n$. Then S is additive, with time constant 0. If we put

$$X_{m,n} = \max(S_{m,n}, 0),$$

we see that X is subadditive, with $\mathbb{E}X_{0,n} = (n/2\pi)^{1/2}$, so $\mu = 0$. Two decompositions of X are given by $Y = 0$, $Z = X$ and $Y = S$, $Z = \max(-S, 0)$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

To see some implications of the above decomposition, suppose that the following limit exists a.s.:

$$B_{m,n} := \lim_{N\to\infty} (X_{m,N} - X_{n,N}) \tag{2.18}$$

and satisfies $\mathbb{E}|B_{0,n}| \le Cn$ for some positive constant C. We claim that $B_{m,n}$ is an additive process in the decomposition above. To see this, note that

$$\begin{aligned} B_{0,n} = \lim_{N\to\infty} (X_{0,N} - X_{n,N}) &= \lim_{N\to\infty} (X_{0,N} - X_{m,N} + X_{m,N} - X_{n,N}) \\ &= B_{0,m} + B_{m,n} \end{aligned}$$

and (b) and (c) from the definition in Theorem 2.2 follow from the stationarity of the process $(X_{m,n})$. For instance,

$$\begin{aligned} B_{m,n} = \lim_{N\to\infty} (X_{m,N} - X_{n,N}) &= \lim_{N\to\infty} (X_{m,N-1} - X_{n,N-1}) \\ &\overset{d}{=} \lim_{N\to\infty} (X_{m+1,N} - X_{n+1,N}) = B_{m+1,n+1}, \end{aligned}$$

where we used the fact that if $X_N \overset{d}{=} Y_N$ and X_N, Y_N converge in distribution to X and Y respectively then $X \overset{d}{=} Y$. The moment bounds follow from the assumption $\mathbb{E}|B_{0,n}| \le Cn$.

Now set

$$Z_{m,n} = X_{m,n} - B_{m,n}.$$

For any integer $N > n$, subadditivity gives $X_{m,n} \ge X_{m,N} - X_{n,N}$ a.s.. Thus, we also have $Z_{m,n} \ge 0$ a.s.. As B is additive, $Z_{m,n}$ is subadditive, and thus it satisfies (b).

The importance of the limit in (2.18) will become clear in Chapter 5. At this point, it would be interesting to determine when we can actually use (2.18).

QUESTION 2.6.1. Find conditions that guarantee the existence of the limit (2.18).

The way that Kingman avoided the problem of existence of the limit in (2.18) was to construct weak averaged limits. It will be beneficial to explain his idea here. Let Ω be the space of all subadditive sequences. A subadditive ergodic process is a random element on Ω, inducing a probability measure $\mathbb{P}$ on this space. For any $x \in \Omega$ we define the shift θx as the element in Ω given by

$$\theta x(m,n) = x(m+1, n+1).$$

Now let f be an element of $L^1(\Omega, \mathbb{P})$ and define $T : L^1(\Omega, \mathbb{P}) \to L^1(\Omega, \mathbb{P})$ as the bounded linear operator given by

$$(Tf)(x) = f(\theta x).$$

Kingman's idea was to construct a function f in $L^1(\Omega, \mathbb{P})$ such that

$$f + Tf + \ldots + T^{n-1} f \le X_{0,n}$$

and $\mathbb{E}f = \mu$. Once this function is constructed, the reader can easily check that the representation follows by taking

$$Y_{m,n} = T^m f + T^{m+1} f + \ldots + T^{n-1} f \tag{2.19}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

and $Z_{m,n} = X_{m,n} - Y_{m,n}$. Note that both sides in (2.19) are random variables in Ω. To construct such f he considered the process

$$f_k = \frac{1}{k}\sum_{i=1}^{k}(X_{0,i} - X_{1,i}) \tag{2.20}$$

and for each $n \geq 1$ its iterates

$$f_k + Tf_k + \ldots + T^{n-1}f_k = \frac{1}{k}\sum_{i=1}^{k+n-1}(X_{a,i} - X_{b,i}) \tag{2.21}$$

where $a = \max(i-k, 0)$ and $b = \min(i, n)$. The reader can now see the connection: as k goes to infinity, (2.20) and (2.21) play the role of $B_{0,1}$ and $B_{0,n}$, respectively from (2.18). The existence of the limit in k turns out to follow from the Bourbaki-Alaoglu theorem. We will come back to this point in Chapter 5. When the limit in (2.18) exists, we can think of it as a "generalized Busemann function" for the subadditive process X.

2.7. Gromov-Hausdorff convergence

FPP is a model of a random (pseudo)metric space. There is a modern way of defining convergence of a sequence of metric spaces, using the Gromov-Hausdorff distance on the space of metric spaces. We follow this route in this section and rephrase the shape theorem in this context. The reader is invited to check [**40**, **43**] for a detailed explanation and historical motivation of the topics we touch on. Although this approach brings a different perspective, these methods have not yet provided significant new progress in FPP on $\mathbb{Z}^d$. However, they were successfully used to extend the shape theorem to FPP in Cayley graphs with polynomial growth, where the subadditive ergodic theorem does not immediately apply [**27**, **167**]. We will need a few definitions before we start, and most of the following is taken from [**40**].

Definition 2.31. A subset S of a metric space X is said to be ϵ-*dense* if every point of X lies in the ϵ-neighborhood of S.

Definition 2.32. An ϵ-relation between two (pseudo)metric spaces X_1 and X_2 is a subset $R \subset X_1 \times X_2$ such that

(1) For $i = 1, 2$, the projection of R to X_i is ϵ-dense.
(2) If $(x_1, x_2), (x_1', x_2') \in R$ then $|d_{X_1}(x_1, x_1') - d_{X_2}(x_2, x_2')| < \epsilon$.

If there exists an ϵ-relation between the metric spaces X_1 and X_2, we say that X_1 and X_2 are ϵ-related and use the notation $X_1 \sim_\epsilon X_2$. When the projection of an ϵ-relation is surjective in both of its coordinates, we say that the relation is surjective and we write $X_1 \simeq_\epsilon X_2$. It is an exercise to show that if $X_1 \sim_\epsilon X_2$ then $X_1 \simeq_{3\epsilon} X_2$.

The Gromov-Hausdorff distance between X_1 and X_2 is defined as:

$$D_H(X_1, X_2) := \inf\{\epsilon > 0 \ : \ X_1 \simeq_\epsilon X_2\}.$$

If there is no ϵ such that $X_1 \simeq_\epsilon X_2$, then $D_H(X_1, X_2)$ defined to be infinity. One can show that D_H satisfies the triangle inequality on the set of metric spaces and thus it is a pseudometric which may take the value infinity.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

DEFINITION 2.33. We say that a sequence of (pseudo)metric spaces (X_n) converges to X in the Gromov-Hausdorff metric, and write $X_n \to X$ if $D_H(X_n, X) \to 0$ as $n \to \infty$.

Gromov-Hausdorff convergence works well in contexts where one deals with sequences of compact metric spaces, but not when applied to non-compact spaces. One disadvantage is that the distance between a compact space and an unbounded set is always infinity. Since the spaces that we care about are not compact, we will need the following alternative definition of convergence.

DEFINITION 2.34. A pointed space (X, d, x) is a pair of a metric space (X, d) and a point $x \in X$. The point x is called the basepoint of the pointed space (X, d, x).

DEFINITION 2.35. A sequence of pointed spaces $((X_n, d_n, x_n))$ converges to a pointed space (X, d, x) if for every $r > 0$ the sequence of closed balls $(B(x_n, r))$ (with induced metrics) converges to the closed ball $B(x, r)$ in the Gromov-Hausdorff metric.

One of the features of pointed Gromov-Hausdorff convergence is that it preserves several properties of the sequence of metric spaces in the limit. Returning to FPP, for each outcome ω, we define a sequence of pseudometric spaces:

$$\Big(X_n, d_n(x, y)\Big) = \left(\frac{1}{n}\mathbb{Z}^d, \frac{T(nx, ny)}{n}\right).$$

The pseudometric space X_n is just the original lattice rescaled by n with the normalized pseudometric; that is, $d_n(0, x) = T(0, nx)/n$. The origin 0 is a point of X_n for all $n \geq 1$. Now recall the limit shape construction from Section 2.3. Given an edge-weight distribution on the edges of $\mathbb{Z}^d$, there exists a norm μ on $\mathbb{R}^d$ such that the unit ball in that norm is the limit shape of the FPP model. The pair $(\mathbb{R}^d, \mu)$ is a normed vector space with distance $d(x, y) = \mu(y - x)$ for $x, y \in \mathbb{R}^d$. We will assume that the passage times have finite exponential moments. This assumption is to make sure the metric satisfies a concentration bound given by Lemma 2.37 below.

The shape theorem translates to the following statement.

THEOREM 2.36. *Assume that $F(0) < p_c$ and $\mathbb{E}e^{\alpha\tau_e} < \infty$ for some $\alpha > 0$. a.s., the sequence $((X_n, d_n, 0))$ converges in the pointed Gromov-Hausdorff sense to $(\mathbb{R}^d, \mu, 0)$.*

PROOF. Fix $r > 0$ rational. We first show that a.s. the balls $B_n := B_n(0, r) \subset X_n$ converge in the Gromov-Hausdorff sense to the ball

$$B := B(0, r) = \{x \in \mathbb{R}^d \ : \ \mu(x) \leq r\}.$$

For this, it suffices to show that for any fixed $\epsilon > 0$ there a.s. exists $n_0 \in \mathbb{N}$ so that, for any $n \geq n_0$ there is an ϵ-relation between B_n and B.

We construct such a relation as follows. Given $0 < \epsilon' < \epsilon$, use Theorem 2.17 to choose $n_0 = n_0(\omega)$ so that for $n \geq n_0$

$$\mathbb{P}\Big((1 - \epsilon')B \subseteq \bar{B}_n \subseteq (1 + \epsilon')B \text{ for all } n \geq n_0\Big) = 1,$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

where $\bar{B}_n = B_n + [-1/(2n), 1/(2n)]^d$. The set R is defined as the union of two sets, $R = R_1 \cup R_2$, where

$$R_1 := \Big\{ (x_1, x_2) \in B_n \times B : x_2 = x_1/(1+\epsilon') \Big\},$$

$$R_2 := \Big\{ (x_1, x_2) \in B_n \times B : (1-\epsilon')x_2 \in x_1 + [-1/(2n), 1/(2n)]^d \Big\}.$$

Note that R is surjective. Indeed, since $B_n \subseteq (1+\epsilon')B$, every element of B_n is related to some element of B through R_1 while $(1-\epsilon')B \subseteq \bar{B}_n$ implies that every element of B is related to at least one element of B_n through R_2. Now take (x_1, x_2) and (x_1', x_2') in R. We have

$$\begin{aligned} |d_n(x_1, x_1') - d(x_2, x_2')| &= |T(nx_1, nx_1')/n - \mu(x_2 - x_2')| \\ &\le |T(nx_1, nx_1')/n - \mu(x_1 - x_1')| \\ &\quad + |\mu(x_1 - x_1') - \mu(x_2 - x_2')|. \end{aligned} \tag{2.22}$$

Let us first look at the second term in the right side of (2.22). If (x_1, x_2) and (x_1', x_2') are both in R_1 we have

$$\mu(x_1 - x_1') - \mu(x_2 - x_2') = (1+\epsilon')\mu(x_2 - x_2') - \mu(x_2 - x_2') = \mu(x_2 - x_2')\epsilon' \le 2r\epsilon'. \tag{2.23}$$

If both pairs of points are in R_2, then $x_2 = (1-\epsilon')^{-1}(x_1 + z_n)$, $x_2' = (1-\epsilon')^{-1}(x_1' + z_n')$ for some z_n, z_n' in $[-1/(2n), 1/(2n)]^d$. We thus obtain

$$\mu(x_2 - x_2') = (1-\epsilon')^{-1}\mu(x_1 - x_1' + z_n - z_n'). \tag{2.24}$$

Since μ is a norm, we can find $n_1 \in \mathbb{N}$ so that for any $n \ge \max(n_0, n_1)$ and any $|w| \le \sqrt{d}/n$,

$$|\mu(z + w) - \mu(z)| \le \epsilon'.$$

As $|z_n - z_n'| \le \sqrt{d}/n$, we have by (2.24) and the triangle inequality

$$\begin{aligned} |\mu(x_1 - x_1') - \mu(x_2 - x_2')| &\le ((1-\epsilon')^{-1} - 1)\mu(x_1 - x_1') + \epsilon'(1-\epsilon')^{-1} \\ &\le 2r((1-\epsilon')^{-1} - 1) + \epsilon'(1-\epsilon')^{-1}. \end{aligned} \tag{2.25}$$

If $(x_1, x_2) \in R_1$ and $(x_1', x_2') \in R_2$, then similarly to (2.23) and (2.25) we obtain

$$|\mu(x_1 - x_1') - \mu(x_2 - x_2')| \le C\epsilon', \tag{2.26}$$

for $C < \infty$. Thus a combination of (2.23), (2.25) and (2.26) tells us that if we choose ϵ' small enough

$$|\mu(x_1 - x_1') - \mu(x_2 - x_2')| < \epsilon/2 \quad \text{for all } n \ge \max\{n_0, n_1\}. \tag{2.27}$$

The first term of (2.22) is controlled by the following concentration bound.

LEMMA 2.37. *Given $\epsilon > 0$ and $M > 0$ there exists $C_1 > 0$ such that for any $n \ge 1$*

$$\mathbb{P}\Big(\exists\, x, y \in [-Mn, Mn]^d \text{ with } |T(x,y)/n - \mu(y - x)| \ge \epsilon/2 \Big) \le \exp(-n^{C_1}).$$

PROOF. This inequality follows from a union bound and any one of many concentration or large deviation inequalities, like Theorem 3.10. $\square$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Combining (2.22), (2.27) and Lemma 2.37, we see that

$$\mathbb{P}\Big(R \text{ is an } \epsilon\text{-relation between } B \text{ and } B_n\Big) \geq 1 - \exp(-n^{C_1}),$$

and thus by taking a countable sequence of $\epsilon_n \to 0$ and using the Borel-Cantelli lemma, we obtain the desired result for each $r \in \mathbb{Q}$. However, if R is a ϵ-relation between $B_n(0, r)$ and $B(0, r)$ then (the restriction of) R is also an ϵ-relation between $B_n(0, r')$ and $B(0, r')$ for any $r' < r$. This last observation suffices to end the proof of Theorem 2.36. □

2.8. Strict convexity of the limit shape

In this section, we explore in more detail the conjecture that, under mild assumptions on F, the limit shape is strictly convex. We also introduce the definition of uniform positive curvature, a concept related to strict convexity. In Chapters 3 and 4, we will discuss important results of Newman where this unproven property of uniformly positive curvature will play a major role. Strict convexity also plays an important part in certain questions regarding the evolution of multi-type stochastic competition models discussed in Chapter 6.

We briefly recall some basic properties of convex sets and convex functions (a classic reference is the book of Rockafellar [**154**]). Let μ be the FPP norm defined in (2.5) and recall that the limit shape $\mathcal{B}$ is the level set

$$\{x \in \mathbb{R}^d : \mu(x) \leq 1\}.$$

Recall that a hyperplane H is a supporting hyperplane for a convex set $\mathcal{B}$ at $x \in \mathcal{B}$ if H contains x and $\mathcal{B}$ intersects at most one component of H^c. By the items discussed after (2.5), μ is a real valued convex function and the following basic properties hold.

(1) $\mathcal{B}$ is a convex subset of $\mathbb{R}^d$.
(2) The set of points of differentiability of μ is dense in $\mathbb{R}^d$, and its complement has zero Lebesgue measure.
(3) The boundary $\partial\mathcal{B}$ of the limit shape is differentiable at all but a countable collection of points of the form $x/\mu(x)$. In other words, there is exactly one supporting hyperplane for $\mathcal{B}$ at such $x/\mu(x)$.

Recall that we call a subset $\mathcal{B}$ of $\mathbb{R}^d$ strictly convex if every line segment connecting any two points of $\mathcal{B}$ is entirely contained, except for its endpoints, in the interior of $\mathcal{B}$. An extreme point x of $\mathcal{B}$ is one such that if $y, z \in \mathcal{B}$ with $x = \lambda y + (1 - \lambda)z$ for some $\lambda \in [0, 1]$, we must have $\lambda = 0$ or 1. An exposed point x of $\mathcal{B}$ is one such that there is a supporting hyperplane for $\mathcal{B}$ that intersects $\mathcal{B}$ only at x.

Let u be a unit vector of $\mathbb{R}^d$ and let H_0 be a hyperplane through the origin such that $u + H_0$ is a supporting hyperplane for $\mu(u)\mathcal{B}$ at u. We introduce a curvature exponent that captures the nature of the boundary of $\mathcal{B}$ in direction u as follows.

DEFINITION 2.38 (Curvature Exponent). Assume that $\partial\mathcal{B}$ is differentiable. The curvature exponent $\kappa(u)$ in the direction u is a real number such that there exist positive constants c, C and ε such that for any $z \in H_0$ with $|z| < \varepsilon$, one has

$$c|z|^{\kappa(u)} \leq \mu(u + z) - \mu(u) \leq C|z|^{\kappa(u)}. \tag{2.28}$$

In other words, if $\kappa(u)$ exists then locally, near the point u, the function $z \in H_0 \mapsto \mu(u + z) - \mu(u)$ is close to $|z|^{\kappa(u)}$; that is, the boundary of the limit shape,

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

viewed from a supporting hyperplane, looks like the graph of the function $x \mapsto |x|^{\kappa(u)}$. Figure 5 illustrates the case $\kappa(u) = 2$.

DEFINITION 2.39 (Uniformly curved). We say that $\mathcal{B}$ is uniformly curved if $\partial\mathcal{B}$ is differentiable and for every unit vector $u \in \mathbb{R}^d$, the left inequality of (2.28) holds with ϵ and c that are uniform in u. A special case is $\kappa(u) = 2$ for all u.

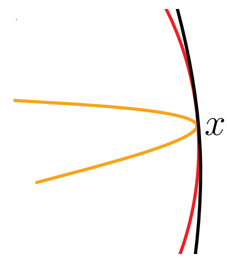


FIGURE 5. Representation of the limit shape (middle curve) in direction x with curvature exponent $\kappa(x) = 2$. The limit shape boundary is delimited by two tangent parabolas, one inside, and the other outside.

It is an exercise to show that if $\mathcal{B}$ is uniformly curved then $\mathcal{B}$ is strictly convex. In the case that $\partial\mathcal{B}$ is not necessarily differentiable, Newman [**147**] gave a general definition of uniform curvature: there exists $C > 0$ such that for all $z_1, z_2 \in \partial\mathcal{B}$ and $z = (1-\lambda)z_1 + \lambda z_2$ with $\lambda \in [0, 1]$,

$$1 - \mu(z) \geq C \min\{\mu(z - z_1), \mu(z - z_2)\}^2. \tag{2.29}$$

Either of these definitions is suitable for the results in this book.

A Euclidean ball is uniformly curved with $\kappa(u) = 2$ for every u. The ℓ^1 ball is not uniformly curved because if u is not in the direction of one of its corners, the lower bound in (2.28) does not hold for u. Unfortunately, uniform curvature has not been proved for the limit shape of any FPP model with i.i.d. passage times. Any advance in the direction of the following question would be a major contribution.

QUESTION 2.8.1. Show that for continuously distributed passage times, the limit shape is uniformly curved.

As we saw, uniform curvature implies strict convexity. A conditional proof of strict convexity was also obtained by Lalley [**137**]. The two hypotheses of Lalley's result, however, seem to be out of reach at this moment. Hypothesis H2 might not be valid, as for instance the Tracy-Widom distribution does not have mean zero. We describe them now.

For u a fixed nonzero vector in $\mathbb{R}^d$ let L_u be the ray through u emanating from the origin.

(H1). For any convex cone $\mathcal{A}$ of $\mathbb{R}^d$ containing the vector u in its interior, and for each $\delta > 0$ there exists $R = R(\delta, \mathcal{A}) < \infty$ such that the following is true: For each point $v \in \mathbb{Z}^d \cap \mathcal{A}$ at distance ≤ 2 from the line L_u, the probability that the time-minimizing path from the origin to v is contained in $\mathcal{A} \cup \{x : \|x\| \leq R\}$ is at least $1 - \delta$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

The second assumption requires a fluctuation theorem for the normalized passage times.

(H2). There exists a mean-zero probability distribution G_u on the real line and a scalar sequence $a(n) \to \infty$ such that as $n \to \infty$

$$\frac{T(0,nu) - n\mu(u)}{a(n)} \Rightarrow G_u.$$

THEOREM 2.40 (Lalley [**137**], Theorem 1). *Let u and v be linearly independent vectors in $\mathbb{R}^d$ and assume hypothesis **(H1)** and **(H2)** for both u and v. Then for each $\lambda \in (0,1)$,*

$$\mu(\lambda u + (1-\lambda)v) < \lambda\mu(u) + (1-\lambda)\mu(v).$$

2.9. Simulations

Although we are celebrating the fiftieth anniversary of the model, simulation studies on FPP were somewhat limited until very recently. Initial work is due to Richardson [**155**] in 1973, where the model with exponentially distributed weights was analyzed. In [**155**] the limit shape $\mathcal{B}$ seemed to be curved, with a shape resembling a circle. As one could imagine, these simulations were restricted due to limitations in computer power. Further investigation (also in the Eden model) came in 1986 in the work of Zabolitzky and Stauffer [**181**]. In particular, the numbers obtained in [**181**] indicate the predicted fluctuation exponents for $\xi = 2/3$ and $\chi = 1/3$ by theoretical physicists [**120, 121, 132**] (see next two chapters for the study of these exponents).

A major contribution was done recently in the work of Alm and Deifjen [**12**] for FPP in two dimensions. Running 19 years of CPU time, in a cluster of 28 Linux machines, they investigate the value of the time constant and the limit shape for several continuous distributions. Their results are consistent with the well-known conjectures for the model, showing for example strict convexity of the limit shape, with a limit shape different from a circle in all cases. The exponents for the standard deviation of hitting times and for the fluctuations of hitting points on lines also matched the predicted values 1/3 and 2/3, respectively. See Figures 6 and 7 for illustrations of simulation results for limit shapes.

The paper of Alm and Deifjen also brought new findings to the table. It seems that the time constant depends primarily on the mean of the minimum edge weight adjacent to the origin, that is,

$$\mathbb{E}_4\tau_e := \mathbb{E}\min\{t_1, t_2, t_3, t_4\},$$

where the $t_i's$ are independent copies of τ_e, at least for continuous distributions which are not too concentrated. They reported that the time constants along the axis and along the diagonal for all simulated distributions have an almost perfect linear relation with $\mathbb{E}_4\tau_e$. They also suggest that if $F(0) = 0$ then $\mu \geq \mathbb{E}_4\tau_e$.

QUESTION 2.9.1 (Alm-Deijfen). Assume $F(0) = 0$. Show that $\mu \geq \mathbb{E}_4\tau_e$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

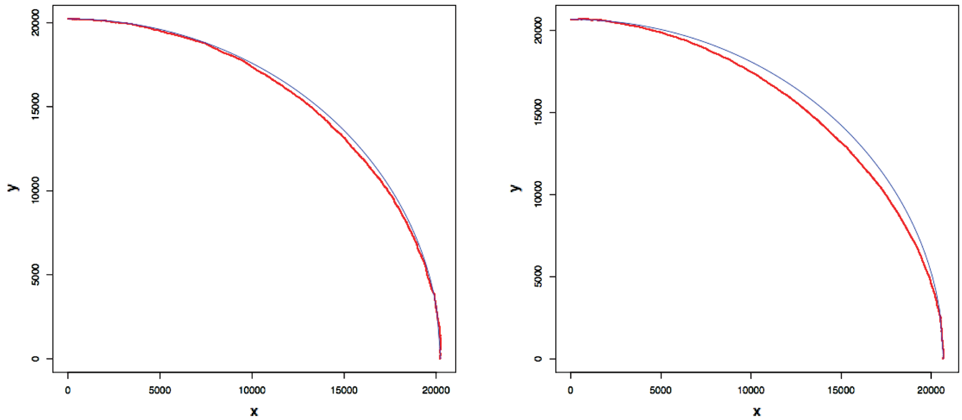


FIGURE 6. Simulation of the ball $B(t)$ for passage times distributed according to a Gamma distribution $\Gamma(k, k)$ with parameters $k = 1$ (left) and $k = 2$ (right) with $t = 20000$. The outer curve is a circle of the same e_1-radius as $B(t)$. [These figures originally appeared in: "First Package Percolation on $\mathbb{Z}^2$: A Simulation Study" (S.-E. Alm and M. Deijfen), Journal of Statistical Physics **161** (2015), no. 3, pp. 657–698. ©2015 Springer Science + Business Media New York. Reprinted with permission of Springer.]

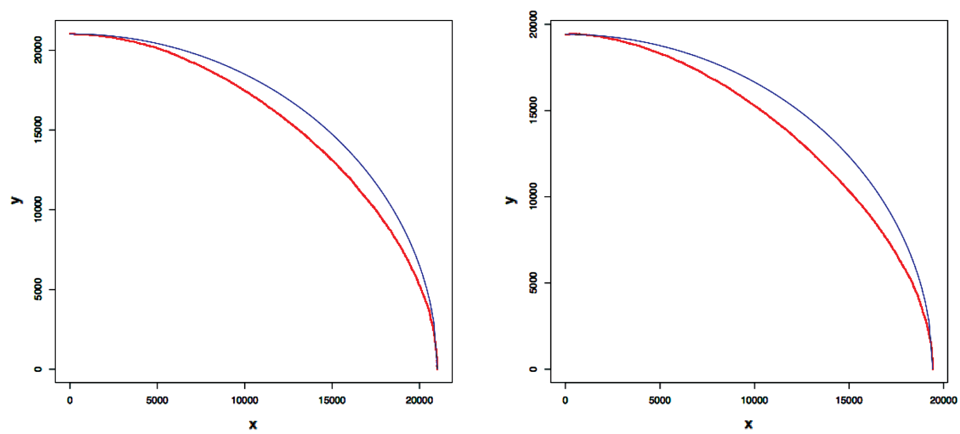


FIGURE 7. Simulation of the ball $B(t)$ for passage times distributed according to a Gamma distribution $\Gamma(k, k)$ with parameters $k = 3$ (left) and $k = 4$ (right) with $t = 20000$. The outer curve is a circle of the same e_1-radius as $B(t)$. [These figures originally appeared in: "First Package Percolation on $\mathbb{Z}^2$: A Simulation Study" (S.-E. Alm and M. Deijfen), Journal of Statistical Physics **161** (2015), no. 3, pp. 657–698. ©2015 Springer Science + Business Media New York. Reprinted with permission of Springer.]

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/03

CHAPTER 3

Fluctuations and concentration bounds

The passage time between 0 and a vertex $x \in \mathbb{Z}^d$ can be approximated (a.s.) as

$$T(0,x) = \mu(x) + o(\|x\|_1),$$

due to the shape theorem. Quantifying the error term is the main subject of this chapter. It has been traditionally analyzed in two pieces:

$$o(\|x\|_1) = \underbrace{T(0,x) - \mathbb{E}T(0,x)}_{random\ fluctuations} + \underbrace{\mathbb{E}T(0,x) - \mu(x)}_{nonrandom\ fluctuations} .$$

The reason this splitting is useful is that the first term can typically be treated using techniques from concentration of measure, whereas the second term is analyzed using, in part, bounds on the first.

3.1. Variance bounds

The most basic control on the random fluctuation term is a variance bound. It has been predicted in the physics literature by simulation [**178**, **181**] and some scaling theory [**121**,**132**] (see also in Kesten [**125**, p. 298]) that there is a dimension-dependent exponent $\chi = \chi(d)$ such that

$$\mathrm{Var}\ T(0,x) \sim |x|^{2\chi}. \tag{3.1}$$

This exponent is expected to be universal; it should not depend on the underlying edge-weight environment, as long as F satisfies some mild moment conditions and the limit shape has no flat edges. The meaning of "$\sim$" has not been made clear. For instance, it could be that as $|x| \to \infty$, the ratio of both sides converges to a constant, is bounded away from 0 and ∞, or even that the variance has the form $|x|^{2\chi+o(1)}$, as in the current estimates for Bernoulli percolation exponents. Regardless, the following dependence on d is predicted for χ:

d	χ
1	1/2
2	1/3 (conjectured)
$d \geq 3$	unknown

TABLE 1. Values of the variance exponent χ as the dimension d varies.

For $d = 1$, the passage time $T(0,x)$ is just a sum of $\|x\|_1$ i.i.d. random variables, and so one has $\chi(1) = 1/2$ under any reasonable definition of "$\sim$." For $d \geq 2$, the infimum in the definition of $T(0,x)$ is predicted to stabilize $T(0,x)$, producing subdiffusive fluctuations with $\chi < 1/2$. It is clear that χ should decrease with dimension, but there is no agreement on whether $\chi(d) = 0$ for all d at least equal

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

to some d_c, and some even debate whether $\chi \to 0$ (see [**149**] and the references therein).

The history of rigorous variance bounds for T begins with Kesten's work [**123**, Theorem 5.16], showing that Var $T(0, ne_1) \le C(n/\log^{1/p} n)^2$ for $p = 9d + 3$. Although this bound is only logarithmically better than a trivial bound (say in the case of bounded weights), the proof is far from trivial. In 1993, Kesten introduced the "method of bounded differences" to FPP, and with this he was able to prove the best current bounds for χ:

$$0 \le \chi(d) \le 1/2 \text{ for all } d \ge 1.$$

We will begin by giving a sketch of his argument using the Efron-Stein inequality.

THEOREM 3.1 (Kesten [**125**]). *Assume $\mathbb{E}\tau_e^2 < \infty$, $F(0) < p_c$, and that the distribution of τ_e is not concentrated at one point. There exist $C_1, C_2 > 0$ such that for all non-zero $x \in \mathbb{Z}^d$,*

$$C_1 \le \text{Var } T(0, x) \le C_2 \|x\|_1.$$

The proof will use the following inequality for functions of independent random variables from [**83, 163**]. We use the notation $x_+ = \max\{0, x\}$.

LEMMA 3.2 (Efron-Stein inequality). *Let $X_1, \ldots, X_n$ be independent and let X_i' be an independent copy of X_i, for $i = 1, \ldots, n$. If f is an L^2 function of $(X_1, \ldots, X_n)$ then*

$$\text{Var } f \le \frac{1}{2} \sum_{i=1}^n \mathbb{E}[(Z_i - Z)^2] = \sum_{i=1}^n \mathbb{E}[(Z_i - Z)_+]^2,$$

where $Z = f(X_1, \ldots, X_n)$ and

$$Z_i = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n).$$

The following proof of the Efron-Stein inequality goes through with minor modifications when f is an L^2 function of infinitely many independent variables $X_1, X_2, \ldots$.

PROOF. The equality follows by symmetry. For the inequality, we follow [**39**, Theorem 3.1]. Write $\mathbb{E}_i$ as the conditional expectation operator conditioned on $(X_j : j \le i)$ (with $\mathbb{E}_0 = \mathbb{E}$). Setting

$$\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z,$$

we have the decomposition

$$Z - \mathbb{E}Z = \sum_{i=1}^n \Delta_i$$

and since $\mathbb{E}\Delta_i \Delta_j = 0$ for $i \ne j$, one has

$$\text{Var } Z = \sum_{i=1}^n \mathbb{E}\Delta_i^2. \tag{3.2}$$

Letting $\mathbb{E}^{(i)}$ be the integral operator over the i-th coordinate, one can use independence to rewrite $\Delta_i = \mathbb{E}_i \left[Z - \mathbb{E}^{(i)} Z\right]$. By the conditional Jensen inequality,

$$\Delta_i^2 \le \mathbb{E}_i \left[\left(Z - \mathbb{E}^{(i)} Z\right)^2\right], \tag{3.3}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

and combining this with (3.2), we obtain

$$\text{Var } Z \leq \sum_{i=1}^{n} \mathbb{E}\left[\left(Z - \mathbb{E}^{(i)} Z\right)^2\right].$$

If $\text{Var}^{(i)}$ is the variance operator corresponding to $\mathbb{E}^{(i)}$, then the right side can be written as $\sum_{i=1}^{n} \mathbb{E}\text{Var}^{(i)}\ Z$. To finish, we use the fact that for i.i.d. random variables X and Y, one has $\text{Var } X = \frac{1}{2}\mathbb{E}(X-Y)^2$. To apply this to our setting, note that Z and Z_i are conditionally i.i.d. given $(X_j : j \neq i)$, and so

$$\text{Var } Z \leq \sum_{i=1}^{n} \mathbb{E}\text{Var}^{(i)}\ Z = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}(Z - Z_i)^2.$$

□

Proof of Theorem 3.1. We now apply the Efron-Stein inequality to the passage time $f = T(0,x)$, noting that the condition $\mathbb{E}\tau_e^2 < \infty$ implies that T has finite second moment. Then

$$\text{Var } T(0,x) \leq \sum_{i=1}^{\infty} \mathbb{E}[(T_i(0,x) - T(0,x))_+]^2,$$

where $e(1), e(2), \ldots$ is any enumeration of the edges and T_i is the passage time in the edge-weight configuration (τ_e) but with the weight $\tau_{e(i)}$ replaced by an independent copy $\tau'_{e(i)}$. Note that $T_i(0,x) > T(0,x)$ only when both $e(i)$ is in $\underline{\text{Geo}}(0,x)$, the intersection of all geodesics from 0 to x in the original edge-weight configuration (τ_e), and $\tau'_{e(i)} > \tau_{e(i)}$. Furthermore, in this case, $T_i(0,x) - T(0,x) \leq \tau'_{e(i)}$. Therefore we obtain the bound

$$\text{Var } T(0,x) \leq \sum_{i=1}^{\infty} \mathbb{E}(\tau'_{e(i)})^2 \mathbf{1}_{\{e(i) \in \underline{\text{Geo}}(0,x)\}}. \tag{3.4}$$

By independence, this equals

$$\mathbb{E}\tau_e^2 \mathbb{E}|\underline{\text{Geo}}(0,x)|.$$

Therefore we can conclude the upper bound for $\text{Var } T(0,x)$ with the following lemma, noting that $\mathbb{E}T(0,x) \leq \mathbb{E}\tau_e \|x\|_1$.

Lemma 3.3. *Assume $F(0) < p_c$. There exists a constant C such that for all $x \in \mathbb{Z}^d$,*

$$\mathbb{E}|\underline{\text{Geo}}(0,x)| \leq C\mathbb{E}T(0,x).$$

Proof. For the proof, we appeal to a more general result, proved later as item 1 of Proposition 4.7. Defining $\text{Geo}(x,y)$ as the self-avoiding geodesic from x to y with the maximal number of edges, we have $\mathbb{E}|\text{Geo}(0,x)| \leq C\mathbb{E}T(0,x)$ for some constant C and all nonzero x. □

The lower bound in Kesten's theorem is easier. Setting Σ to be the sigma-algebra generated by the $2d$ edge-weights for edges adjacent to 0, one has

$$\text{Var } T(0,x) \geq \mathbb{E}\left(\mathbb{E}[T(0,x) \mid \Sigma] - \mathbb{E}T(0,x)\right)^2 = \text{Var}\Big[\mathbb{E}[T(0,x) \mid \Sigma]\Big].$$

Let $t'_1, \ldots, t'_{2d}$ be independent copies of the edge-weights of edges adjacent to the origin and set $\Sigma' = \sigma(t'_1, \ldots, t'_{2d})$. Let $T'(0,x)$ be the passage time from 0 to x

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

when we replace the edge-weights of edges adjacent to the origin by the t_i's. Then the right side of the display above is equal to

$$\frac{1}{2}\mathbb{E}\Big[\mathbb{E}[T(0,x)|\Sigma]-\mathbb{E}[T'(0,x)|\Sigma']\Big]^2.$$

Now pick $a<b$ such that $\mathbb{P}(\tau_e<a)>0$ and $\mathbb{P}(\tau_e>b)>0$. By considering the events $A=\{\tau_{\{0,y\}}<a$ for all $y\in\mathbb{Z}^d$ with $\|y\|_1=1\}$ and B the same event but with $<a$ replaced by $>b$ and with $\tau_{\{0,y\}}$'s replaced by t_i's, one has the lower bound

$$\frac{1}{2}\mathbb{E}\Big[\big[\mathbb{E}[T(0,x)|\Sigma]-\mathbb{E}[T'(0,x)|\Sigma']\big]\mathbf{1}_{A\cap B}\Big]^2\geq C(b-a)^2>0$$

independently of x. $\square$

We end by restating three open questions discussed at the beginning of the section.

QUESTION 3.1.1. Show that for any $d\geq 2$, under suitable conditions on F, $\chi<1/2$. If $d=2$, show that $\chi=1/3$.

QUESTION 3.1.2. Determine whether or not

$$\lim_{d\to\infty}\chi(d)=0.$$

QUESTION 3.1.3. For suitable d, show that $\chi>0$.

3.2. Logarithmic improvement to variance upper bound

The main tools used in the proof of the upper bound of Theorem 3.1 were (a) $\mathbb{E}|\underline{\mathrm{Geo}}(0,x)|\leq C\|x\|_1$ and (b) $T_i(0,x)-T(0,x)\leq\tau_{e(i)}'$. To improve the variance upper bound we will need to use one more piece of information: for most edges e, the probability that e is in a geodesic from 0 to x is small (in x). Another way to say this is that each edge has small *influence* on the variable $T(0,x)$. This statement requires $d\geq 2$, since in $d=1$, each edge has high influence (there is only one path from 0 to x in that case).

The first proof of sublinear variance for $T(0,x)$ was due to Benjamini-Kalai-Schramm [**26**] in 2003 and applied only to τ_e's that are Bernoulli: there exist a,b with $0<a<b<\infty$ such that τ_e takes values a or b with probability 1/2. This specific distribution was needed to take advantage of Talagrand's influence inequality on the hypercube [**165**]. The second proof was due to Benaïm-Rossignol [**25**] in 2008 and applied to distributions in the nearly-Gamma class: those distributions that satisfy a logarithmic-Sobolev inequality similar to that for the Gamma distribution. Their methods were based on entropy and used an inequality due to Falik-Samorodnitsky [**84**] which replaced Talagrand's inequality (a similar inequality was derived by Rossignol [**156**]).

The most recent proof is due to Damron-Hanson-Sosoe [**66**] in 2014 and applies to all distributions with $2+\log$ moments. Their method follows that of Benaïm-Rossignol, but replaces the representation of an edge-weight as a push-forward of a Gaussian variable with a representation using a Bernoulli encoding. That is, each edge-weight is encoded as an infinite sequence of 0/1-valued random variables, and the Gross two-point inequality [**104**] is used instead of the Gaussian log-Sobolev inequality to provide entropy bounds.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

3.2.1. The BKS proof of sublinear variance. We begin by discussing the original sublinear bound of BKS. Assume that the edge-weights (τ_e) satisfy $\mathbb{P}(\tau_e = a) = 1/2 = \mathbb{P}(\tau_e = b)$ for some a, b with $0 < a < b < \infty$. The main idea is to apply the following inequality to the function $f = T(0, x)$. If $\omega \in \{a, b\}^n$, write $\hat{\omega}_j$ for the element of $\{a, b\}^n$ which differs from ω exactly in the j-th coordinate.

THEOREM 3.4 (Talagrand [**165**]). *There is a constant C such that for all functions $f : \{a, b\}^n \to \mathbb{R}$,*

$$\mathrm{Var}\ f \le C \sum_{i=1}^{n} \frac{\mathbb{E}(\rho_j f)^2}{1 + \log \frac{\sqrt{\mathbb{E}(\rho_j f)^2}}{\mathbb{E}|\rho_j f|}},$$

where $\{a, b\}^n$ is endowed with the uniform measure, and $(\rho_j f)(\omega)$ is $\frac{1}{2}(f(\omega) - f(\hat{\omega}_j))$.

By Jensen's inequality, the denominator in Talagrand's inequality is always at least 1, so the right side is bounded above by $C \sum_{i=1}^{n} \mathbb{E}(\rho_j f)^2$. This sum is seen to be of the same order as the right side of the Efron-Stein inequality (see Lemma 3.2). Therefore Talagrand's inequality is an improvement over the Efron-Stein inequality in situations in which the denominator is large; that is, when $\mathbb{E}(\rho_j f)^2$ is much larger than $(\mathbb{E}|\rho_j f|)^2$. If we consider f's with the property that $\rho_j f$ is of order 1 on an event A_j and 0 otherwise (and this should roughly be true for f equal to the passage time $T(0, x)$), we find the requirement that the event A_j have very small probability. That is, we may expect a logarithmic improvement over the Efron-Stein bound if the influence of the j-th coordinate (the probability that $\rho_j f$ is nonzero) is at most n^{-c} for some positive c and all j.

BKS were not able to show that influences of edge-weights on $T(0, x)$ are uniformly small (they state that it is sufficient to prove that the probability that an edge is in a geodesic from 0 to x is at most $C|x|^{-1/C}$ with the exception of at most $C|x|/\log|x|$ edges, for some constant $C > 0$ [**26**, p. 1973]), and this motivated the "midpoint problem," which we now state.

QUESTION 3.2.1. Show that for continuous F and $d \ge 2$,

$$\lim_{n \to \infty} \mathbb{P}((n/2)e_1 \in \overline{\mathrm{GEO}}(0, ne_1)) = 0.$$

See Remark 5.26 for partial results related to question 3.2.1, and [**5**] for more recent work.

As a result, they used an averaging trick, considering an averaged version of $T(0, x)$ instead, and for this version they could show small influences. We detail a more recent version of this trick in the next section (see (3.12)), using the machinery of Benaïm-Rossignol in a more general setting. For now, let us explain how BKS could use Talagrand's inequality quickly to show sublinear variance for a translation invariant version of FPP on a torus.

Consider the side length n discrete torus, $[0, n]^d \cap \mathbb{Z}^d$, with opposite sides identified. On the edges, we place Bernoulli edge-weights as above, taking values a and b. Let $f = T_n$ be the minimal passage time among all paths which wind around the torus once in the e_1-direction. If we enumerate the edges $e(1), e(2), \ldots$ of the torus and consider f as a function of the edge-weights $(\tau_{e(i)})$, then $\rho_j f < 0$ if and only if both $\tau_{e(j)} = a$ and $e(j)$ is in $\mathcal{P}_n$, the intersection of all minimal such paths. Since symmetry and independence imply $\mathbb{P}(\rho_j f < 0) = \mathbb{P}(\rho_j f \neq 0, \tau_{e(j)} = a) = \frac{1}{2}\mathbb{P}(\rho_j f \neq$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

0), we can estimate using the Cauchy-Schwarz inequality:

$$\begin{aligned}\mathbb{E}|\rho_j f| \leq \sqrt{\mathbb{E}(\rho_j f)^2 \mathbb{P}(\rho_j f \neq 0)} &= \sqrt{2\mathbb{E}(\rho_j f)^2 \mathbb{P}(\tau_{e(j)} = a, e(j) \in \mathcal{P}_n)} \\ &\leq \sqrt{2\mathbb{E}(\rho_j f)^2 \mathbb{P}(e(j) \in \mathcal{P}_n)}.\end{aligned}$$

Therefore

$$\frac{\sqrt{\mathbb{E}(\rho_j f)^2}}{\mathbb{E}|\rho_j f|} \geq \frac{1}{\sqrt{2\mathbb{P}(e(j) \in \mathcal{P}_n)}}.$$

Note that any minimizing path Γ must satisfy $a|\Gamma| \leq T(\Gamma) = T_n \leq bn$, where the right inequality holds by choosing the deterministic path proceeding in the e_1 direction. Therefore $|\mathcal{P}_n| \leq \frac{b}{a} n$ and by translation invariance, if we write E_n for the set of edges in the torus,

$$\mathbb{P}(e(j) \in \mathcal{P}_n) = \frac{1}{|E_n|} \sum_j \mathbb{P}(e(j) \in \mathcal{P}_n) \leq \frac{bn}{a|E_n|} \leq Cn^{-(d-1)}.$$

So by applying Theorem 3.4 and the bound $\mathbb{E}(\rho_j f)^2 \leq (b-a)^2 \mathbb{P}(\rho_j f \neq 0) = 2(b-a)^2 \mathbb{P}(\tau_{e(j)} = a, e(j) \in \mathcal{P}_n)$, we obtain

$$\text{Var } T_n \leq \frac{C}{\log n} \sum_j \mathbb{P}(\tau_{e(j)} = a, e(j) \in \mathcal{P}_n) \leq C \frac{n}{\log n}.$$

3.2.2. Extending the BKS result to general weights. Below we state the sublinear variance bound from [**66**], but we will sketch the proof only in the simplest case (uniform $[0,1]$ weights), following [**25**], and indicating where complications arise in extending the argument.

THEOREM 3.5 (Damron-Hanson-Sosoe [**66**]). *For $d \geq 2$, suppose $F(0) < p_c$ and $\mathbb{E}\tau_e^2 (\log \tau_e)_+ < \infty$. There exists a constant C such that for all $x \in \mathbb{Z}^d$ with $\|x\|_1 > 1$,*

$$\text{Var } T(0,x) \leq C \frac{\|x\|_1}{\log \|x\|_1}.$$

To prove the theorem above, we will use the Falik-Samorodnitsky inequality. Recall that the entropy of a nonnegative random variable X is defined as

$$\text{Ent } X = \mathbb{E}X \log X - \mathbb{E}X \log \mathbb{E}X.$$

Now, let $\nu_n := \nu \times \ldots \times \nu$ be the uniform measure on $[0,1]^n$. In what follows, expectation is with respect to ν_n. For k with $1 \leq k \leq n$, define Σ_k to be the sigma-algebra generated by the first k coordinates in $\mathbb{R}^n$, with Σ_0 equal to the trivial sigma-algebra. Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that $\mathbb{E}f^2 < \infty$. Last, define the martingale difference

$$\Delta_k = \mathbb{E}[f \mid \Sigma_k] - \mathbb{E}[f \mid \Sigma_{k-1}].$$

THEOREM 3.6 (Falik-Samorodnitsky [**84**]). *Let $f : \mathbb{R}^n \to \mathbb{R}$ have nonzero and finite variance. Then*

$$\text{Var } f \ \log \left[\frac{\text{Var } f}{\sum_{k=1}^n \left(\mathbb{E}|\Delta_k|\right)^2} \right] \leq \sum_{k=1}^n \text{Ent } \Delta_k^2. \tag{3.5}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Before we prove (3.5), a few words of comment are needed. First, note that by the martingale decomposition of the variance, just as in (3.2),

$$\mathrm{Var}\ f = \sum_{k=1}^{n} \mathbb{E}\Delta_k^2.$$

Applying Jensen's inequality,

$$\mathrm{Var}\ f \geq \sum_{k=1}^{n} (\mathbb{E}|\Delta_k|)^2 , \tag{3.6}$$

so the term inside the logarithm in (3.5) is greater than or equal to 1.

As with Talagrand's influence inequality (Theorem 3.4), (3.5) is useful to obtain log-sublinear bounds if one can show that the right side of (3.6) is of lower order than Var f. We will see that this is the case in our setting as it can be shown to be at most of order $|x|^{-\alpha}\mathrm{Var}\ T(0,x)$ for some $\alpha > 0$.

Last, Equation (3.5) first appeared in a paper of Falik-Samorodnitsky [**84**, Equation (3)] as a functional version of an edge-isoperimetric inequality for Boolean functions.

We start the proof of (3.5) with the following lemma.

LEMMA 3.7. *Let f be a nonnegative function on a probability space such that* $\mathrm{Ent}(f^2) < \infty$. *Then,*

$$\mathbb{E}f^2 \log \frac{\mathbb{E}f^2}{(\mathbb{E}f)^2} \leq \mathrm{Ent}(f^2). \tag{3.7}$$

If f is identically zero we interpret the left side of (3.7) to be 0.

PROOF. As the inequality is preserved if we multiply f by any positive constant, we can assume that $\mathbb{E}f^2 = 1$. In this case, the inequality reads

$$-\log(\mathbb{E}f)^2 \leq \mathbb{E}f^2 \log f^2,$$

or

$$0 \leq \mathbb{E}f^2 \log(f\mathbb{E}f).$$

However, on the event $f > 0$, we can use the fact that $1 - x \leq \log x^{-1}$ with $x = (f\mathbb{E}f)^{-1}$ to obtain

$$0 = \mathbb{E}f^2 - 1 = \mathbb{E}\left[f^2\left[1 - \frac{1}{f\mathbb{E}f}\right]\mathbf{1}_{\{f>0\}}\right] \leq \mathbb{E}f^2 \log(f\mathbb{E}f).$$

Therefore, (3.7) holds. □

PROOF OF THEOREM 3.6. We use Lemma 3.7 on each Δ_k:

$$\sum_{k=1}^{n} \mathrm{Ent}(\Delta_k^2) \geq \sum_{k=1}^{n} \mathbb{E}(\Delta_k^2) \log \frac{\mathbb{E}(\Delta_k^2)}{(\mathbb{E}|\Delta_k|)^2} = -(\mathrm{Var}\ f) \sum_{k=1}^{n} \frac{\mathbb{E}(\Delta_k^2)}{\mathrm{Var}\ f} \log \frac{(\mathbb{E}|\Delta_k|)^2}{\mathbb{E}(\Delta_k^2)}.$$

As $\mathrm{Var}\ f = \sum_{k=1}^{n} \mathbb{E}(\Delta_k^2)$, we can apply Jensen's inequality with the function $x \mapsto -\log x$ to get the lower bound

$$-(\mathrm{Var}\ f) \log \sum_{k=1}^{n} \frac{\mathbb{E}(\Delta_k^2)}{\mathrm{Var}\ f} \frac{(\mathbb{E}|\Delta_k|)^2}{\mathbb{E}(\Delta_k^2)},$$

which is the left side of (3.5). □

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

With the Falik-Samorodnitsky inequality in our hands, we turn back to the proof of Theorem 3.5. As mentioned before, we will prove this theorem under the assumption that the passage times are uniformly distributed in the interval $[0,1]$. This assumption allow us to use the fact that this probability measure satisfies a log-Sobolev inequality (see [**141**, Theorem 8.14] for the uniform case and [**39**, Chapter 5] for an introduction and applications). Precisely, for any $n \geq 1$, under uniform measure on $[0,1]^n$, there exists $C > 0$ such that for any $f = f(x_1, \dots, x_n) : [0,1]^n \to \mathbb{R}$ that is smooth, one has

$$\operatorname{Ent}(f^2) \leq C \sum_{i=1}^{n} \mathbb{E} \left(\frac{\partial}{\partial x_i} f \right)^2 . \tag{3.8}$$

The above inequality combined with Theorem 3.6 gives

$$\operatorname{Var} f \ \log \left[\frac{\operatorname{Var} f}{\sum_{k=1}^{n} \left(\mathbb{E}|\Delta_k| \right)^2} \right] \leq C \sum_{k=1}^{n} \sum_{i=1}^{n} \mathbb{E} \left(\frac{\partial}{\partial x_i} \Delta_k \right)^2 . \tag{3.9}$$

If $i > k$, the derivative $\frac{\partial}{\partial x_i} \Delta_k$ is 0. If $i = k$, then the derivative is $\mathbb{E}[\frac{\partial}{\partial x_i} f \mid \Sigma_i]$. If $i < k$, the derivative is $\mathbb{E}[\frac{\partial}{\partial x_i} f \mid \Sigma_k] - \mathbb{E}[\frac{\partial}{\partial x_i} f \mid \Sigma_{k-1}]$. Now we use orthogonality of martingale differences to compute the double sum above as

$$\begin{aligned}
&\sum_{k=1}^{n} \sum_{i=1}^{n} \mathbb{E} \left(\frac{\partial}{\partial x_i} \Delta_k \right)^2 \\
&= \sum_{k=1}^{n} \sum_{i=1}^{k} \mathbb{E} \left(\frac{\partial}{\partial x_i} \Delta_k \right)^2 \\
&= \sum_{i=1}^{n} \sum_{k=i}^{n} \mathbb{E} \left(\frac{\partial}{\partial x_i} \Delta_k \right)^2 \\
&= \sum_{i=1}^{n} \left[\mathbb{E} \left(\mathbb{E} \left[\frac{\partial}{\partial x_i} f \;\middle|\; \Sigma_i \right] \right)^2 \right. \\
&\qquad \left. + \sum_{k=i+1}^{n} \mathbb{E} \left(\mathbb{E} \left[\frac{\partial}{\partial x_i} f \;\middle|\; \Sigma_k \right] - \mathbb{E} \left[\frac{\partial}{\partial x_i} f \;\middle|\; \Sigma_{k-1} \right] \right)^2 \right] \\
&= \sum_{i=1}^{n} \mathbb{E} \left(\mathbb{E} \left[\frac{\partial}{\partial x_i} f \;\middle|\; \Sigma_n \right] \right)^2 \\
&= \sum_{i=1}^{n} \mathbb{E} \left(\frac{\partial}{\partial x_i} f \right)^2 .
\end{aligned}$$

Combining this with (3.9), we obtain

$$\operatorname{Var} f \ \log \left[\frac{\operatorname{Var} f}{\sum_{k=1}^{n} \left(\mathbb{E}|\Delta_k| \right)^2} \right] \leq C \sum_{i=1}^{n} \mathbb{E} \left(\frac{\partial}{\partial x_i} f \right)^2 . \tag{3.10}$$

Proof of Theorem 3.5. We now apply (3.10) to the passage time $T(0,x)$, noting that since we assumed that our edge-weights are uniform, $T(0,x)$ is bounded,

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

and so we can extend the above inequality with $n \to \infty$ as

$$\operatorname{Var} T(0,x) \ \log\left[\frac{\operatorname{Var} T(0,x)}{\sum_{k=1}^{\infty}(\mathbb{E}|\Delta_k|)^2}\right] \leq C\sum_{i=1}^{\infty}\mathbb{E}\left(\frac{\partial}{\partial \tau_{e(i)}}T(0,x)\right)^2. \tag{3.11}$$

The derivative on the right is in the sense of distributions (since the passage time is not a smooth function of the edge-weights) and it is relative to the i-th edge-weight, where we have enumerated the edges in the lattice as $e(1), e(2), \ldots$. When the edge-weights are not bounded, one needs to argue that n can be taken to ∞ more carefully, imposing that T has at least $2+\epsilon$ moments (and this is guaranteed by existence of $2+\log$ moments for τ_e), to exploit uniform integrability.

Next we use the fact that

$$\frac{\partial}{\partial \tau_{e(i)}}T(0,x) = \mathbf{1}_{\{e(i)\in\underline{\mathrm{GEO}}(0,x)\}},$$

where we recall that $\underline{\mathrm{GEO}}(0,x)$ is the intersection of all geodesics from 0 to x. This holds Lebesgue a.s., which is satisfactory for us since the weights are uniform. So we obtain an upper bound for the right side of (3.11) of

$$C\mathbb{E}|\underline{\mathrm{GEO}}(0,x)| \leq C\|x\|_1.$$

Here we have used Lemma 3.3. Note that this is the same bound we obtained from the Efron-Stein inequality (in Kesten's method in the last section) but now the advantage is that we have an extra factor of $\log[\cdots]$ on the left of (3.11).

We are now left to show that $\sum_{k=1}^{\infty}(\mathbb{E}|\Delta_k|)^2$ is at most $\|x\|_1^a$ for some $a<1$. If we succeed in this, then (3.11) implies sublinear variance. Indeed, in that case, either $\operatorname{Var} T(0,x)$ is already $\leq \|x\|_1^{1-\epsilon}$ for $\epsilon=(1-a)/2$, or it is not, in which case, the term $\log[\cdots]$ is at least order $\log\|x\|_1$, and this completes the proof.

Unfortunately, it is not known how to show the bound on $\sum_{k=1}^{\infty}(\mathbb{E}|\Delta_k|)^2$. The reason is that it is at most of order $\sum_{k=1}^{\infty}(\mathbb{P}(e(k)\in\mathrm{GEO}(0,x)))^2$ (where $\mathrm{GEO}(0,x)$ is the self-avoiding geodesic from 0 to x with the most number of edges) and the only information we have on these probabilities is

$$\sum_{k=1}^{\infty}\mathbb{P}(e(k)\in\mathrm{GEO}(0,x)) = \mathbb{E}|\mathrm{GEO}(0,x)| \asymp \|x\|_1.$$

If the geodesic prefers to take certain nearly deterministic edges (say in a small tube centered on an ℓ^2-geodesic from 0 to x), then the sum of squares can be of order $\|x\|_1$. The work of Benjamini-Kalai-Schramm [**26**] introduced an averaging trick to get around this. The main realization is that if the system is translation-invariant, we can give the appropriate inequality, as in Section 3.2.1. First, we can restrict attention to a box around 0 of size $C\|x\|_1$ for a large constant C. If we pretend that all the probabilities $\mathbb{P}(e(k)\in\mathrm{GEO}(0,x))$ are equal, they must be of order $\|x\|_1^{1-d}$. Plugging this in gives the correct bound.

So we consider an averaged passage time (this form of averaging was used in [**10, 162**])

$$F_m = \frac{1}{Z_m}\sum_{\|z\|_1\leq m} T(z,z+x), \tag{3.12}$$

where the sum is over integer sites only and $m = \left\lfloor \|x\|_1^{1/4}\right\rfloor$. The term Z_m is the number of elements in the sum. By Jensen's inequality, we can still obtain the same

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

upper bound as in (3.11), using $f = F_m$:

$$\sum_{i=1}^{\infty} \mathbb{E}\left(\frac{1}{Z_m}\sum_{\|z\|_1\le m}\frac{\partial}{\partial \tau_{e(i)}}T(z,z+x)\right)^2 \le \frac{1}{Z_m}\sum_{\|z\|_1\le m}\sum_{i=1}^{\infty}\mathbb{P}(e(i)\in \underline{\mathrm{GEO}}(z,z+x)),$$

which is bounded by $C\|x\|_1$. Furthermore, F_m is not too different from $T(0,x)$: by the triangle inequality applied twice,

$$\|F_m - T(0,x)\|_2 \le \frac{1}{Z_m}\sum_{\|z\|_1\le m}\|T(z,z+x)-T(0,x)\|_2 \le 2\max_{\|z\|_1\le m}\|T(0,z)\|_2.$$

By our bound on the weights, then,

$$\|F_m - T(0,x)\|_2 \le 2\|x\|_1^{1/4},$$

which is $o(\|x\|_1^{1/2}/\log\|x\|_1)$, so it suffices to bound Var F_m. By the arguments in the beginning of the proof, we need only show that $\sum_{k=1}^{\infty}(\mathbb{E}|\Delta_k|)^2 \le \|x\|_1^a$ for some $a<1$, where Δ_k is now the martingale difference associated to F_m.

Similarly to (3.3), if $\mathbb{E}^{(k)}$ is expectation relative to the edge-weight $\tau_{e(k)}$, then

$$\mathbb{E}|\Delta_k| \le \frac{2}{Z_m}\sum_{\|z\|_1\le m}\mathbb{E}(\mathbb{E}^{(k)}T(z,z+x)-T(z,z+x))_+ \le \frac{2}{Z_m}\sum_{\|z\|_1\le m}\mathbb{P}(e(k)\in\underline{\mathrm{GEO}}(z,z+x)).$$

By translation invariance, we obtain the upper bound

$$\mathbb{E}|\Delta_k| \le \frac{2}{Z_m}\sum_{\|z\|_1\le m}\mathbb{P}(e(k)-z\in\underline{\mathrm{GEO}}(0,x)) \le Cm/Z_m. \tag{3.13}$$

Here we have used item 2 of Proposition 4.7, which implies that there exists $C>0$ such that for all nonzero $x\in\mathbb{Z}^d$ and all sets E of edges, one has

$$\mathbb{E}|\underline{\mathrm{GEO}}(0,x)\cap E| \le C\mathbb{E}\max_{y,z\in V}T(y,z),$$

where V is the set of endpoints of edges in E. Because our passage times τ_e are bounded by 1, the above is bounded by $C\mathrm{diam}\, E$. We apply this in (3.13) using E as the set of edges of the form $e(k)-z$ for $\|z\|_1\le m$.

Continuing from (3.13), we have

$$\sum_{k=1}^{\infty}(\mathbb{E}|\Delta_k|)^2 \le \frac{2Cm}{Z_m^2}\sum_{\|z\|_1\le m}\sum_{k=1}^{\infty}\mathbb{P}(e(k)\in\underline{\mathrm{GEO}}(z,z+x)) \le \frac{Cm}{Z_m}\|x\|_1.$$

This is bounded by $Cm^{1-d}\|x\|_1 \le C\|x\|_1^{3/4}$ and completes the proof. □

In the general case (assuming only $\mathbb{E}\tau_e^2(\log\tau_e)_+<\infty$), the difference in the proof is in the entropy bound (3.8), since no log-Sobolev inequality is available. By

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

writing each τ_e as the push-forward of an infinite sequence $\omega_e = (\omega_{e,1}, \omega_{e,2}, \ldots)$ of Bernoulli random variables, one can apply the Gross two-point entropy bound for

$$\sum_{k=1}^{\infty} \text{Ent } \Delta_k^2 \le C \sum_{k=1}^{\infty} \sum_{e} \mathbb{E} \left(\Delta_{e,k} F_m\right)^2 ,$$

where $\Delta_{e,k}$ is the discrete derivative operator relative to $\omega_{e,k}$. To give the upper bound $C\|x\|_1$ for the right side, one needs a careful analysis of these discrete derivatives and tools from the theory of greedy lattice animals. See [**65**] for details.

3.3. Logarithmic improvement to variance lower bound

3.3.1. The Newman-Piza lower bound. The following theorem by Newman-Piza in 1995 represents the state of the art for lower-bounding the variance of $T(0,x)$. It improves on Kesten's lower bound by a factor of $\log \|x\|_1$.

THEOREM 3.8 (Newman-Piza [**149**]). *Let $d = 2$ and let I be the infimum of the support of F. Assume $\mathbb{E}\tau_e^2 < \infty$ and* Var $\tau_e > 0$. *Assume in addition that one of the following two conditions is satisfied:*

$$I = 0 \text{ and } F(0) < p_c$$

$$I > 0 \text{ and } F(I) < \vec{p}_c.$$

There is a constant $B > 0$ such that

$$\text{Var } T(0,x) \ge B \log \|x\|_1 \tag{3.14}$$

for all nonzero $x \in \mathbb{Z}^2$.

In the case when τ_e is exponential with mean 1, the $\log n$ lower bound was also obtained, using different methods, by Pemantle and Peres [**150**]. Theorem 3.8 was extended to distributions in $\mathcal{M}_p$ (whose limit shape has a flat edge — see Section 2.5) first in the e_1-direction by Zhang [**188**] and then for all directions outside the percolation cone by Auffinger and Damron [**19**] (see also Kubota [**134**, Corollary 1.4], who reduced the moment condition of [**19**]). It is important to note, however, that inside the percolation cone, the variance of the passage time is of order constant [**188**], and this is a strong version of $\chi = 0$ in those directions. A related result is by Zhang [**184**], which shows a version of logarithmic divergence of fluctuations for the entire growing ball in dimensions $d \ge 3$ and distributions with Bernoulli $0,1$ edge-weights.

To justify Theorem 3.8, we will give a proof in a special case, when the τ_e's are Bernoulli, taking values 0 and 1. The general setting of Newman-Piza can be handled by a similar approach, and we briefly describe it after the proof. Our proof in the Bernoulli case will follow that of Newman-Piza (see "Special case and general approach" in [**149**, Section 3]), but replacing the estimate (3.6) from their paper with a more intuitive geometric argument.

PROOF OF THEOREM 3.8 IN THE BERNOULLI CASE. The proof will be split into two steps. In the first, as in the proof of the upper bound for the variance, we use a martingale representation for the variance to lower bound Var $T(0,x)$ by a sum of "influences" of edge variables on $T(0,x)$. In the second, we use an annular construction to show that the contribution to the variance given by influences of edges in dyadic annuli is at least a constant, and since the number of annuli will be of the order of $\log |x|$, the total variance must be at least $C \log |x|$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Assume for this proof that for $p \in (1-p_c, 1)$, the edge-weights are Bernoulli(p) distributed:

$$\mathbb{P}(\tau_e = 1) = p = 1 - \mathbb{P}(\tau_e = 0).$$

Enumerate the edges $(e(i))$ of $\mathbb{Z}^2$ in a spiral order starting from the origin and define a filtration (Σ_i), where Σ_i is the sigma-algebra generated by the weights $\tau_{e(1)}, \ldots, \tau_{e(i)}$, and Σ_0 is trivial. Writing $T = T(0,x)$, we may use L^2-orthogonality of martingale differences to find

$$\operatorname{Var} T = \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbb{E}[T \mid \Sigma_i] - \mathbb{E}[T \mid \Sigma_{i-1}]\right]^2. \tag{3.15}$$

The summand can be expressed in terms of the influence that the edge-weight $\tau_{e(i)}$ has on the variable T. Indeed, the first step (which we will prove in the next paragraph) is to show that if we define the event

$$F_i = \left\{\tau_{e(i)} = 1 \text{ and } e(i) \in \overline{\text{Geo}}(0,x)\right\},$$

where $\overline{\text{Geo}}(0,x)$ is the union of all geodesics from 0 to x, then

$$\mathbb{E}\left[\mathbb{E}[T \mid \Sigma_i] - \mathbb{E}[T \mid \Sigma_{i-1}]\right]^2 \geq p(1-p)\mathbb{P}(F_i)^2, \tag{3.16}$$

so that, along with (3.15), we have

$$\operatorname{Var} T \geq p(1-p) \sum_{i=1}^{\infty} \mathbb{P}(F_i)^2. \tag{3.17}$$

When F_i occurs, decreasing the value of $\tau_{e(i)}$ from 1 to 0 necessarily decreases the passage time T by 1. Therefore this inequality lower bounds the variance of T in terms of the influence of edge-weights $\tau_{e(i)}$.

To prove (3.16), we express, for a given i, our edge-weight configuration (τ_e) as $(\tau_{e(i)}, \hat{\tau}_i)$, where $\hat{\tau}_i$ is the restriction of (τ_e) to all edges except for $e(i)$, and we set for $\delta = 0, 1$, the variable T_i^δ to be the passage time from 0 to x in the configuration in which the weight at $e(i)$ is forced to be δ, and the weights at edges not equal to $e(i)$ are the same as those in our original configuration; that is, $T_i^\delta = T(\delta, \hat{\tau}_i)$. Then

$$H_i := T_i^1 - T_i^0$$

satisfies

$$H_i = \begin{cases} 1 & \text{if } e(i) \in \overline{\text{Geo}}(0,x) \text{ in } (1, \hat{\tau}_i) \\ 0 & \text{otherwise} \end{cases}.$$

Furthermore, we may write $T = T_i^0 + H_i \tau_{e(i)}$ and because T_i^0 and H_i depend only on $\hat{\tau}_i$ (and are therefore independent of $\tau_{e(i)}$), one has

$$\begin{aligned}
&\mathbb{E}[T \mid \Sigma_i] - \mathbb{E}[T \mid \Sigma_{i-1}] \\
&= \mathbb{E}[T_i^0 + H_i \tau_{e(i)} \mid \Sigma_i] - \mathbb{E}[T_i^0 + H_i \tau_{e(i)} \mid \Sigma_{i-1}] \\
&= \mathbb{E}[T_i^0 \mid \Sigma_i] + \tau_{e(i)} \mathbb{E}[H_i \mid \Sigma_i] - \mathbb{E}[T_i^0 \mid \Sigma_{i-1}] - \mathbb{E}\tau_{e(i)} \mathbb{E}[H_i \mid \Sigma_{i-1}] \\
&= \mathbb{E}[T_i^0 \mid \Sigma_{i-1}] + \tau_{e(i)} \mathbb{E}[H_i \mid \Sigma_{i-1}] - \mathbb{E}[T_i^0 \mid \Sigma_{i-1}] - \mathbb{E}\tau_{e(i)} \mathbb{E}[H_i \mid \Sigma_{i-1}] \\
&= (\tau_{e(i)} - \mathbb{E}\tau_{e(i)}) \mathbb{E}[H_i \mid \Sigma_{i-1}].
\end{aligned}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

So using independence, the left side of (3.16) equals

$$\begin{aligned}\mathbb{E}\left((\tau_{e(i)} - \mathbb{E}\tau_{e(i)})\mathbb{E}[H_i \mid \Sigma_{i-1}]\right)^2 &= \mathbb{E}\left(\tau_{e(i)} - \mathbb{E}\tau_{e(i)}\right)^2 \mathbb{E}\left(\mathbb{P}[H_i = 1 \mid \Sigma_{i-1}]\right)^2 \\ &= p(1-p)\mathbb{E}\left(\mathbb{P}[H_i = 1 \mid \Sigma_{i-1}]\right)^2.\end{aligned}$$

By independence and Jensen's inequality, we can relate the right term to the event F_i as

$$\mathbb{E}\left(\mathbb{P}[H_i = 1 \mid \Sigma_{i-1}]\right)^2 = \mathbb{E}\left(\frac{1}{p^2}\mathbb{P}(F_i \mid \Sigma_{i-1})^2\right) \geq \frac{1}{p^2}\mathbb{P}(F_i)^2 \geq \mathbb{P}(F_i)^2,$$

and this completes the proof of (3.16).

The second step is to give a lower bound for the right side of (3.17) using an annular construction. Note first that if $\mathcal{I} \subset \mathbb{N}$ is finite, then by Jensen's inequality,

$$\begin{aligned}\sum_{i\in\mathcal{I}} \mathbb{P}(F_i)^2 = |\mathcal{I}| \cdot \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \mathbb{P}(F_i)^2 &\geq |\mathcal{I}|\left(\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\mathbb{P}(F_i)\right)^2 \\ &= \frac{1}{|\mathcal{I}|}\left(\sum_{i\in\mathcal{I}}\mathbb{P}(F_i)\right)^2. \qquad (3.18)\end{aligned}$$

We will apply this inequality to the indices corresponding to roughly $\log|x|$ many annuli centered at 0 and show that the contribution from the terms for each annulus is at least a positive constant. If we let $\mathcal{B}_i$ be the set of edges with both endpoints in $[-2^i, 2^i]^2$, we then put $\mathcal{A}_i = \mathcal{B}_i \setminus \mathcal{B}_{i-1}$ and $\mathcal{A}_0 = \mathcal{B}_0$. Setting $\mathfrak{I} = \left\lfloor \log_2 \frac{\|x\|_\infty}{2} \right\rfloor$, we obtain from (3.17) and (3.18)

$$\begin{aligned}\mathrm{Var}\, T \geq p(1-p)\sum_{i=1}^{\mathfrak{I}} \sum_{j:e(j)\in\mathcal{A}_i} \mathbb{P}(F_j)^2 &\geq p(1-p)\sum_{i=1}^{\mathfrak{I}} \frac{1}{|\mathcal{A}_i|}\left(\sum_{j:e(j)\in\mathcal{A}_i}\mathbb{P}(F_j)\right)^2 \\ &\geq p(1-p)\sum_{i=1}^{\mathfrak{I}}\frac{1}{|\mathcal{A}_i|}\left(\mathbb{E}\mathcal{T}_i\right)^2, \qquad (3.19)\end{aligned}$$

where $\mathcal{T}_i$ is the maximum over all geodesics Γ from 0 to x of the quantity $\sum_{e\in\mathcal{A}_i\cap\Gamma}\tau_e$.

If $i \leq \mathfrak{I}$, then any geodesic Γ from 0 to x must cross the annulus $\mathcal{A}_i$, so for such i, we lower bound

$$\mathbb{E}\mathcal{T}_i \geq \mathbb{E}\min_{\Gamma}\sum_{e\in\Gamma}\tau_e, \qquad (3.20)$$

where the minimum is over all paths Γ that begin on the inner boundary of $\mathcal{A}_i$, end on the outer boundary, and stay entirely in $\mathcal{A}_i$. One can now derive for some constant $D_1 > 0$

$$\mathbb{E}\min_{\Gamma}\sum_{e\in\Gamma}\tau_e \geq D_1 2^i. \qquad (3.21)$$

In fact, this inequality is implied by the shape theorem, but for simplicity, we will use Kesten's lemma, which we state as Lemma 4.5, and which says that because

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

$F(0) < p_c$, there are constants $a, D_2 > 0$ such that for all $n \geq 1$,

$$\mathbb{P}\Big(\exists \text{ self-avoiding } \Gamma \text{ starting at } 0 \text{ with } |\Gamma| \geq n \text{ but with } T(\Gamma) < an \Big) \leq e^{-D_2 n}. \tag{3.22}$$

Since any Γ as in (3.21) has at least $D_3 2^i$ many edges, we may sum over all of the at most $D_4 2^i$ possible starting points of such Γ for

$$\mathbb{P}\left(\min_{\Gamma} \sum_{e \in \Gamma} \tau_e < a D_3 2^i \right) \leq D_4 2^i e^{-D_2 D_3 2^i} \leq D_5 e^{-D_6 2^i},$$

and from this (3.21) follows.

Placing (3.21) into (3.20), and then back into (3.19), we obtain

$$\text{Var } T \geq p(1-p) \sum_{i=1}^{\mathfrak{I}} \frac{1}{|\mathcal{A}_i|} (D_1 2^i)^2.$$

Since $|\mathcal{A}_i| \leq D_7 2^{2i}$ for some D_7, this sum is bounded below by $D_8 \log \|x\|_1$ for some $D_8 > 0$, and this completes the proof. $\square$

In the general setting of Newman-Piza, the proof is similar. The first step is to show the analogue of (3.17), that

$$\text{Var } T \geq C \sum_{i=1}^{\infty} \mathbb{P}(F_i)^2, \tag{3.23}$$

where F_i is the event that $e(i)$ (in the spiral ordering of edges $(e(i))$ of $\mathbb{Z}^2$) is in a geodesic from 0 to x, but instead of having $\tau_{e(i)} = 1$, we require that $\tau_{e(i)} \geq I + \delta$, where I is the infimum of the distribution of τ_e, and δ is a small number. This is accomplished again using a martingale decomposition but with a more detailed analysis of the relevant conditional expectations. See [**149**, Theorem 8].

After establishing (3.23), we again split the sum over dyadic annuli, but one must proceed differently depending on the nature of the distribution of τ_e at I. If $\mathbb{P}(\tau_e = I) \geq p_c$, then $I > 0$, and one must use oriented percolation techniques (see [**149**, p. 995-998]), and even differentiability properties of the shape in the case that the distribution of τ_e is in the class $\mathcal{M}_p$ (see [**19**, Section 4]). In the case that $\mathbb{P}(\tau_e = I) < p_c$, one writes as before (see (3.19))

$$\text{Var } T \geq C \sum_{i=1}^{\mathfrak{I}} \frac{1}{|\mathcal{A}_i|} (\mathbb{E}\mathcal{T}_i)^2. \tag{3.24}$$

Here, the variable $\mathcal{T}_i$ is defined slightly differently, by using auxiliary weights (σ_e) defined as

$$\sigma_e = \begin{cases} 0 & \text{if } \tau_e < I + \delta \\ 1 & \text{otherwise} \end{cases}.$$

Then $\mathcal{T}_i$ is defined as the maximum over all geodesics Γ from 0 to x of the quantity $\sum_{e \in \mathcal{A}_i \cap \Gamma} \sigma_e$. By applying Kesten's result (3.22) again with δ small enough that $\mathbb{P}(\tau_e \leq I + \delta) < p_c$, we obtain $\mathbb{E}\mathcal{T}_i \geq D_1 2^i$, and placing this back in (3.24), we obtain the logarithmic bound as before.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

It is crucial for the above proof that the dimension $d = 2$. Indeed, if we try the proof in dimensions $d \geq 3$, then it goes through to the end for

$$\text{Var } T \geq C \sum_{i=1}^{\mathfrak{I}} \frac{1}{|\mathcal{A}_i|} (D_1 2^i)^2.$$

However now $|\mathcal{A}_i|$ is of order 2^{di} and since $d \geq 3$, the sum is uniformly bounded.

We note as well that the Newman-Piza strategy breaks down on the d-dimensional torus. So we propose the following question.

QUESTION 3.3.1. Consider the two-dimensional torus of side length n: the graph $[0, n]^2 \cap \mathbb{Z}^2$ with opposite sides identified, and let T_n be the minimal passage time of all paths that wind around the torus once in the e_1-direction. Show that for some $C > 0$ and all $n \geq 1$,

$$\text{Var } T_n \geq C \log n.$$

3.3.2. Exact decompositions of variance: Fourier-Walsh. The martingale decomposition (3.15) gives an exact expression for Var $T(0, x)$ in terms of types of edge-weight influences, but in certain cases, we can find a more explicit one. Here we will describe the Fourier-Walsh basis for the hypercube, which allows for a variance formula that shows the relation between the Newman-Piza lower bound and the variance. The interested reader can see [**39**, Section 5.8] for more on the Fourier-Walsh basis and [**39**, Section 9.4] for a derivation of Talagrand's influence inequality (Theorem 3.4) using the Fourier-Walsh basis.

Suppose that our edge-weights satisfy $\mathbb{P}(\tau_e = 1) = 1/2 = \mathbb{P}(\tau_e = 2)$. Then any geodesic from 0 to x must have at most $2\|x\|_1$ number of edges (since such Γ must satisfy $|\Gamma| \leq T(\Gamma) = T(0, x) \leq 2\|x\|_1$), so $T(0, x)$ is a function of finitely many (say n) edge-weights, and is therefore an element of the space $L^2(\{1, 2\}^n)$ with the uniform measure. It is useful then to find an orthonormal basis of this space with which to represent the passage time.

For simplicity, we identify our space $L^2(\{1, 2\}^n)$ with $L^2(\{-1, +1\}^n)$ by mapping $1 \mapsto -1$ and $2 \mapsto +1$, and express an element of this space as $\omega = (\omega_i)_{1 \leq i \leq n}$.

DEFINITION 3.9. The Fourier-Walsh basis for $L^2(\{-1, +1\}^n)$ is the set $\mathcal{S} = \{\chi_S : S \subset \{1, \ldots, n\}\}$, where

$$\chi_S(\omega) = \prod_{i \in S} \omega_i,$$

with the convention that the empty product is 1.

Note that $\mathcal{S}$ contains 2^n elements, which is the dimension of $L^2(\{-1, +1\}^n)$, so to check that it is an orthonormal basis, we must only show that the elements are orthogonal and have norm 1. To this end, we use independence to compute (where Δ is the symmetric difference operator)

$$\begin{aligned}
\langle \chi_S, \chi_T \rangle = \mathbb{E}(\chi_S \chi_T) = \mathbb{E}\left(\prod_{i \in S} \omega_i \prod_{j \in T} \omega_j \right) &= \mathbb{E} \prod_{i \in S \Delta T} \omega_i \ \mathbb{E} \prod_{j \in S \cap T} \omega_j^2 \\
&= \mathbb{E} \prod_{i \in S \Delta T} \omega_i \\
&= \begin{cases} 0 & \text{if } S \neq T \\ 1 & \text{if } S = T \end{cases}.
\end{aligned}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Since $\mathcal{S}$ is an orthonormal basis, we can represent any $f : \{-1,+1\}^n \to \mathbb{R}$ as

$$f = \sum_S \hat{f}(S)\chi_S, \text{ where } \hat{f}(S) = \langle f, \chi_S\rangle = \mathbb{E} f\chi_S.$$

Then $\hat{f}(S)$ is called the S*-th* Fourier-Walsh coefficient for f.

Here are some basic properties of the Fourier-Walsh coefficients.

(1) $\hat{f}(\emptyset) = \mathbb{E}f$. To see this, note that $\chi_\emptyset \equiv 1$, so

$$\hat{f}(\emptyset) = \langle f, \chi_\emptyset\rangle = \mathbb{E} f\chi_\emptyset = \mathbb{E}f.$$

(2) $\mathbb{E}f^2 = \sum_S \hat{f}(S)^2$. This is a restatement of Parseval's formula: using orthogonality,

$$\mathbb{E}f^2 = \mathbb{E}\left(\sum_S \hat{f}(S)\chi_S \sum_T \hat{f}(T)\chi_T\right) = \sum_{S,T}\hat{f}(S)\hat{f}(T)\langle\chi_S,\chi_T\rangle = \sum_S \hat{f}(S)^2.$$

(3) $\mathrm{Var}\, f = \sum_{S\neq\emptyset}\hat{f}(S)^2$. This follows from the previous two items:

$$\mathrm{Var}\, f = \mathbb{E}f^2 - (\mathbb{E}f)^2 = \sum_S \hat{f}(S)^2 - \hat{f}(\emptyset)^2.$$

Using the exact decomposition for the variance in item 3, we can show the Newman-Piza lower bound (3.17). First if we set $f = T(0,x)$, we can compute the Fourier-Walsh coefficients $\hat{f}(S)$ corresponding to sets $S = \{e(i)\}$ of size 1:

$$\mathbb{E} f\chi_S = \frac{1}{2}\mathbb{E}\left[f(2,\omega_{i^c}) - f(1,\omega_{i^c})\right],$$

where we have expressed the edge-weight configuration ω as (ω_i,ω_{i^c}), and ω_{i^c} is the restriction of ω to edges $e(j)$ for $j \neq i$. As usual, by independence we can write the above as

$$\begin{aligned}\mathbb{E} f\chi_S &= \mathbb{E}\left[(f(2,\omega_{i^c}) - f(1,\omega_{i^c}))\mathbf{1}_{\{\omega_i=2\}}\right]\\ &= \mathbb{E}\left[(f(\omega) - f(1,\omega_{i^c}))\mathbf{1}_{\{\omega_i=2\}}\right]\\ &= \mathbb{P}(e(i)\in\overline{\mathrm{GEO}}(0,x) \text{ and } \tau_{e(i)} = 2),\end{aligned} \tag{3.25}$$

where we recall that $\overline{\mathrm{GEO}}(0,x)$ is the union of all geodesics from 0 to x. Using our formula for the variance in item 3 above, we obtain

$$\begin{aligned}\mathrm{Var}\, T(0,x) &= \sum_{S\neq\emptyset}\hat{f}(S)^2\\ &\geq \sum_{|S|=1}(\mathbb{E}f\chi_S)^2\\ &= \sum_i \mathbb{P}(e(i)\in\overline{\mathrm{GEO}}(0,x) \text{ and } \tau_{e_i} = 2)^2.\end{aligned}$$

This is the form of the Newman-Piza bound (3.17).

In this formulation, we can see that the Newman-Piza bound ignores all Fourier-Walsh coefficients for sets S with $|S| \geq 2$. We can also give a heuristic argument that this bound is not likely to be optimal; that is, the majority of the variance of $T(0,x)$ likely comes from coefficients corresponding to sets with many edges. To see why, we consider the case of continuous edge-weights, and take $f = T(0,ne_1)$. We further make the reasonable assumption that each edge intersecting the hyperplane $\{x\cdot e_1 = n/2\}$ is roughly equally likely to be in the geodesic from 0 to ne_1, as long

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

as the edge has maximal distance at most n^ξ from the e_1-axis. This is a strong version of assuming the existence of a "wandering exponent" ξ, which we discuss in detail in Section 4.2. Therefore such an edge should have roughly probability $1/n^{\xi(d-1)}$ to be in the geodesic.

For $0 \le k \le n/2$, an edge intersecting the hyperplane $\{x \cdot e_1 = k\}$ with distance at most k^ξ from the e_1-axis likely has probability about $1/k^{\xi(d-1)}$ to be in the geodesic from 0 to ne_1. Since there are order $k^{\xi(d-1)}$ edges of this type, the Newman-Piza lower bound should be at most

$$\sum_e \mathbb{P}(e \in \textsc{Geo}(0,x))^2 \lesssim 2\sum_{k=1}^{n/2} k^{\xi(d-1)}k^{-2\xi(d-1)} \lesssim \sum_{k=1}^{n} \frac{1}{k^{\xi(d-1)}} \lesssim n^{1-\xi(d-1)}.$$

If we assume the scaling relation $\chi = 2\xi - 1$, where χ is the variance exponent (see Section 4.3), we obtain $\xi = \frac{\chi+1}{2}$, and so the upper bound for the Newman-Piza lower bound is of order

$$n^{1-\frac{\chi+1}{2}(d-1)} \ll n^{2\chi} \text{ whenever } \chi > \frac{3-d}{d+3}.$$

The bound $\frac{3-d}{d+3}$ is negative for $d \ge 4$, so the Newman-Piza lower bound must be non-optimal in that case. For $d = 3$, the condition becomes $\chi > 0$, which is expected to be true, and for $d = 2$, there is strong evidence for $\chi = 1/3$, which is larger than $\frac{3-d}{d+3} = 1/5$. We conclude that the Newman-Piza lower bound is likely not of the true order of the variance for any dimension $d \ge 2$, and any progress toward stronger lower bounds must involve analysis of the dependence of $T(0,x)$ on multiple edges at once.

3.4. Concentration bounds

3.4.1. Subdiffusive concentration. In addition to sublinear variance bounds, there has been work to establish concentration inequalities for $T(0,x)$ on the scale $(|x|/\log|x|)^{1/2}$. These have so far only been exponential inequalities, not Gaussian ones.

As in the case of sublinear variance, the first such exponential inequality was not for general distributions, only those in the "nearly Gamma" class. The result we present below is from [**65**], and removes the nearly Gamma assumption, requiring only moment conditions. Note that the condition for the lower-tail inequality is weaker than that for the upper-tail inequality.

Theorem 3.10 (Damron-Hanson-Sosoe [**65**]). *Let $d \ge 2$ and suppose that $F(0) < p_c$. If $\mathbb{E}e^{\alpha\tau_e} < \infty$ for some $\alpha > 0$ then there exist $c_1, c_2 > 0$ such that for all $x \in \mathbb{Z}^d$ with $\|x\|_1 > 1$,*

$$\mathbb{P}\left(T(0,x) - \mathbb{E}T(0,x) \ge \lambda\sqrt{\frac{\|x\|_1}{\log\|x\|_1}}\right) \le c_1 e^{-c_2\lambda} \text{ for } \lambda \ge 0.$$

If $\mathbb{E}\tau_e^2(\log\tau_e)_+ < \infty$, then for all $x \in \mathbb{Z}^d$ with $\|x\|_1 > 1$,

$$\mathbb{P}\left(T(0,x) - \mathbb{E}T(0,x) \le -\lambda\sqrt{\frac{\|x\|_1}{\log\|x\|_1}}\right) \le c_1 e^{-c_2\lambda} \text{ for } \lambda \ge 0.$$

The main strategy is again due to Benaïm-Rossignol and follows the same lines as their proof of sublinear variance. The proof of the general case in [**65**] again

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

involves a Bernoulli encoding and estimating discrete derivatives after applying the two-point entropy estimate. Either way, one defines the averaged passage time

$$F_m = \frac{1}{Z_m} \sum_{\|z\|_1 \le m} T(z, z+x),$$

where the sum is over integer sites only, $m = \left\lfloor \|x\|_1^{1/8} \right\rfloor$, and Z_m is the number of terms in the sum. One can show [**65**, Section 2.1] that it suffices to derive the concentration inequality for F_m.

The main idea is to obtain a variance estimate for an exponential function of F_m analogous to the one obtained for T. By following the sublinear variance strategy from Section 3.2.2, with more technical difficulty, one obtains the following inequality: for some $C > 0$,

$$\text{Var } e^{\lambda F_m/2} \le K\lambda^2 \mathbb{E} e^{\lambda F_m} < \infty \text{ for } |\lambda| < \frac{1}{2\sqrt{K}} \text{ and } \|x\|_1 > 1, \tag{3.26}$$

where $K = \frac{C\|x\|_1}{\log \|x\|_1}$. Note the similarity to the entropy bound obtained below in (3.30) in the proof of Talagrand's concentration theorem, Theorem 3.12. The above inequality can be thought of as weaker, due to the presence of the variance instead of entropy, but stronger due to the logarithmic factor in the parameter K.

The above variance estimate is turned into an exponential concentration bound using the "iteration method." The following comes from [**39**, p. 70-71].

PROPOSITION 3.11. *If* $\mathbb{E}Z = 0$ *and for some constants* B, C *satisfying* $0 < C \le B$,

$$\text{Var } e^{tZ/2} \le Ct^2 \mathbb{E} e^{tZ} < \infty \textit{ for } t \in (0, B^{-1/2}),$$

then putting $\psi_Z(t) = \log \mathbb{E} e^{tZ}$, *one has*

$$\psi_Z(t) \le -2\log(1 - Ct^2) \textit{ for } t \in (0, B^{-1/2}).$$

PROOF. Beginning with

$$\mathbb{E} e^{tZ} - \left(\mathbb{E} e^{tZ/2}\right)^2 = \text{Var } e^{tZ/2} \le Ct^2 \mathbb{E} e^{tZ},$$

we obtain

$$\mathbb{E} e^{tZ} \le \frac{1}{1 - Ct^2} \left(\mathbb{E} e^{tZ/2}\right)^2,$$

or

$$\psi_Z(t) \le -\log(1 - Ct^2) + 2\psi_Z(t/2).$$

By induction, for $n \ge 1$,

$$\psi_Z(t) \le -\sum_{k=1}^{n} 2^{k-1} \log(1 - C(t/2^{k-1})^2) + 2^n \psi_Z(t/2^n).$$

Because $\mathbb{E}Z = 0$,

$$2^n \psi_Z(t/2^n) = t\frac{\psi_Z(t/2^n) - \psi_Z(0)}{t/2^n} \to t\psi_Z'(0) = 0,$$

so

$$\psi_Z(t) \le -\sum_{k=1}^{\infty} 2^{k-1} \log(1 - C(t/2^{k-1})^2). \tag{3.27}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

To bound these terms, we use the fact that $-u^{-1}\log(1-u)$ is non-decreasing in $u \in (0,1)$. Therefore

$$-C^{-1}(t/2^{k-1})^{-2}\log(1-C(t/2^{k-1})^2) \le -C^{-1}t^{-2}\log(1-Ct^2) \text{ for } k \ge 1.$$

Rewritten,

$$-2^{k-1}\log(1-C(t/2^{k-1})^2) \le -2^{-(k-1)}\log(1-Ct^2).$$

Placing this in (3.27),

$$\psi_Z(t) \le -\log(1-Ct^2)\sum_{k=1}^{\infty} 2^{-(k-1)} = -2\log(1-Ct^2).$$

□

To make the last bound more clear, we can use the inequality

$$-\log(1-u) \le \frac{u}{1-u} \text{ for } u \in [0,1),$$

which follows from the mean value theorem. So we obtain

$$\psi_Z(t) \le \frac{2Ct^2}{1-Ct^2} \le \frac{4Ct^2}{2(1-\sqrt{C}t)} \text{ for } t \in (0, B^{-1/2}). \tag{3.28}$$

In our case, we write (3.26) equivalently as

$$\mathrm{Var}\; e^{\lambda(F_m - \mathbb{E}F_m)/2} \le K\lambda^2 \mathbb{E}e^{\lambda(F_m-\mathbb{E}F_m)}$$

and apply (3.28) with $Z = F_m - \mathbb{E}F_m$, and $B = C = K$ to get

$$\log \mathbb{E}e^{t(F_m-\mathbb{E}F_m)} \le \frac{4Kt^2}{2(1-\sqrt{K}t)} \text{ for } t \in (0, K^{-1/2}).$$

A similar bound holds if $F_m - \mathbb{E}F_m$ is replaced by $\mathbb{E}F_m - F_m$, so

$$\log \mathbb{E}e^{t|F_m-\mathbb{E}F_m|} \le \frac{8Kt^2}{2(1-\sqrt{K}t)} \text{ for } t \in (0, K^{-1/2}).$$

By using the Markov inequality,

$$\mathbb{P}(X \ge x) = \mathbb{P}(e^{\lambda X} \ge e^{\lambda x}) \le e^{-\lambda x}\mathbb{E}e^{\lambda X},$$

which is valid for $\lambda \ge 0$, and optimizing over λ, one can then complete the first inequality of Theorem 3.10.

3.4.2. Talagrand's theorem via the entropy method. Here we will give the concentration argument from the appendix of [**65**, Corollary A.5]. The goal will be to give a Gaussian concentration inequality for the passage time about its mean assuming certain moment conditions for the edge weights. The lower-tail inequality comes from [**70**]. Our aim will be to prove the following result. It was initially established by Talagrand in [**166**], using different methods.

THEOREM 3.12 (Talagrand [**166**]). *Let $d \ge 2$. Assuming $F(0) < p_c$ and $\mathbb{E}e^{\alpha\tau_e} < \infty$ for some $\alpha > 0$, there exist $C_1, C_2 > 0$ such that*

$$\mathbb{P}\left(T(0,x) - \mathbb{E}T(0,x) \ge t\sqrt{\|x\|_1}\right) \le e^{-C_1 t^2} \text{ for } t \in \left(0, C_2\sqrt{\|x\|_1}\right).$$

If $F(0) < p_c$ and $\mathbb{E}Y^2 < \infty$, where Y is the minimum of d i.i.d. copies of τ_e, then also

$$\mathbb{P}\left(T(0,x) - \mathbb{E}T(0,x) \le -t\sqrt{\|x\|_1}\right) \le e^{-C_1 t^2} \text{ for all } t \ge 0.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

PROOF. Write $T = T(0,x)$. We will show only the upper-tail inequality, and we will assume $\mathbb{E}e^{\alpha\tau_e} < \infty$ for all $\alpha > 0$ for simplicity. The idea will be to set $\psi(\lambda) = \log \mathbb{E}e^{\lambda(T-\mathbb{E}T)}$ and to note that it suffices to show for some $C, C' > 0$ independent of x,

$$\psi(\lambda) \leq C\|x\|_1\lambda^2 \text{ for } \lambda \in (0, C'). \tag{3.29}$$

Indeed, one has for $\lambda > 0$

$$\begin{aligned}\mathbb{P}(T - \mathbb{E}T \geq t\sqrt{\|x\|_1}) \leq \mathbb{P}(e^{\lambda(T-\mathbb{E}T)} \geq e^{t\lambda\sqrt{\|x\|_1}}) &\leq e^{\psi(\lambda)-t\lambda\sqrt{\|x\|_1}} \\ &\leq e^{C\|x\|_1\lambda^2 - t\lambda\sqrt{\|x\|_1}}.\end{aligned}$$

Choosing $\lambda = t/2C\sqrt{\|x\|_1}$ would then complete the proof.

To show the bound (3.29), we will use the "Herbst argument." Setting

$$\text{Ent } X = \mathbb{E}X\log X - \mathbb{E}X\log\mathbb{E}X$$

for a nonnegative random variable X, we will aim to show for some $C, C' > 0$ independent of x, (compare to (3.26))

$$\text{Ent } e^{\lambda T} \leq C\|x\|_1\lambda^2\mathbb{E}e^{\lambda T} \text{ for } \lambda \in (0, C'). \tag{3.30}$$

This implies that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\psi(\lambda)}{\lambda}\right) \leq C\|x\|_1 \text{ for } \lambda \in (0, C'),$$

and integrating this inequality gives (3.29).

So we focus on proving (3.30). Enumerating the edge variables (in any deterministic order) as $(\tau_{e(1)}, \tau_{e(2)}, \ldots)$, then

$$\text{Ent } e^{\lambda T} \leq \sum_{i=1}^{\infty} \mathbb{E}\text{ Ent}_i\ e^{\lambda T}, \tag{3.31}$$

where Ent_i is entropy relative to only the edge-weight $\tau_{e(i)}$. This is known as "tensorization of entropy" [**39**, Theorem 4.10]. We now apply a "modified Log-Sobolev inequality" of Boucheron-Lugosi-Massart [**39**, Theorem 6.15].

LEMMA 3.13 (Symmetrized modified LSI). *Let $q(x) = x(e^x - 1)$. If X is a random variable and X' is an independent copy of X, then for all $t \in \mathbb{R}$,*

$$\text{Ent } e^{tX} \leq \mathbb{E}\left[e^{tX}q(\lambda(X' - X)_+)\right].$$

Use the symmetrized modified LSI in (3.31):

$$\text{Ent } e^{\lambda T} \leq \sum_{i=1}^{\infty} \mathbb{E}e^{\lambda T}q(\lambda(T_i' - T)_+). \tag{3.32}$$

Here T_i' is the passage time from 0 to x in the edge-weight configuration in which the weight $\tau_{e(i)}$ is replaced by an independent copy $\tau_{e(i)}'$, and all other edge-weights remain the same. As in the argument for (3.4), we know that $T_i' - T$ is only positive if $e(i)$ is in $\underline{\text{GEO}}(0,x)$, the intersection of all geodesics from 0 to x, in the original edge-weight configuration. Therefore, we can give the upper bound

$$\text{Ent } e^{\lambda T} \leq \sum_{i=1}^{\infty} \mathbb{E}e^{\lambda T}q(\lambda(T_i' - T)_+)\mathbf{1}_{\{e(i)\in\underline{\text{GEO}}(0,x)\}}.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

The function $x \mapsto q(x)$ is monotone increasing for $x \geq 0$ so using the bound $T_i' - T \leq \tau_{e(i)}'$ and independence, we have

$$\text{(3.33)}\quad \text{Ent } e^{\lambda T} \leq \sum_{i=1}^{\infty} \mathbb{E} e^{\lambda T} q(\lambda \tau_{e(i)}') \mathbf{1}_{\{e(i) \in \underline{\mathrm{Geo}}(0,x)\}} = \mathbb{E} q(\lambda \tau_e)\ \mathbb{E}\left[e^{\lambda T} |\underline{\mathrm{Geo}}(0,x)|\right].$$

To apply the Herbst argument we would like to decouple $e^{\lambda T}$ from $|\underline{\mathrm{Geo}}(0,x)|$. Unfortunately, the variable $|\underline{\mathrm{Geo}}(0,x)|$ is not bounded, so we cannot just pull it out. So we use a variational characterization of entropy [**39**, Remark 4.4]:

$$\text{Ent } X = \sup\{\mathbb{E} XY : \mathbb{E} e^Y \leq 1\},$$

which implies that for $X \geq 0$ and any Y,

$$\mathbb{E} XY \leq \text{Ent } X + \mathbb{E} X \log \mathbb{E} e^Y.$$

We therefore write for arbitrary $a > 0$,

$$\mathbb{E} e^{\lambda T} |\underline{\mathrm{Geo}}(0,x)| \leq a \text{ Ent } e^{\lambda T} + a\ \mathbb{E} e^{\lambda T} \log \mathbb{E} \exp\left(\frac{|\underline{\mathrm{Geo}}(0,x)|}{a}\right).$$

Combining with (3.33), if $a\mathbb{E} q(\lambda \tau_e) < 1$,

$$\text{Ent } e^{\lambda T}(1 - a\mathbb{E} q(\lambda \tau_e)) \leq a\mathbb{E} q(\lambda \tau_e) \mathbb{E} e^{\lambda T} \log \mathbb{E} \exp\left(\frac{|\underline{\mathrm{Geo}}(0,x)|}{a}\right),$$

or

$$\text{Ent } e^{\lambda T} \leq \mathbb{E} e^{\lambda T} \frac{a\mathbb{E} q(\lambda \tau_e)}{1 - a\mathbb{E} q(\lambda \tau_e)} \log \mathbb{E} \exp\left(\frac{|\underline{\mathrm{Geo}}(0,x)|}{a}\right).$$

To control these terms we will need a lemma from [**65**, Corollary 1.4(2)] which follows from our later equation (4.3) and which is a strong version of a geodesic length bound.

LEMMA 3.14. *Assuming $\mathbb{E} e^{\alpha \tau_e} < \infty$ for all $\alpha > 0$, there exist $a, c_1 > 0$ such that*

$$\log \mathbb{E} \exp\left(\frac{|\underline{\mathrm{Geo}}(0,x)|}{a}\right) \leq c_1 \|x\|_1 \textit{ for all } x \in \mathbb{Z}^d.$$

Applying the lemma,

$$\text{Ent } e^{\lambda T} \leq c_1 \|x\|_1 \frac{a\mathbb{E} q(\lambda \tau_e)}{1 - a\mathbb{E} q(\lambda \tau_e)}\ \mathbb{E} e^{\lambda T} \text{ if } a\mathbb{E} q(\lambda \tau_e) < 1,$$

or

$$\text{(3.34)}\qquad \text{Ent } e^{\lambda T} \leq 2c_1 \|x\|_1 a\mathbb{E} q(\lambda \tau_e) \mathbb{E} e^{\lambda T} \text{ if } a\mathbb{E} q(\lambda \tau_e) < 1/2.$$

Note that by the dominated convergence theorem,

$$\lim_{\lambda \downarrow 0} \frac{\mathbb{E} q(\lambda \tau_e)}{\lambda^2} = \lim_{\lambda \downarrow 0} \mathbb{E}\left(\frac{\tau_e(e^{\lambda \tau_e} - 1)}{\lambda}\right) = \mathbb{E} \tau_e^2.$$

So we can find $c_2 > 0$ such that if $\lambda \in (0, c_2)$ then $\mathbb{E} q(\lambda \tau_e) \leq 2\lambda^2 \mathbb{E} \tau_e^2$. For such λ, note that the condition $a\mathbb{E} q(\lambda \tau_e) < 1/2$ from (3.34) holds if $2a\lambda^2 \mathbb{E} \tau_e^2 < 1/2$, which occurs if λ is smaller than some positive $c_3 < c_2$. So we obtain the desired equation (3.30): for some $c_4 > 0$,

$$\text{Ent } e^{\lambda T} \leq 4c_1 \|x\|_1 \lambda^2 \mathbb{E} \tau_e^2 \mathbb{E} e^{\lambda T} \leq \lambda^2 c_4 \|x\|_1 \mathbb{E} e^{\lambda T} \text{ for } \lambda \in (0, c_3).$$

□

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

3.5. Convergence of the mean for subadditive ergodic processes

We return to the perspective adopted in Section 2.6. Let X be a subadditive ergodic process satisfying the hypotheses of Theorem 2.2. As discussed previously, since the leading order (shape theorem) behavior of FPP is established using the general framework of the subadditive ergodic theorem, one would hope that general abstract arguments for subadditive sequences could be used to establish limit theorems and other sharper results. Since the class of processes satisfying Theorem 2.2 is much too large to characterize the exponents and limiting behavior of FPP, we will have to impose additional axioms in order to derive useful results. Furthermore, many results will require us to go beyond considering subadditive sequences X as above and instead consider the d-dimensional structure of the model.

We will restrict our attention here to the convergence of the mean $\mathbb{E}X_{0,n}/n$ of our subadditive processes to their limiting μ (henceforth, we assume that the sequence is ergodic). Recall that in the FPP setting, the fluctuations of $T(0,x) - \mu(x)$ can be written as:

$$o(\|x\|_1) = \underbrace{T(0,x) - \mathbb{E}T(0,x)}_{random\ fluctuations} + \underbrace{\mathbb{E}T(0,x) - \mu(x)}_{nonrandom\ fluctuations} .$$

Earlier in this chapter, we discussed bounds on the random fluctuations in FPP. We therefore shift to the problem of convergence of the mean in order to characterize the other error term in the convergence to the FPP limit shape.

3.5.1. Nonrandom fluctuations for subadditive sequences . All of the existing methods for controlling the rate of convergence of the mean require also some control of the corresponding random fluctuations. As mentioned before, it is expected that the random fluctuations in FPP are governed by the fluctuation exponent χ. It is reasonable to postulate the existence of a similar exponent γ for the nonrandom fluctuations. Specifically, one expects that (for a suitable definition of "$\sim$")

$$T(0,x) - \mathbb{E}T(0,x) \sim \|x\|_1^{\chi}, \quad \mathbb{E}T(0,x) - \mu(x) \sim \|x\|_1^{\gamma}$$

for some exponents χ and γ. This motivates the following definitions from [**21**], made for any subadditive ergodic sequence $(X_{m,n})$ satisfying the hypotheses of Theorem 2.2.

DEFINITION 3.15. The exponents $\underline{\gamma}$ and $\overline{\gamma}$ are defined as

$$\underline{\gamma} = \liminf_n \frac{\log(\mathbb{E}X_{0,n} - n\mu)}{\log n} \text{ and } \overline{\gamma} = \limsup_n \frac{\log(\mathbb{E}X_{0,n} - n\mu)}{\log n}.$$

Here, recall μ is the almost-sure limit $\lim_n X_{0,n}/n$; we make the convention that $\log 0 = -\infty$.

DEFINITION 3.16. For $p > 0$, the fluctuation exponents $\underline{\chi}_p$ and $\overline{\chi}_p$ are defined using the L^p norm as

$$\underline{\chi}_p = \liminf_n \frac{\log \|X_{0,n} - \mathbb{E}X_{0,n}\|_p}{\log n} \quad \text{and} \quad \overline{\chi}_p = \limsup_n \frac{\log \|X_{0,n} - \mathbb{E}X_{0,n}\|_p}{\log n}.$$

Note that $\underline{\chi}_p \leq \overline{\chi}_p$ and by Jensen's inequality,

$$\underline{\chi}_p \leq \underline{\chi}_q \text{ and } \overline{\chi}_p \leq \overline{\chi}_q \text{ if } p \leq q.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

REMARK 3.17. The definitions above only specify the first order growth of the fluctuation terms. For instance, in FPP with $X_{m,n} = T(me_1, ne_1)$, if there exist positive constants c, C such that

$$cn^{2\chi} \leq \operatorname{Var} T(0, ne_1) \leq Cn^{2\chi} \text{ for all large } n,$$

then $\underline{\chi}_2 = \overline{\chi}_2 = \chi$, and similarly for $p \neq 2$ and $\underline{\gamma}, \overline{\gamma}$.

Our first result will be a lower bound of the form "$\gamma \geq \chi$" under some assumptions on the sequence which are natural in the case of FPP. We explain these and the intuition. Recall that the χ exponent in FPP is, except in the one-dimensional case, expected to be strictly smaller than 1/2. Suppose that we had a subadditive ergodic sequence $X_{0,n}$ for which χ and γ exponents existed in a moderately strong sense (i.e. $\overline{\chi}_p = \underline{\chi}_q$ for all q, $p \geq 1$ and $\overline{\gamma} = \underline{\gamma}$). Assume also for a contradiction that $\chi < 1/2$ but $\gamma < \chi$. Last, assume the sequence $(X_{Km,Kn})$ is weakly dependent for large K in a sense which we leave imprecise.

For any positive integer K, we can write using subadditivity

$$\begin{aligned}
[X_{0,n} - \mathbb{E}X_{0,n}] &\leq \sum_{i=1}^{n/K} [X_{(i-1)K,iK} - \mathbb{E}X_{0,K}] + (n/K)\mathbb{E}X_{0,K} - \mathbb{E}X_{0,n} \\
&\leq \sum_{i=1}^{n/K} [X_{(i-1)K,iK} - \mathbb{E}X_{0,K}] + (n/K)[\mathbb{E}X_{0,K} - K\mu]. \qquad (3.35)
\end{aligned}$$

If the sequence $X_{m,n}$ is suitably weakly dependent, then we could expect that the sum on the right-hand side of (3.35) obeys a central limit theorem, and in particular frequently takes values on the order of $-n^{1/2}K^{\chi-1/2}$. The second term is of order $nK^{\gamma-1}$. In particular, if we were to choose K on the order of $n^{1-\delta}$ for $0 < \delta \ll 1$, the first term would dominate, giving that $X_{0,n} - \mathbb{E}X_{0,n}$ typically has lower tail fluctuations of order $-n^{\chi+\delta(1/2-\chi)}$. This is a contradiction to the definition of χ under the assumption that $\chi < 1/2$.

We are now ready to state some rigorous results following the line of reasoning above; these results appeared in [**21**]. In order to avoid precisely describing the weak dependence axioms needed, we restrict to the case of FPP on $\mathbb{Z}^d$ (here, the "weak dependence axiom" amounts to control of the diameter of geodesics – see Theorem 4.8). In the language of FPP, this result says that given existence of the fluctuation exponent χ, the nonrandom fluctuations are at least of the same order as the standard deviation of the passage time.

In the FPP results below, we consider exponents defined as in Definitions 3.15 and 3.16 for some arbitrary fixed direction x (that is, $\overline{\gamma} = \limsup[\mathbb{E}T(0, nx) - n\mu(x)]/\log n$, etc.).

THEOREM 3.18 (Auffinger-Damron-Hanson [**21**]). *Consider the case of FPP on $\mathbb{Z}^d$ for $d \geq 2$. Assume $\mathbb{E}\tau_e^{2+\delta} < \infty$ for some $\delta > 0$.*

(1) *If $\chi := \underline{\chi}_2 = \overline{\chi}_{2+\delta}$, then*

$$\begin{cases} \underline{\gamma} \geq \chi & \text{if } \chi \neq 1/2 \\ \overline{\gamma} \geq \chi - \frac{1}{2}(2^{\beta^{-1}} - 1) & \text{if } \operatorname{Var} T(0, nx) = O\left(\frac{n}{(\log n)^\beta}\right) \text{ for some } \beta > 0 \end{cases}.$$

(2) *If $\chi := \overline{\chi}_2 = \overline{\chi}_{2+\delta}$, then if $\chi < 1/2$, $\overline{\gamma} \geq \chi$.*

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

REMARK 3.19. The result above should be compared with the classical example of a sum of i.i.d. random variables with finite mean and finite $2+\delta$ moment. In this case, $\chi = 1/2$ and $\overline{\gamma} = \underline{\gamma} = -\infty$ and these exponents do not satisfy the conclusions of the theorem above. Note that this example is equivalent to FPP on $\mathbb{Z}$ with the appropriate moment conditions.

Another theorem of [**21**] gives the bound $\overline{\gamma} \geq -1/2$ under minimal assumptions. This result should be compared to [**125**, Theorem 1], where it is shown that $\underline{\gamma} \geq -1$.

THEOREM 3.20 (Auffinger-Damron-Hanson [**21**]). *Consider FPP on $\mathbb{Z}^d$ for $d \geq 2$. Assume $F(0) < p_c$, that the distribution of τ_e is not concentrated at a point, and that*

$$\mathbb{E}e^{\alpha\tau_e} < \infty \text{ for some } \alpha > 0. \tag{3.36}$$

One has the bound $\overline{\gamma} \geq -1/2$: for any nonzero $x \in \mathbb{Z}^d$ and $\epsilon > 0$,

$$\mathbb{E}T(0, nx) - n\mu(x) \geq n^{-\frac{1}{2}-\epsilon} \text{ for infinitely many } n.$$

One may be interested in the sharpness of the result $\gamma \geq \chi$. In the case of FPP on $\mathbb{Z}^d$ ($d \geq 2$), it is reasonable to expect the result is sharp — that is, $\gamma = \chi$ — as explained in the following remark.

REMARK 3.21. Alexander [**9**] has noted that if $\hat{\chi}$ is any number such that for some $a > 0$,

$$\mathbb{P}\left(|T(0,y) - \mathbb{E}T(0,y)| \geq \lambda \|y\|_1^{\hat{\chi}}\right) \leq e^{-a\lambda} \text{ for all } \lambda \geq 0,\ y \in \mathbb{Z}^d,$$

then $\overline{\gamma} \leq \hat{\chi}$. This follows from an extension of his method for controlling nonrandom fluctuations, discussed in Section 3.5.3 below. Note that if this exponential inequality holds for some $\hat{\chi}$, then $\overline{\chi}_p \leq \hat{\chi}$ for all $p > 0$. Combining these observations with Theorem 3.18 above, if $\hat{\chi}$ can be taken to be $\chi := \hat{\chi} = \underline{\chi}_2$, then

$$\gamma := \underline{\gamma} = \overline{\gamma} = \chi$$

when $\chi < 1/2$ and $\overline{\gamma} = \chi$ under the assumption $\mathrm{Var}\ T(0, nx) = O(n/(\log n)^\beta)$ for every $\beta > 0$.

QUESTION 3.5.1. Show that $\underline{\chi}_2 = \overline{\chi}_{2+\delta}$ for some $\delta > 0$.

A solution for the question above has the following consequence that improves Theorem 3.20.

REMARK 3.22. Combining Theorem 3.5 with Theorem 3.18, if $\underline{\chi}_2 = \overline{\chi}_{2+\delta}$ for some $\delta > 0$ and $F(0) < p_c$ holds, then $\overline{\gamma} \geq 0$.

We now return to the general setting of subadditive ergodic sequences $(X_{m,n})$. Our goal is to find extensions to the axioms of Theorem 2.2 which could guarantee upper bounds on the exponent γ, complementing the lower bounds given above.

Note that by subadditivity, for any n,

$$\mathbb{E}X_{0,2n} \leq \mathbb{E}X_{0,n} + \mathbb{E}X_{n,2n} = 2\mathbb{E}X_{0,n}.$$

The magnitude of nonrandom fluctuations is related to the degree to which the above inequality is strict. Let us abbreviate $\mathbb{E}X_{0,n} =: H(n)$. As a first step, assume the above inequality were actually an equality; then

$$\mu = \lim_k H(2^k n)/(2^k n) = \lim_k H(n)/n = H(n)/n,$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

in which case we would have $H(n) = n\mu$ and therefore γ would be $-\infty$.

In fact, one can give upper bounds on $H(n) - n\mu$ assuming just some weaker quantitative control on $H(2n) - 2H(n)$. To illustrate this, assume that for some function $j : \mathbb{N} \to [0, \infty)$, one has

$$H(2n) \geq 2H(n) - j(n). \tag{3.37}$$

We then have the following lemma from [**115**, Lemma 4.2].

LEMMA 3.23. *Suppose that j satisfies $\Psi := \limsup_n j(2n)/j(n) < 2$, $j(n)/n \to 0$, and* (3.37). *Then for $c > 1/(2-\Psi)$ and all large n, one has $H(n) \leq n\mu + cj(n)$.*

PROOF. For c as above, set $a(n) = H(n) - cj(n)$. Then for large n,

$$a(2n) = H(2n) - cj(2n) \geq 2H(n) - 2cj(n)\left(\frac{1}{2c} + \frac{j(2n)}{2j(n)}\right) \geq 2a(n).$$

So for $m \geq 1$,

$$\frac{a(2^m n)}{2^m n} \geq \frac{a(n)}{n}.$$

Taking $m \to \infty$, we obtain

$$\frac{a(n)}{n} \leq \mu,$$

or

$$H(n) \leq n\mu + cj(n).$$

□

We would like to apply Lemma 3.23 to bound nonrandom fluctuations in FPP, taking $H(n) = \mathbb{E}T(0, nx)$ and $\mu = \mu(x)$. By that lemma, if we could show that $H(n)$ is close to being a linear function of n, we can upper bound $H(n) - n\mu(x)$. This approach is most successful in the axis direction e_1 for symmetry reasons, and a version of this method is the basis for several such bounds (see Zhang [**187**], Rhee [**153**], and another reflection trick in Alexander [**8**]). One typically tries to use the fact that the passage time along geodesics is additive. Define the events

$$\begin{aligned} A &= \{T(0, ne_1) - \mathbb{E}T(0, ne_1) \geq -j(n)/3\} \\ &\quad \cap \{T(ne_1, 2ne_1) - \mathbb{E}T(0, ne_1) \geq -j(n)/3\}, \\ B &= \{T(0, 2ne_1) \leq 2\mathbb{E}T(0, ne_1) + j(n)/3\}, \\ C &= \{\text{a geodesic from } 0 \text{ to } 2ne_1 \text{ passes through } ne_1\}. \end{aligned}$$

Then if $A \cap B \cap C$ has positive probability, on this event, one has

$$T(0, 2ne_1) = T(0, ne_1) + T(ne_1, 2ne_1)$$

and therefore

$$\mathbb{E}T(0, 2ne_1) \geq 2\mathbb{E}T(0, ne_1) - j(n).$$

It is reasonable that C has probability at least some negative power in n given the heuristics for geodesic wandering, which would — along with strong bounds on the probabilities of A and B, provided by concentration results on the passage time — complete the argument. Unfortunately, not enough is known about geodesic wandering to make this argument work in general directions. In practice, showing this positive probability statement has been done using a reflection trick, and unfortunately this method typically does not work in non-axis directions.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

3.5.2. Rate of convergence of the expected ball. Since the axioms of Lemma 3.23 are not easy to verify in FPP, one is led to generalize from subadditive sequences to some broader framework. Setting $h(x) := \mathbb{E}T(0,x)$, it is obvious that h has more than the structure of a subadditive sequence: it is subadditive as a function of $\mathbb{Z}^d$. Since the time constant and h itself depend on the entire d-dimensional structure of the lattice, it is reasonable to hope that one could progress by considering the geometry induced on the lattice by h.

We can think of h as a metric given by $h(x,y) = h(y-x)$, and although the metric space $(\mathbb{Z}^d, h)$ does not typically have geodesics, we would like to show that it has approximate midpoints. Recall that to implement the approach outlined following Lemma 3.23, we would like to show that ne_1 is nearly a midpoint between 0 and $2ne_1$. A realization of Tessera is that we do not have to choose ne_1: it actually suffices to find skeletons of paths $0 = x_0, x_1, \ldots, x_k = 2ne_1$ which are approximately geodesics. That is, we would like to give a lower bound for

$$h(2ne_1) - \sum_{i=0}^{k-1} h(x_{i+1} - x_i),$$

for some suitably chosen skeleton. We follow [**167**] for the following definitions and results.

Definition 3.24. Let $N : [0,\infty) \to (0,\infty)$ be an increasing function such that $\lim_{\alpha\to\infty} N(\alpha) = \infty$. A metric space (X,d) is called SAG(N) (strongly asymptotically geodesic) if there exists $\alpha_0 \geq 0$ such that for all integers $m \geq 1$, and for all $x, y \in X$ such that $d(x,y)/m \geq \alpha_0$, there exists a sequence $x = x_0, \ldots, x_m = y$ satisfying, for all $0 \leq i \leq m-1$,

$$\frac{d(x,y)}{m}\left(1 - \frac{1}{N(\alpha)}\right) \leq d(x_i, x_{i+1}) \leq \frac{d(x,y)}{m}\left(1 + \frac{1}{N(\alpha)}\right),$$

where $\alpha = d(x,y)/m$; and for all $x \in X$ and large enough r,

$$B\left(x, \left(1 + \frac{1}{N(r)} r\right)\right) \subset [B(x,r)]_{\frac{6r}{N(r)}}.$$

Here, $B(x,r)$ is the r-neighborhood of x and for any $T > 0$ and set $S \subset X$, $[S]_T$ is the T-neighborhood of S.

We briefly try to motivate the above definition in the setting of $(\mathbb{Z}^d, h)$. Although h is not geodesic, the passage time metric T is. Given suitably strong concentration results, one can fix a realization of edge weights and hope to approximate the passage times between vertices in a geodesic by the expected passage time between them, up to a smaller order correction. Breaking this geodesic into a skeleton of m roughly equally spaced vertices would then furnish the sequence (x_i) in Definition 3.24. The second condition of Definition 3.24 guarantees that there are no large “jumps” in $B(x,r)$ as one increases r, which is also natural to expect in FPP.

The SAG condition guarantees that the metric d obeys a shape theorem with an explicit rate of convergence estimate, which we now describe. We say that an increasing function $\phi : [0,\infty) \to (0,\infty)$ is sublinearly doubling if there exists a function $\eta : [0,\infty) \to [0,\infty)$ such that $\lim_{\lambda\to\infty} \eta(\lambda)/\lambda = 0$ and for all $\lambda > 0$, $\phi(\lambda r) \leq \eta(\lambda)\phi(r)$. The following theorem gives an equivalence between asymptotic

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

geodesicity in the form of SAG(N) and a shape theorem with error term on the order N^{-1}.

THEOREM 3.25 (Tessera [**167**]). *Let δ be some translation-invariant metric on $\mathbb{Z}^d$ and $\phi : [0,\infty) \to [1,\infty)$ be an increasing, sublinearly doubling function. The following are equivalent.*

(1) *There exists $c > 0$ such that δ is SAG(N) with*
$$N(\alpha) \geq c\phi(\alpha),$$
for α large enough.

(2) *There exists a norm $\|\cdot\|$ on $\mathbb{R}^d$ and $C > 0$ such that for all large enough n,*
$$B_{\|\cdot\|}\left(0, n - \frac{Cn}{\phi(n)}\right) \cap \mathbb{Z}^d \subset B_\delta(0,n) \subset B_{\|\cdot\|}\left(0, n + \frac{Cn}{\phi(n)}\right).$$

A version of SAG with exponent $1/2$ was shown for FPP in [**167**]. Namely, if the edge weights have finite exponential moments and $F(0) < p_c$, then h satisfies SAG(N) for
$$N(\alpha) = c(\alpha/(\log \alpha))^{1/2},$$
for some $c > 0$. Therefore there exist $C > 0$ and $n_0 \geq 1$ such that for any $n \geq n_0$,
$$(n - C(n\log n)^{1/2})\mathcal{B} \cap \mathbb{Z}^d \subseteq \bar{B}(0,n) \subseteq (n + C(n\log n)^{1/2})\mathcal{B}, \tag{3.38}$$
where $\bar{B}(0,n) := \{x \in \mathbb{Z}^d : h(x) \leq n\}$.

Let us conclude this subsection by giving some account of why Theorem 3.25 should hold; we will freely make simplifying assumptions. We work in the metric δ. Let m and n be arbitrary positive integers, and recall that the convergence in Theorem 3.25 is convergence under the Hausdorff metric d_H, where for two compact sets A and B
$$d_H(A,B) = \inf\{r > 0 : A \subseteq [B]_r,\ B \subseteq [A]_r\}.$$
We will try to show that $d_H(n^{-1}B(0,n),\ (mn)^{-1}B(0,mn))$ is small; taking m to infinity gives a corresponding bound for $d_H(n^{-1}B(0,n),\ B_{\|\cdot\|}(0,n))$.

A natural way to use the SAG assumption is to try to build the ball $B(0,mn)$ out of m copies of the ball $B(0,n)$. This is because, given x at distance mn from 0, we can find a skeleton sequence of vertices (x_i) as in Definition 3.24 each at distance $\sim n$ from each other. This reasoning may be made rigorous as in the following lemma, adapted from [**167**, Lemma 5.1]:

LEMMA 3.26. *There exists a C such that, for all m and n positive integers,*
$$d_H\left((mn)^{-1}B(0,n)^m,\ (mn)^{-1}B(0,mn)\right) \leq C/\phi(n).$$

In the above, $A + B = \{a + b : a \in A,\ b \in B\}$ and $A^2 = A + A$, etc. Using the fact that d_H is a metric, we see
$$d_H\left(\frac{B(0,n)}{n}, \frac{B(0,mn)}{mn}\right) \leq d_H\left(\frac{B(0,n)}{n}, \frac{B(0,n)^m}{mn}\right) + d_H\left(\frac{B(0,n)^m}{mn}, \frac{B(0,mn)}{mn}\right),$$
and the second term on the right-hand side is small by Lemma 3.26.

In other words, we see that $B(0,mn)$ is approximately equal to m copies of $B(0,n)$ strung together, and it remains to show that $mB(0,n)$ is also approximately

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

equal to this set. Note that if $B(0,n)$ were a convex subset of $\mathbb{R}^n$, then $B(0,n)^m = mB(0,n)$, and we would be done. So the only way this line of attack could fail would be if $B(0,n)$ were irregular enough to be significantly different from its convex hull.

This complication is dealt with in [**167**] by inductively controlling the Hausdorff distance between balls $B(0,n)$ and their convex hulls $\widehat{B}(0,n)$. A crucial fact used is that in $\mathbb{R}^d$, Carathéodory's Theorem for convex hulls guarantees that $d_H(B(0,n)^m, \widehat{B}(0,n)^m)$ can be bounded by a uniform constant multiple of $d_H(B(0,n), \widehat{B}(0,n))$.

3.5.3. Alexander's method . The first upper bounds on nonrandom fluctuations go back to the work of Kesten [**125**, Theorem 1], who established the bound

$$\mathbb{E}T(0,ne_1) \le n\mu(e_1) + Cn^{5/6}\log n.$$

The right side has since been improved to $n\mu(e_1) + C(n\log n)^{1/2}$; see (3.38) and (3.44). An extraordinary contribution was the theory developed by Alexander in [**7**]. It concerns the rate of convergence of a (deterministic) nonnegative function h defined on the lattice $\mathbb{Z}^d$ to its limit g. That is, putting $h(x) = \mathbb{E}T(0,x)$ as in the last section, we assume

$$h(x+y) \le h(x) + h(y), \quad x, y \in \mathbb{Z}^d,$$

and set

$$g(x) := \lim_{n\to\infty} h(nx)/n.$$

By subadditivity, $g(x) \le h(x)$ for all $x \in \mathbb{Z}^d$. Alexander's theory has a very geometric flavor and explores the fact that the function h is defined on the whole space $\mathbb{Z}^d$; h is not only a (one-dimensional) subadditive sequence. In his point of view, one way to guarantee an upper bound on the rate of convergence of $h(nx)/n$ is to request h to have "good skeleton paths" towards nx in $\mathbb{Z}^d$. It turns out that this condition is possible to verify in FPP and other models and suitable to bound the rate of convergence.

To describe his approach, let $0 = v_0, v_1, v_2, \ldots, v_k = x$ be points in $\mathbb{Z}^d$. We will soon think of the $v_i's$ as the skeleton of a path from 0 to x. From subadditivity, we have

$$h(x) = h\bigg(\sum_{i=1}^k (v_i - v_{i-1})\bigg) \le \sum_{i=1}^k h\bigg(v_i - v_{i-1}\bigg).$$

Imagine for a moment that we can choose these v_i's in such a way that

$$h(v_i - v_{i-1}) = g(v_i - v_{i-1}) + o(|x|^a), \tag{3.39}$$

for some $a < 1$. This would lead to

$$h(x) - g(x) \le \sum_{i=1}^k \big(g(v_i - v_{i-1})\big) - g(x) + ko(|x|^a). \tag{3.40}$$

Unfortunately, (3.40) is not enough for our purpose, as we still need to control the difference $\sum_{i=1}^k \big(g(v_i - v_{i-1})\big) - g(x)$, even if k is a small power of $|x|$. Moreover, it is not clear how to find v_i's satisfying (3.39).

To overcome these problems, we will make two modifications. First, instead of requesting (3.39), we will replace g by the projection g_x of g on the line that goes through 0 and x. This projection is linear and satisfies $g_x(nx) = g(nx)$ for any $n \in \mathbb{N}$. In Alexander's description, the value $g_x(y)$ is the amount of progress toward

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

x made by a vector increment of y. The difference $h(y) - g_x(y)$ should be thought as the error or the inefficiency associated with such an increment. An increment y will be called good if y satisfies $h(y) \le g_x(y) + C|x|^a$ for a given $C > 0$.

The change from g to g_x helps us to get rid of the sum in the right side of (3.40). However, we still need to find the skeleton v_i's of good increments. The main idea and contribution of [**7**] is to come up with an assumption on h that allows us to roughly do so. It will suffice to verify the weaker condition that x (divided by some constant larger than one) is in the convex hull of good increments. This weaker condition can be verified in FPP and many other models. We will make this statement precise in the next definition.

Let Φ denote the set of all positive nondecreasing functions on $(1, \infty)$ and fix $a < 1$. For $\phi \in \Phi$, define $Q_x = Q_x(a, \phi, C, K)$ to be the set of good increments with exponent a, correction factor ϕ in direction x, that is,

$$Q_x := \left\{ y \in \mathbb{Z}^d : |y| \le K|x|,\ h(y) \le g_x(y) + C|x|^a \phi(|x|),\ g_x(y) \le g(x) \right\}.$$

The condition $g_x(y) \le g(x)$ is to avoid "overshooting" x, and the condition $|y| \le K|x|$ ensures Q_x is not too large. For a collection of points in $\mathbb{R}^d$, let $\mathrm{Co}(\cdot)$ denote the convex hull.

Definition 3.27 (CHAP). The function h satisfies the convex hull approximation property (or CHAP) with exponent a and correction ϕ (and constants C, K) if there exists $L > 1$ such that

$$x/\alpha \in \mathrm{Co}(Q_x(a, \phi, C, K)) \text{ for some } \alpha \in [1, L],$$

for all $x \in \mathbb{Q}^d$ with $|x|$ sufficiently large.

We will soon discuss how one would verify that a subadditive function h satisfies CHAP and why it is natural to work with Definition 3.27. The next definition is exactly the control on the rate of convergence of h.

Definition 3.28 (GAP). For $a > 0$ and $\phi \in \Phi$ we say that h satisfies the general approximation property (GAP) with exponent a and correction factor ϕ if there exist $M > 1$ and $C > 0$ such that for all $x \in \mathbb{Z}^d$ and $|x| \ge M$,

$$g(x) \le h(x) \le g(x) + C|x|^a \phi(|x|).$$

Alexander's main result is the following.

Theorem 3.29 (Alexander [**7**], Theorem 1.8). *Suppose h is a nonnegative subadditive function on $\mathbb{Z}^d$. If h satisfies CHAP with exponent $a > 0$ and correction ϕ, then it satisfies GAP with exponent a and correction ϕ.*

We will now sketch the proof of Theorem 3.29. The following fixed point argument is the core of the proof.

Step 1. Suppose h satisfies $\mathrm{CHAP}(a, \phi)$ and $\mathrm{GAP}(b, \phi)$ with $b \in (a, 1]$. Then h satisfies $\mathrm{GAP}(b', \phi)$ with $b' = b/(1 + b - a) < b$.

Sketch of the proof of Step 1. Suppose $|x|$ is large. Take $q \in \mathbb{Q}$ also large but smaller than $|x|$. Applying $\mathrm{CHAP}(a, \phi)$ to x/q we can express $x/(\alpha q)$ as a convex combination of $d + 1$ good increments $y_{q,i}$, $1 \le i \le d + 1$. Now, use the increments

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

$y_{q,i}$ to decompose

$$x = \sum_{i=1}^{d+1} \alpha_{q,i} y_{q,i} + x_{(r)}$$

with $\alpha_{q,i} \in \mathbb{N} \cup \{0\}$ and some remainder $x_{(r)}$. By subadditivity and the fact that $g(x) = g_x(x)$, we have

$$\begin{aligned} h(x) - g(x) &\leq \sum_{i=1}^{d+1} \alpha_{q,i} h(y_{q,i}) + h(x_{(r)}) - g_x(x - x_{(r)}) - g_x(x_{(r)}) \\ &= \sum_{i=1}^{d+1} \alpha_{q,i} \big(h(y_{q,i}) - g_x(y_{q,i})\big) + h(x_{(r)}) - g_x(x_{(r)}). \end{aligned} \tag{3.41}$$

Our choice of the $y_{q,i}$'s implies that the sum on the right side of (3.41) is bounded by

$$C^{d+1} q^{1-a} |x|^a \phi(|x|),$$

while the remainder term $h(x_{(r)}) - g_x(x_{(r)})$ is controlled using $\mathrm{GAP}(b, \phi)$. Optimizing over q gives the desired result. (The optimal q turns out to be of the order of a small power of $|x|$.) The details can be found in [**7**, Proposition 2.1].

STEP 2. By hypothesis, h satisfies $\mathrm{GAP}(1, \phi)$ with $\phi = 1$. If $a \geq 1$, there is nothing to prove. If $a < 1$, it is possible to iterate STEP 1 to obtain $\mathrm{GAP}(a, \phi)$ as the fixed point of the function $x \to x/(1 + x - a)$ is at a. (It must be shown that the constants C, K, M do not change after each iteration.)

Let us now end this subsection by saying a few words about why it is natural to expect CHAP and how one proves $\mathrm{CHAP}(a, \phi)$ in FPP. The first ingredient is the following geometric lemma that defines the skeleton $\{v_i\}$ discussed before. The proof is short, so we include it here.

LEMMA 3.30 (Lemma 1.6, [**7**]). *Let h be a nonnegative subadditive function on $\mathbb{Z}^d$ and let $c > 1$. Suppose that for each $x \in \mathbb{Q}^d$, there exist $n \geq 1$, a lattice path Γ from 0 to nx and a sequence of sites $0 = v_0, v_1, \ldots, v_m = nx$ in Γ such that $m \leq cn$ and $v_i - v_{i-1} \in Q_x(a, \phi, C, K)$ for all $1 \leq i \leq m$. Then h satisfies CHAP(a, ϕ).*

PROOF. We need to find $\alpha \geq 1$ so that

$$x/\alpha \in \mathrm{Co}(Q_x).$$

Let $a(n, y)$ be the number of indices i such that $v_i - v_{i-1} = y$. Then, by hypothesis,

$$nx = \sum_{y \in Q_x} a(n, y) y$$

and applying g_x to the equation above

$$ng(x) = \sum_{y \in Q_x} a(n, y) g_x(y) \leq g(x) \sum_{y \in Q_x} a(n, y)$$

so

$$n \leq \sum_{y \in Q_x} a(n, y).$$

On the other hand, by hypothesis,

$$n^{-1} \sum_{y \in Q_x} a(n, y) \leq c.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Taking $\alpha = n^{-1}\sum_{y\in Q_x} a(n,y)$ and $L = c$ gives us CHAP. □

Now consider the $m+1$ sites in the lemma above. One can inductively find a sequence of marked sites for any path Γ from 0 to nx. One starts with $v_0 = 0$ and chooses z_i as the first site in Γ after v_{i-1} such that $z_i - v_{i-1} \notin Q_x$. Now take v_i as the last vertex visited by Γ before z_i. We call the sequence of marked sites, obtained from a self-avoiding path in this way, the Q_x-skeleton of Γ. The difficulty is to control the number of vertices in a Q_x-skeleton. This is the role of the next proposition, which combined with Lemma 3.30 ends the proof of CHAP$(1/2, \log|x|)$.

PROPOSITION 3.31. *Assume $F(0) < p_c$ and $\mathbb{E}e^{\alpha\tau_e} < \infty$ for some $\alpha > 0$. Then there exists $\mathfrak{M} > 0$ so that if $|x| > \mathfrak{M}$ then for sufficiently large n there exists a lattice path from 0 to nx with $Q_x(1/2, \log|x|)$-skeleton of $2n+1$ or fewer vertices.*

PROOF SKETCH. We will take one geodesic from 0 to nx as our lattice path. To prove Proposition 3.31, Alexander makes use of an estimate of Kesten [**123**, (4.13)] and his skeleton construction. The argument goes as follows, where here the constant C will change from line to line.

Consider a sequence of points (v_i) in $\mathbb{Z}^d$ so that the increments $v_i - v_{i-1}$ are the Q_x-skeleton of some lattice path Γ. Let

$$Y_i = \mathbb{E}T(v_{i-1}, v_i) - T(v_{i-1}, v_i).$$

The hypothesis $\mathbb{E}e^{\lambda\tau_e} < \infty$ is sufficient (see Theorem 3.12) to find a constant $C > 0$ so that

$$\mathbb{P}\Big(|Y_i| > \lambda|v_i - v_{i-1}|^{1/2}\Big) \le (1/C)\exp\Big(-C\lambda\Big).$$

In particular, this combined with the fact that $|v_i - v_{i-1}| \le 2d|x| =: K|x|$ implies the existence of a $\beta > 0$ so that $\mathbb{E}\exp(\beta Y_i/|x|^{1/2}))$ is bounded. Now, imagine that the set of edges used by paths from v_i to v_{i+1} were disjoint and deterministic. Then the Y_i's would be independent. Moreover, taking exponentials and using Markov's inequality we obtain

$$\mathbb{P}\bigg(\sum_{i=0}^{m-1} Y_i \ge Cm|x|^{1/2}\log|x|\bigg) \le (1/C)\exp(-Cm\log|x|). \tag{3.42}$$

The fact that we assumed the use of disjoint edges and independence (illegally) can be handled by taking independent copies Y_i' of Y_i and using two extensions of the BK inequality (this is done in Kesten [**123**, eq. (4.13)]); see also Theorem 3.39 below. On the other hand, by geometric considerations, there are at most $C|x|^{md}$ ways to choose $v_i, i = 1, \ldots, m$, with $|v_i - v_{i-1}| \le 2d|x|$. Thus, a union bound implies

$$\begin{aligned}\mathbb{P}\bigg(\sum_{i=0}^{m-1} Y_i \ge Cm|x|^{1/2}\log|x| \text{ for some } m \ge 1 \text{ and some } Q_x\\ \text{-skeleton with } m \text{ vertices}\bigg) \le (1/C)\exp(-Cm\log|x|).\end{aligned} \tag{3.43}$$

Now take a Q_x-skeleton with m vertices of a geodesic from 0 to nx. By definition of Q_x, it follows that $n \le m$. The estimate (3.43) above implies that we can choose

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

n large enough so that with probability going to one

$$\sum_{i=0}^{m-1} \mathbb{E}T(v_i, v_{i+1}) - T(0, nx) = \sum_{i=0}^{m-1} Y_i \leq Cm|x|^{1/2} \log |x|.$$

Hence, by the convergence to the time constant, with high probability when n is large,

$$\sum_{i=0}^{m-1} \mathbb{E}T(v_i, v_{i+1}) \leq n\mu(x) + Cm|x|^{1/2} \log |x|.$$

Now the proof ends by arguing as follows. One splits the set of increments $(v_{i+1} - v_i)$ into two classes: short and long. The short increments are those that are in the skeleton because the vertex z_{i+1} has the property that

$$h(z_{i+1} - v_i) \geq g_x(z_{i+1} - v_i) + C'|x|^{1/2} \log |x|,$$

where C' is the constant C from the definition of Q_x. The long increments are those that are in the skeleton because either $|z_{i+1} - v_i| \geq K|x|$ or $g_x(z_{i+1} - v_i) \geq g(x)$. The short increments can be shown (see [**7**, eq. (3.16)]) to comprise at most $m/4$ of the terms of the sum on the left, so long as C' is chosen large enough relative to the constant C appearing on the right above. The reason is that if $|x|$ is large, each short increment will contribute at least $g_x(v_{i+1} - v_i) + (C'/2)|x|^{1/2} \log |x|$ to the sum on the left, whereas all increments contribute at least $g_x(v_{i+1} - v_i)$. After observing that

$$\sum_{i=1}^{m-1} g_x(v_{i+1} - v_i) = \mu(nx) = n\mu(x),$$

this leaves

$$|\text{short increments}|(C'/2)|x|^{1/2} \log |x| \leq Cm|x|^{1/2} \log |x|,$$

giving that the number of short increments is at most $m/4$ if C' is large. As for the long increments, there must be at least $3m/4$ of them, and each can be shown to contribute at least $5g(x)/6$ to the sum. One can therefore conclude that there are at most $6n/5 + m/8$ long increments (see the top of p. 48 of [**7**]). Putting these together with the appropriate constants leads to $m \leq 2n$ for n large and $|x| \geq \mathfrak{M}$. □

We finish this section with a summary of the current state of the art on (a) nonrandom fluctuation upper bounds and (b) the convergence rate to the limit shape. For (a), Alexander's methods were used in the low moment case in [**70**] along with a concentration inequality for the lower tail of $T(0, x)$ to obtain the following.

THEOREM 3.32. *Assume $F(0) < p_c$ and $\mathbb{E}Y^2 < \infty$, where Y is the minimum of d independent copies of τ_e. There exists $C = C(F, d)$, such that for all $x \in \mathbb{Z}^d$ with $|x| > 1$,*

$$\mu(x) \leq \mathbb{E}T(0, x) \leq \mu(x) + C(|x| \log |x|)^{1/2}. \tag{3.44}$$

For (b) above, the following version of convergence rate bounds under low moments also appears in [**70**].

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

THEOREM 3.33. *Assume $F(0) < p_c$. If $\mathbb{E}Y^2 < \infty$, where Y is the minimum of d independent copies of τ_e, then there exists $C > 0$ such that a.s.,*

$$B(t) \subset (t + C(t \log t)^{1/2})\mathcal{B} \text{ for all large } t.$$

If in addition $\mathbb{E}\tau_e^\alpha < \infty$ for some $\alpha > 1 + 1/d$, then a.s.

$$(t - C(\log t)^4 t^{1/2})\mathcal{B} \subset B(t) \text{ for all large } t.$$

The inner bound given here can be improved under finite exponential moments, as mentioned in (3.38).

3.6. Large deviations

In this section we discuss some large deviation bounds for FPP. For a given $\epsilon > 0$, our first goal is to estimate the following probabilities

$$\begin{aligned} p_n^u(\epsilon) &:= \mathbb{P}\Big(T(0, ne_1) > n(\mu(e_1) + \epsilon)\Big) \\ p_n^\ell(\epsilon) &:= \mathbb{P}\Big(T(0, ne_1) < n(\mu(e_1) - \epsilon)\Big), \end{aligned} \tag{3.45}$$

for n large. If we assume (2.1), Theorem 2.1 implies that both $p_n^u(\epsilon)$ and $p_n^\ell(\epsilon)$ go to zero as $n \to \infty$. Stronger assumptions will allow us to derive exponentially small upper bounds for these probabilities.

Typically, these bounds will exhibit a strong asymmetry: the behavior of p_n^u is radically different from that of p_n^ℓ. Indeed, if we want to increase the passage time $T(0, ne_1)$, we may need to increase the passage time of every path from 0 to ne_1. If the passage times are bounded, this roughly amounts to increasing almost every edge in a box with diameter of order n. Thus one should expect:

$$p_n^u(\epsilon) \approx \exp(-n^d I_u(\epsilon)). \tag{3.46}$$

On the other hand, to decrease the passage time between two points, it suffices to force the passage time of a single path to be very low. Thus,

$$p_n^l(\epsilon) \approx \exp(-n I_\ell(\epsilon)), \tag{3.47}$$

as we should roughly decrease the passage time of order n edges.

The reader may have seen this asymmetry in large deviations in other areas of probability. In random matrix theory, for instance, the smallest eigenvalue of a $n \times n$ GOE matrix satisfies a large deviation principle of speed n if forced out of the bulk; on the other hand, the probability that the smallest eigenvalue is above the bottom edge of the equilibrium measure is of order $\exp(-cn^2)$. The intuitive reason for this asymmetry is that in the latter case, the smallest eigenvalue has to push other eigenvalues inside the bulk; in the former, it is free to move away by itself.

Before stating the results, let us make a remark. The boundedness assumption is important for (3.46). For example, suppose that τ_e is exponentially distributed with mean 1. We know that if all $2d$ edges connected to the origin take values larger than $n(\mu(e_1) + \epsilon)$, then $T(0, ne_1) \geq n(\mu(e_1) + \epsilon)$ and

$$p_n^u(\epsilon) \geq \mathbb{P}\Big(\tau_e \geq n(\mu(e_1) + \epsilon)\Big)^{2d} = \exp\Big(-2nd(\mu(e_1) + \epsilon)\Big),$$

so (3.46) does not hold.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Let Y denote the minimum of $2d$ independent random variables distributed as τ_e. The example above illustrates that the behavior of p_n^u is intimately related to the tail of the distribution of Y. In fact, as in Lemma 2.3, one can relate the moments of Y to summability of probabilities of large deviations away from the time constant. This is summarized in the theorem below, taken from Ahlberg [**3**].

THEOREM 3.34. *For every $\alpha > 0$, $\epsilon > 0$ and $d \geq 2$,*

$$\mathbb{E}\big[Y^\alpha\big] < \infty \iff \sum_{z \in \mathbb{Z}^d} \|z\|^{\alpha-d}\mathbb{P}\big(\big|T(0,z) - \mu(z)\big| > \epsilon\|z\|\big) < \infty.$$

We move to large deviation lower bounds and we will assume that $F(0) < p_c$ so that $\mu(e_1) > 0$ by Theorem 2.5. Let

$$\beta = \sup\left\{x : F(\mu(e_1) - x) > 0\right\}.$$

THEOREM 3.35 (Kesten [**123**],Theorem 5.2). *For any $\epsilon > 0$, there exist a positive constant $A(\epsilon)$ and $I_\ell(\epsilon) \in (0, \infty]$ such that for any $n \geq 1$*

$$p_n^\ell(\epsilon) \leq A(\epsilon)\exp(-nI_\ell(\epsilon)).$$

Furthermore,

$$\lim_n \frac{1}{n} \log p_n^\ell(\epsilon) = -I_\ell(\epsilon), \tag{3.48}$$

where $0 < I_\ell(\epsilon) < \infty$ for any $0 < \epsilon < \beta$ and $I_\ell(\epsilon) = \infty$ for $\beta \leq \epsilon$. Setting $I_\ell(\epsilon) = 0$ for $\epsilon \leq 0$, the function $I_\ell : (-\infty, \beta) \to \mathbb{R}$ is convex on its domain and strictly increasing on $[0, \beta)$.

REMARK 3.36. Theorem 3.35 requires no assumptions on the moments of the edge-weight distribution. In particular, it holds even when (2.10) is not satisfied and $\mu(e_1)$ is only defined through convergence in probability. In this case, $\mu(e_1)$ in the definition of $p_n^\ell(\epsilon)$ is replaced by the shell time constant $\bar{\mu}(e_1) := \lim_n \hat{T}(0, ne_1)/n$ from Section 2.4.

REMARK 3.37. The function $I_\ell(\epsilon)$ should depend on d and on the underlying edge-weight distribution F. As far as we know, there is no explicit expression for I_ℓ as a function of F.

As a sketch of the proof of Theorem 3.35, we will now explain how to obtain bounds of the form

$$A_1(\epsilon)\exp(-nJ_1(\epsilon)) \leq p_n^\ell(\epsilon) \leq A_2(\epsilon)\exp(-nJ_2(\epsilon)),$$

for positive functions A_1, A_2, J_1, J_2, when $d = 2$ and in the simpler case that the distribution of the passage times satisfies

$$\mathbb{P}(\tau_e = 1) = 1/2 = \mathbb{P}(\tau_e = 2).$$

Here, we know that by Theorem 2.12,

$$1 < \mu(e_1) < 2.$$

Let ϵ be such that $0 < \epsilon < \mu(e_1) - 1$. By taking the straight path from 0 to ne_1 we see that for n sufficiently large with probability at least $e^{-n \log 2}$ it is possible to construct a configuration such that $T(0, ne_1) < n(\mu(e_1) - \epsilon)$. Thus we obtain the bound

$$p_n^\ell(\epsilon) \geq A_1(\epsilon)e^{-nJ_1(\epsilon)},$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

for some $A_1, J_1 > 0$.

To obtain an upper bound on p_n^ℓ we will estimate the passage time from below. The idea is as follows. We extract from a geodesic $\Gamma(0, ne_1)$ subpaths Γ_i, $i = 1, \ldots, \alpha n$, $\alpha \in (0,1)$, that are edge-disjoint. Writing $T_i = T(\Gamma_i)$, we obtain

$$T(0, ne_1) \geq \sum_{i=1}^{\lfloor \alpha n \rfloor} T_i$$

so that for any $t > 0$,

$$\mathbb{P}\left(T(0, ne_1) < t\right) \leq \mathbb{P}\left(\sum_{i=1}^{\lfloor \alpha n \rfloor} T_i < t\right).$$

Our hope is that this construction leads to a family (T_i) of nearly independent random variables. If so, we could use classic deviation bounds for sum of independent variables. Fortunately, life is not so simple and we get to do some extra work. To start, instead of just considering the geodesic $\Gamma(0, ne_1)$, we consider each path from 0 to ne_1.

The main tool for this construction is a version of the van den Berg-Kesten (BK) inequality. This inequality allow us to replace the T_i's by independent copies with nice asymptotic bounds. We visit this inequality in the next section and complete the sketch of the proof of Theorem 3.35.

3.6.1. The BK-Reimer and Harris-FKG inequalities. Consider a finite probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$, $\mathcal{F}$ is the collection of all subsets of Ω, and $\mathbb{P}$ is a product of n probability measures

$$\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2 \times \ldots \times \mathbb{P}_n.$$

Given a set $S \subseteq \{1, \ldots, n\}$, we say that two configurations $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ and $\omega' = (\omega_1', \ldots, \omega_n') \in \Omega$ are equal on S if $\omega_i = \omega_i'$ for all $i \in S$. We denote this by $\omega \stackrel{S}{=} \omega'$.

We say that an event $A \in \mathcal{F}$ occurs on the set S in the configuration ω if A occurs using only the random variables over S; that is, if A occurs independently of the values ω_i, $i \in S^c$. Let $A|_S$ be the collection of such ω:

$$A|_S = \{\omega \in A : \omega' \stackrel{S}{=} \omega \text{ implies } \omega' \in A\}.$$

Two events $A_1, A_2 \in \mathcal{F}$ are said to occur disjointly, denoted by $A_1 \circ A_2$, if there are two disjoint sets on which they occur:

$$A_1 \circ A_2 = \{\omega : \exists\, S_1, S_2 \subset \{1, \ldots, n\}, S_1 \cap S_2 = \emptyset,\ \omega \in A_1|_{S_1} \cap A_2|_{S_2}\}. \tag{3.49}$$

The BK-Reimer inequality states that

$$\mathbb{P}(A \circ B) \leq \mathbb{P}(A)\mathbb{P}(B). \tag{3.50}$$

The inequality was conjectured (and proved in the increasing event case) by van den Berg-Kesten in [**30**] and proved by Reimer [**152**] after important simplifications by van den Berg-Fiebig [**29**] and Fishburn-Shepp [**87**].

Although it is not necessarily true that $(A \circ B) \circ C = A \circ (B \circ C)$, using (3.49) for the definition of $A_1 \circ \ldots \circ A_k$, one obtains the extension of (3.50) as

$$\mathbb{P}(A_1 \circ A_2 \circ \ldots \circ A_k) \leq \prod_{\ell=1}^{k} \mathbb{P}(A_\ell). \tag{3.51}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Inequality (3.51) can be improved if we assume that we are dealing with increasing (or decreasing) events. Recall that an event A is called increasing if the indicator function $\mathbf{1}_A(\omega)$ is an increasing function of each $\omega_i, i = 1, \ldots, n$. If A and B are increasing events then the Harris-FKG inequality states [**39**, Theorem 2.15]

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A)\mathbb{P}(B). \tag{3.52}$$

We will use a version of inequality (3.51) in the following form.

LEMMA 3.38. *[Kesten [**123**], Theorem 4.8] Let $\{A(k,i) : k, i \geq 1\}$ be a family of increasing (or decreasing) events and let $\{A'(k,i) : k, i \geq 1\}$ be a another family of events on some probability space $(\Omega', \mathbb{P}')$ such that*

(1) *For each fixed i, the joint distribution of $\{A(k,i) : k \geq 1\}$ under $\mathbb{P}$ is the same as the joint distribution of $\{A'(k,i) : k \geq 1\}$ under $\mathbb{P}'$.*
(2) *The families $\{A'(k,i) : k \geq 1\}$, $i = 1, 2, \ldots$ are independent under $\mathbb{P}'$.*

Then, for arbitrary $n(k) < \infty$ one has

$$\mathbb{P}\left(\bigcup_{k\geq 1} A(k,1) \circ A(k,2) \circ \ldots \circ A(k, n(k))\right) \leq \mathbb{P}'\left(\bigcup_{k\geq 1} \bigcap_{i=1}^{n(k)} A'(k,i)\right). \tag{3.53}$$

Note that in the case where $A(k,i) = \emptyset$ for all $k \geq 2$, equation (3.53) simplifies to (3.51). We will use the lemma above in the following form to study the passage time from 0 to x. Given sets of vertices of $\mathbb{Z}^d$, $V(l,j)$, $l \geq 1, 1 \leq j \leq n(l)+1$, with $V(l,0) = \{0\}$ and $V(l, n(l)+1) = \{x\}$, define

$$\Pi_l(0,x,t) := \Big\{\Gamma : \Gamma \text{ is a path from 0 to } x \text{ that passes successively} \\ \text{through } V(l,1), \ldots, V(l,n(l)) \text{ and has } T(\Gamma) < t\Big\}.$$

LEMMA 3.39 (eq. (4.13) in [**123**]). *For any $t > 0$,*

$$\mathbb{P}\Big(\bigcup_{l\geq 1} \Pi_l(0,x,t) \neq \emptyset\Big) \leq \sum_{l\geq 1} \mathbb{P}\Big(\sum_{j=0}^{n(l)} T'(l,j) < t\Big), \tag{3.54}$$

where $T'(l,j)$ are all independent with $T'(l,j) \stackrel{d}{=} T(V(l,j), V(l,j+1))$.

PROOF. Write the left side of (3.54) as

$$\mathbb{P}\left(\bigcup_{l\geq 1} \bigcup_{m\geq 1} A(m,l,1) \circ A(m,l,2) \circ \ldots \circ A(m,l,n(d))\right) \tag{3.55}$$

where

$$A(m,l,i) = \{T(V(l,i), V(l,i+1)) < t(m,l,i)\}$$

and the union over m runs over all choices of $t(m,l,i)$ as rational numbers which satisfy

$$\sum_{i=1}^{n(i)} t(m,l,i) < t.$$

The representation (3.55) follows from the fact that any path Γ from 0 to x through $V(l,1), \ldots, V(l,n(l))$ can be decomposed into edge disjoint pieces Γ_i from $V(l,i)$ to

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

$V(l, i+1)$ with $T(\Gamma) = \sum T(\Gamma_i)$. The inequality in (3.54) is now just an application of Lemma 3.38 for each fixed l. $\square$

Lemma 3.39 was used by Kesten to obtain his large deviation bounds. Here we apply it to obtain a bound on deviations in all directions simultaneously, and not only for coordinate directions. This comes with no cost in the presentation. We assume $\mu \neq 0$. Consider the ball of radius t in the metric μ,

$$\mathcal{B}(t) = \{z \in \mathbb{Z}^d : \mu(z) \leq t\},$$

and its complement $\mathcal{B}(t)^c = \{z \in \mathbb{Z}^d : \mu(z) > t\}$. A version of the bound below first appeared in [**3**].

LEMMA 3.40. *Let $X^{(q)}_{s,s+m}$ for $q = 1, 2, \ldots$ denote independent random variables distributed as $T(\mathcal{B}(s), \mathcal{B}(s+m)^c)$. There exists $C < \infty$ such that for every $n \geq m \geq s \geq 1$ and $t > 0$ we have*

$$\mathbb{P}\big(T\big(0, \mathcal{B}(n)^c\big) < t\big) \leq \sum_{Q \geq n/(m+Cs)-1} n^{d-1} \left(C\frac{m}{s}\right)^{d(Q-1)} \mathbb{P}\left(\sum_{q=1}^{Q} X^{(q)}_{s,s+m} < t\right).$$

PROOF. Fix s, m so that $1 \leq s \leq m \leq n$. Pick $z \in \mathbb{Z}^d$ such that $\mu(z) > n$. Let $\Gamma = \Gamma(z)$ be a self-avoiding path connecting the origin to z. We will choose a subsequence $v_0, v_1, \ldots, v_Q$ of the vertices in Γ as follows. Set $v_0 = 0$. Given v_q, let v_{q+1} be the first vertex in Γ succeeding v_q such that

$$\mu(v_{q+1} - v_q) > m + 2s.$$

When no such vertex exists, stop and set $Q = q$. To find a lower bound on Q, note that by the triangle inequality

$$n < \mu(z) \leq \mu(z - v_Q) + \mu(v_Q) \leq \mu(z - v_Q) + \sum_{q=0}^{Q-1} \mu(v_{q+1} - v_q). \tag{3.56}$$

By convexity and our choice of v_q's, we also know that

$$\mu(v_{q+1} - v_q) \leq m + 2s + \mu(e_1)$$

and

$$\mu(z - v_Q) \leq m + 2s.$$

Thus we obtain from (3.56),

$$n \leq (Q+1)\big(m + 2s + \mu(e_1)\big). \tag{3.57}$$

Now, pick $r > 0$ such that $[-r, r]^d \subseteq \mathcal{B}$ and tile $\mathbb{Z}^d$ with copies of the box $(-rs, rs]^d$ such that each box is centered at a point in $\mathbb{Z}^d$, and each point in $\mathbb{Z}^d$ is contained in precisely one box. Let Λ_q denote the box that contains v_q, and let w_q denote the center of Λ_q.

Choose the tiling in such a way that $w_0 = v_0 = 0$. Denote by Γ_q the part of the path Γ that connects v_q and v_{q+1}. As Γ is self-avoiding, for $q_1 \neq q_2$, the two pieces Γ_{q_1} and Γ_{q_2} are edge-disjoint. By construction, v_q is contained in the copy of $\mathcal{B}(s)$ centered at w_q, while v_{q+1} is not contained in the copy of $\mathcal{B}(s+m)$ centered at w_q. That is,

$$\mu(v_q - w_q) \leq s \quad \text{and} \quad \mu(v_{q+1} - w_q) > s + m. \tag{3.58}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Moreover, the points $w_0, w_1, \dots, w_{Q-1}$ have to satisfy

$$\mu(w_{q+1} - w_q) \leq m + 4s + \mu(e_1). \tag{3.59}$$

Let W_Q denote the set of all sequences $(w_0, w_1, \dots, w_{Q-1})$ such that $w_0 = 0$, each w_q is the center of some box Λ_q, and w_q and w_{q+1} satisfies (3.59) for each $q = 0, 1, \dots, Q-2$.

Given $x > 0$, Q and $w = (w_0, w_1, \dots, w_{Q-1}) \in W_Q$, let $A(t, w)$ denote the event that there exists a path Γ from the origin to z with edge-disjoint segments $\Gamma_0, \Gamma_1, \dots, \Gamma_{Q-1}$ such that:

(1) $\sum_{q=0}^{Q-1} T(\Gamma_q) < t$,

(2) the endpoints v_q and v_{q+1} of Γ_q satisfy (3.58), for each $q = 0, 1, \dots, Q-1$.

Since $T(\Gamma) \geq \sum_{q=0}^{Q-1} T(\Gamma_q)$, together with (3.57), we obtain that

$$\{T(0, z) < t\} \subseteq \bigcup_{Q \geq n/(m+bs)-1} \bigcup_{w \in W_Q} A(t, w), \tag{3.60}$$

where $b = 2 + \mu(e_1)$. Note that given w_q, the passage time of any path between two vertices v and v' such that $\mu(v - w_q) \leq s$ and $\mu(v' - w_q) > s + m$ is larger than $T(\mathcal{B}(s), \mathcal{B}(s+m)^c)$. Hence, via Lemma 3.39, and choosing the sets $V(i, j)$ as the vertices v_i, for each $w \in W_Q$ it is possible to bound the probability of the event $A(t, w)$ from above by

$$\mathbb{P}\big(X_{s,s+m}^{(1)} + X_{s,s+m}^{(2)} + \cdots + X_{s,s+m}^{(Q)} < t\big). \tag{3.61}$$

It remains to count the number of elements $(w_0, w_1, \dots, w_{Q-1})$ in W_Q. Assuming that w_q has already been chosen, the number of choices for w_{q+1} is restricted by (3.59). In particular, w_{q+1} has to be contained in a cube centered at w_q and whose side length is a multiple of $(5 + \mu(e_1))m$. This cube is intersected by at most $(Cm/s)^d$ boxes of the form $(-rs, rs]^d$ in the tiling of $\mathbb{Z}^d$, for some $C < \infty$. Since w_{q+1} is the center of one of these boxes, this is also an upper bound for its number of choices. Consequently, the total number of choices for $w_1, w_2, \dots, w_{Q-1}$ is at most $(Cm/s)^{d(Q-1)}$. Together with (3.60) and (3.61), we conclude that

$$\mathbb{P}\big(T(0, z) < t\big) \leq \sum_{Q+1 \geq n/(m+Cs)} \left(C\frac{m}{s}\right)^{d(Q-1)} \mathbb{P}\left(\sum_{q=1}^{Q} X_{s,s+m}^{(q)} < t\right),$$

for some $C < \infty$. The lemma follows observing that the number of $z \in \mathbb{Z}^d$ that satisfies $\mu(z) > n$ and has a neighbor within $\mathcal{B}(n)$ is of order n^{d-1}.

□

Given Lemma 3.40 it is now easy to complete the proof of Theorem 3.35. For any $\epsilon > 0$, by Markov's inequality, and Lemma 3.40 we have

$$\begin{aligned} &\mathbb{P}\Big(T(0, ne_1) < n(\mu(e_1) - \epsilon)\Big) \\ &\leq \sum_{Q+1 \geq n/(m+Cs)} \left(C\frac{m}{s}\right)^{d(Q-1)} \inf_{\gamma \geq 0} \left\{ e^{\gamma n(\mu(e_1)-\epsilon)} (\mathbb{E} e^{\gamma X_{s,s+m}})^Q \right\}. \end{aligned}$$

A judicious choice of s, m and Q (see [**125**, Page 196]) leads to the bound

$$\mathbb{P}(T(0, ne_1) < n(\mu(e_1) - \epsilon)) \leq e^{2\gamma m \mu(e_1)} 2^{d+1-dn/2m},$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

which ends the proof of the lower deviation estimate.

3.6.2. Upper large deviation bounds. We now turn to upper large deviation bounds. As remarked before, it does not suffice to have finite exponential moments to derive (3.46). The most natural assumption is to require the passage times to be bounded. This was the assumption used by Kesten in [**123**].

THEOREM 3.41 (Kesten [**123**], Theorem 5.9). *Assume that* $\mathbb{E}e^{\alpha\tau_e} < \infty$ *for some* $\alpha > 0$. *Then for each* $\epsilon > 0$ *there exist constants* $A_1 = A_1(\epsilon)$ *and* $B_1 = B_1(\epsilon) > 0$ *such that for all* $n \geq 1$,

$$\mathbb{P}\big(T(0, ne_1) > n(\mu(e_1) + \epsilon)\big) \leq A_1 e^{-B_1 n}. \tag{3.62}$$

If τ_e *is bounded with probability one, then for all* $\epsilon > 0$ *there exist constants* $A_2 = A_2(\epsilon)$ *and* $B_2 = B_2(\epsilon) > 0$ *such that for all* $n \geq 1$

$$\mathbb{P}\big(T(0, ne_1) > n(\mu(e_1) + \epsilon)\big) \leq A_2 e^{-B_2 n^d}. \tag{3.63}$$

Note that Kesten's bound (3.63) for bounded passage times is not of the form (3.46). Proving (3.46) is still an open question.

QUESTION 3.6.1. Assume that the passage times are bounded and not concentrated at a single point. Show that there exists a convex function $I_u(\epsilon)$ such that the following limit holds:

$$\lim_{n\to\infty} \frac{1}{n^d} \log p_n^u(\epsilon) = I_u(\epsilon).$$

Theorem 3.41 was extended by Cranston-Gauthier-Mountford. In [**64**], the authors found necessary and sufficient conditions on the distribution of edge weights so that $\limsup_n n^{-d} \log p_n^u(\epsilon) < 0$. Their result is the most recent contribution concerning large deviation bounds of passage times.

THEOREM 3.42 (Cranston-Gauthier-Mountford [**64**], Theorem 1.3). *Assume that for positive* $M_0 < \infty$ *that there is a positive increasing function* f *so that*

$$\log \mathbb{P}\big(\tau_e > x\big) = -x^d f(x), \quad x > M_0.$$

Then for every sufficiently small $\epsilon > 0$,

$$\limsup_n \frac{1}{n^d} \log \mathbb{P}\bigg(T(0, ne_1) \geq n(\mu(e_1) + \epsilon)\bigg) < 0$$

if and only if

$$\sum_{n=1}^{\infty} \frac{1}{f(2^n)^{1/(d-1)}} < \infty.$$

If instead of considering the sequence of point-to-point passage times $(T(0, ne_1))$, we take the box-to-plane passage times

$$G(0, n) = \inf\bigg\{T(\Gamma) : \Gamma \text{ a path from } \{0\} \times [-n, n]^{d-1} \text{ to } \{n\} \times \mathbb{Z}^{d-1}\bigg\},$$

then it is shown in [**103**] that

$$\lim_n \frac{G(0, n)}{n} = \lim_n \frac{T(0, ne_1)}{n} = \mu(e_1).$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

In this case, assuming $F(0) < p_c$, Chow and Zhang [**59**] showed the existence of convex functions $c(\epsilon)$ and $c'(\epsilon)$ so that

$$\lim_n \frac{1}{n^d} \log \mathbb{P}\Big(G(0,n) \geq n(\mu(e_1)+\epsilon)\Big) = c(\epsilon)$$

and

$$\lim_n \frac{1}{n} \log \mathbb{P}\Big(G(0,n) \leq n(\mu(e_1)-\epsilon)\Big) = c'(\epsilon).$$

We finish this section proving equation (3.62) of Theorem 3.41 when $d = 2$. For $1 \leq m \leq n$ and $k \geq 1$ let $T_{m,n}(k)$ be the first-passage time from me_1 to ne_1 over paths that do not leave the strip $\{(x,y) \in \mathbb{Z}^2 : |y| \leq k\}$. As $T(0,ne_1) \leq T_{0,n}(k)$, it suffices to show the estimate for $T_{0,n}(k)$. First note that for any fixed ℓ we have

$$\lim_{k\to\infty} \mathbb{E}T_{0,\ell}(k) = \mathbb{E}T(0,\ell e_1)$$

so we can choose a k large enough for which

$$\mathbb{E}T_{0,\ell}(k) \leq \ell(\mu(e_1) + 3\epsilon/5).$$

Now,

$$T_{0,\ell m}(k) \leq \sum_{j=0}^{m-1} T_{j\ell,(j+1)\ell}(k)$$

and the random variables $T_{j\ell,(j+1)\ell}(k)$ are almost independent with the same distribution as $T_{0,\ell}(k)$ (to get independence one can restrict to paths that also stay in vertical strips). Applying the Chernoff inequality, we find constant A_1 and B_1 so that

$$\mathbb{P}\Big(T_{0,m\ell}(k) \geq m\ell(\mu(e_1)+3\epsilon/5)\Big) \leq A_1 \exp(-B_1 m), \quad m \geq 1.$$

Writing $n = m\ell + q$, with $0 \leq q < \ell$ we use $T_{0,n}(k) \leq T_{0,m\ell}(k) + T_{m\ell,n}(k)$ and

$$\mathbb{P}\Big(T_{m\ell,n}(k) \geq \epsilon/5\Big) \leq e^{-\gamma\epsilon/5}(\mathbb{E}\exp(\gamma\tau_e))^q$$

to get

$$\mathbb{P}\Big(T_{0,n}(k) \geq n(\mu(e_1)+4\epsilon/5)\Big) \leq A\exp(-Bn), \quad n \geq 1$$

for suitable $A, B > 0$.

3.7. Cases where Gaussian fluctuations appear

In this section, we discuss a few cases where Gaussian fluctuations are known or expected to appear. The first example is critical FPP, where the edge-weights are taken with $F(0) = p_c$, exactly at the percolation threshold. The second case is FPP in a thin cylinder, where geodesics from 0 to ne_1 are constrained to be in cylinders of cross section n^α for some small $\alpha > 0$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

3.7.1. Critical FPP. From Kesten's work (Theorems 2.5 and 2.6), we know that

$$p := \mathbb{P}(\tau_e = 0) \geq p_c \Leftrightarrow \mu(x) = 0 \text{ for all } x.$$

Most theorems in FPP have the assumption that the above probability is strictly less than p_c for this reason. Namely, if $p > p_c$, then one can show that as $|x| \to \infty$, the family of passage times $T(0,x)$ is stochastically bounded above by a random variable that does not depend on x. The reason is that to travel from 0 to x, one simply needs to go from 0 to the infinite cluster of zero-weight edges, then from there to x. On the other hand, if $p = p_c$, then very little is known. This is due to the fact that one of the outstanding problems in probability theory is to show that there is no infinite cluster of open edges in bond percolation on $\mathbb{Z}^d$ at the critical point. Therefore, we do not know if vertices can take advantage of an infinite cluster of zero-weight edges when $p = p_c$, and so we do not even know the growth rate of $T(0,x)$ as $|x| \to \infty$.

The first result on the critical case came from Chayes [**54**], who showed in the Bernoulli case:

THEOREM 3.43 (Chayes [**54**]). *Let $d \geq 3$ and τ_e satisfy $\mathbb{P}(\tau_e = 0) = p_c = 1 - \mathbb{P}(\tau_e = 1)$. If $\epsilon > 0$, then*

$$\lim_{n\to\infty} \frac{T(0, ne_1)}{n^\epsilon} = 0 \text{ a.s.}.$$

See [**123**, Remark 3] also, where Kesten claims that Chayes's argument can be extended to

$$T(0, ne_1) \leq \exp(C\sqrt{\log n})$$

for large n a.s..

Although this is the extent of knowledge on the critical case in general, much more progress has been made in two dimensions. In this case, Zhang [**185**] showed that different critical edge-weight distributions can display completely different asymptotic behavior: defining the passage time to infinity as

$$\rho = \lim_{n\to\infty} T(0, \partial B(n)), \tag{3.64}$$

where $B(n) = [-n,n]^2$, there are some critical distributions for which $\rho = \infty$ a.s., and there are some for which $\rho < \infty$ a.s.. He constructed such distributions by comparing critical FPP to a model called the Chayes-Chayes-Durrett incipient infinite cluster [**56**]. This is an inhomogeneous bond percolation model which is tuned to resemble a large Bernoulli percolation cluster at criticality. A complete characterization was given recently in [**71**], using methods related to $2d$ invasion percolation. For its statement, let F be the distribution function of τ_e and set

$$F^{-1}(y) = \inf\{x : F(x) \geq t\} \text{ for } t \geq 0.$$

THEOREM 3.44 (Damron-Lam-Wang [**71**]). *For $d = 2$, one has $\rho < \infty$ a.s. if and only if $\sum_{n=2}^{\infty} F^{-1}(p_c + 2^{-n}) < \infty$.*

The theorem implies that if F is concentrated enough near 0, then the passage time to infinity can be finite. Otherwise, if there are too many large edge-weights, the passage time to infinity will be infinite. For example, $\rho < \infty$ when F decays to p_c like a polynomial near 0 (say $F(\epsilon) \sim p_c + \epsilon^\alpha$ for some $\alpha > 0$), but can be infinity if F decays to p_c like an exponential (with large enough rate) near 0 (say $F(\epsilon) \sim p_c + e^{-\epsilon^{-\beta}}$ for $\beta \geq 1$).

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

In the case that $\rho = \infty$, one can ask about the rate of growth of $T(0, \partial B(n))$ as $n \to \infty$. The first result of this type was due to Chayes-Chayes-Durrett [**55**], who showed that in the critical Bernoulli case,

$$\mathbb{E}T(0, \partial B(n)) \asymp \log n.$$

The general case was given in [**71**]. Let Y be the minimum of 4 i.i.d. copies of τ_e.

THEOREM 3.45 (Damron-Lam-Wang [**71**]). *For $d = 2$, suppose that $F(0) = p_c$ and $\mathbb{E}Y^\alpha < \infty$ for some $\alpha > 1/4$. There exists $C > 0$ such that for all n,*

$$\frac{1}{C}\mathbb{E}T(0, \partial B(2^n)) \leq \sum_{k=2}^{n} F^{-1}(p_c + 2^{-k}) \leq C\mathbb{E}T(0, \partial B(2^n)).$$

If $\mathbb{E}Y^\alpha < \infty$ for some $\alpha > 1/2$, then there exists $C' > 0$ such that for all n,

$$\frac{1}{C'}\mathrm{Var}\ T(0, \partial B(2^n)) \leq \sum_{k=2}^{n} \left(F^{-1}(p_c + 2^{-k})\right)^2 \leq C'\mathrm{Var}\ T(0, \partial B(2^n)).$$

In the case $F(0) < p_c$, establishing limiting laws for $T(0, ne_1)$ is far off, except in some related exactly solvable models (see Section 7.2.3). Kesten and Zhang, however, showed that in the Bernoulli critical case in two dimensions, one has a Gaussian central limit theorem on scale $\sqrt{\log n}$. Their theorem applies to all distributions with a gap near zero; that is, ones which have $F(0) = p_c$ and $F(a) = p_c$ for some $a > 0$. The reason for this condition is precisely the uncertain behavior of ρ in the critical case.

THEOREM 3.46 (Kesten-Zhang [**127**]). *Let $d = 2$. Suppose that for some $C > 0$, $F(0) = F(C) = 1/2$. Also assume that $\mathbb{E}\tau_e^{4+\delta} < \infty$ for some $\delta > 0$. Then there exist positive constants C_1, C_2, γ_n with*

$$C_1(\log n)^{1/2} \leq \gamma_n \leq C_2(\log n)^{1/2}$$

such that for all $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\big(\gamma_n^{-1}\big[T(0, ne_1) - \mathbb{E}T(0, ne_1)\big] \leq t\big) = \mathbb{P}(Z \leq t)$$

where Z is a standard Gaussian random variable.

Theorem 3.46 was extended to general distributions in [**71**]. Note that if $\sum_k F^{-1}(p_c + 2^{-k}) = \infty$ but $\sum_k (F^{-1}(p_c + 2^{-k}))^2 < \infty$, then the variance of the passage time converges, whereas the mean diverges. In this case, we do not get a Gaussian limit, but rather $T(0, \partial B(n)) - \mathbb{E}T(0, \partial B(n))$ converges to some other variable.

THEOREM 3.47 (Damron-Lam-Wang [**71**]). *For $d = 2$, suppose that $F(0) = p_c$, and $\sum_{k=2}^\infty F^{-1}(p_c + 2^{-k}) = \infty$. Suppose further that $\mathbb{E}Y^\alpha < \infty$ for some $\alpha > 1/2$.*

(1) *If $\sum_{k=2}^\infty (F^{-1}(p_c + 2^{-k}))^2 < \infty$, then there is a random variable Z with $\mathbb{E}Z = 0$ and $\mathbb{E}Z^2 < \infty$ such that as $n \to \infty$,*

$$T(0, \partial B(n)) - \mathbb{E}T(0, \partial B(n)) \to Z \text{ a.s. and in } L^2.$$

(2) *If $\sum_{k=2}^\infty (F^{-1}(p_c + 2^{-k}))^2 = \infty$, then as $n \to \infty$,*

$$\frac{T(0, \partial B(n)) - \mathbb{E}T(0, \partial B(n))}{\sqrt{\mathrm{Var}\ T(0, \partial B(n))}} \Rightarrow N(0, 1).$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

In the case of site FPP on the $2d$ triangular lattice with Bernoulli weights ($\mathbb{P}(t_v = 0) = p_c = 1 - \mathbb{P}(t_v = 1)$), Yao [**180**] used the conformal loop ensemble to find the exact values of the constants appearing in the first-order asymptotics of the mean and variance of $T(0, \partial B(n))$.

THEOREM 3.48 (Yao [**180**]). *Consider site FPP on the triangular lattice with weights (t_v) satisfying $\mathbb{P}(t_v = 0) = p_c = 1 - \mathbb{P}(t_v = 1)$. The following statements hold:*

$$\lim_{n\to\infty} \frac{T(0, \partial B(n))}{\log n} = \frac{1}{2\sqrt{3}\pi} \ a.s.,$$
$$\lim_{n\to\infty} \frac{\mathbb{E}T(0, \partial B(n))}{\log n} = \frac{1}{2\sqrt{3}\pi},$$
$$\lim_{n\to\infty} \frac{\mathrm{Var}\ T(0, \partial B(n))}{\log n} = \frac{4}{3\sqrt{3}\pi} - \frac{1}{\pi^2}.$$

As mentioned before, there is a big difference between what is known in two dimensions and in three and higher dimensions. This is due mostly to the lack of knowledge about critical percolation in three and higher dimensions. A natural question is then:

QUESTION 3.7.1. Let $d = 3$ and $\mathbb{P}(\tau_e = 0) = p_c = 1 - \mathbb{P}(\tau_e = 1)$. Is it true that

$$\mathbb{E}T(0, \partial B(n)) \asymp \log n?$$

The two-dimensional proof of this fact uses that with uniformly positive probability in n, every path which crosses $B(n) \setminus B(n/2)$ must contain at least one edge which does not have zero weight. This is known not to be true in high dimensions [**6**, Theorem 4], so it is conceivable that for large d, one has $T(0, \partial B(n)) \ll \log n$ even in the Bernoulli case.

3.7.2. Passage time in small cylinders. A second case where Gaussian fluctuations are present is when we consider passage times of paths constrained in a thin cylinder. Precisely, for $h > 0$, let $a_n(h)$ be the first-passage time from 0 to the point ne_1 in the graph $(\mathbb{Z} \times [-h, h]^{d-1}) \cap \mathbb{Z}^d$; that is, define

$$a_n(h) := \inf\{T(\Gamma) \ : \ \Gamma \text{ is a path from 0 to } ne_1 \text{ in } \mathbb{Z} \times [-h, h]^{d-1}\}.$$

If h is small enough, the constrained geodesic from 0 to ne_1 should not coincide with the one in the full integer lattice. In this case, we have Gaussian fluctuations for $a_n(h)$ after proper centering and scaling. (See related work in [**1**].)

THEOREM 3.49 (Chatterjee-Dey [**52**]). *Suppose that $\mathbb{E}\tau_e^{2+\delta} < \infty$ for some $\delta > 0$. Assume that $h = h(n)$ satisfies $h(n) = o(n^\alpha)$ with*

$$\alpha < \frac{1}{d + 1 + 2(d-1)/\delta}. \tag{3.65}$$

Then we have for all $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{a_n(h_n) - \mathbb{E}a_n(h_n)}{\sqrt{\mathrm{Var}(a_n(h_n))}} \le t \right) = \mathbb{P}(Z \le t) \tag{3.66}$$

where Z is a standard Gaussian random variable.

The theorem above gives rise to the following question, raised in [**52**] and partially answered in sections 8 and 9 of that paper.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

QUESTION 3.7.2. Is (3.65) optimal? More explicitly, let

$$\gamma_F(d) := \sup\left\{\alpha : (3.66) \text{ holds for } h_n = n^\alpha\right\}.$$

What is the value of $\gamma_F(d)$?

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/04

CHAPTER 4

Geodesics

Recall the definition of a geodesic from x to y.

DEFINITION 4.1. A path Γ from x to y with $T(\Gamma) = T(x, y)$ is called a *geodesic* from x to y.

Note that a subpath of a geodesic is also a geodesic. We have three different definitions of the set of geodesics between two points; these all coincide when there is only one geodesic between a pair of points:

$$\begin{aligned}\underline{\mathrm{GEO}}(x,y) &= \text{(edge) intersection of all geodesics from } x \text{ to } y,\\ \mathrm{GEO}(x,y) &= \text{self-avoiding geodesic from } x \text{ to } y\\ &\qquad\qquad \text{with maximal number of edges,}\\ \overline{\mathrm{GEO}}(x,y) &= \text{(edge) union of all geodesics from } x \text{ to } y.\end{aligned}$$

Soon, we will deal with infinite geodesics, so we will sometimes use the term finite geodesic for a geodesic between two points. We will also require geodesics from points to sets: if A is a subset of $\mathbb{Z}^d$, then a geodesic from x to A is a geodesic from x to y, where y is the vertex of A minimizing $T(x, y)$. If A is a subset of $\mathbb{R}^d$, then we identify each point z of A with the vertex z' of $\mathbb{Z}^d$ such that $z \in z' + [0, 1)^d$, and define $\hat{A}$ to be the union of all such vertices; the set of all geodesics from x to A is defined to be the set of all geodesics from x to $\hat{A}$. In any case, we use the terms $\underline{\mathrm{GEO}}(x, A)$, $\mathrm{GEO}(x, A)$, and $\overline{\mathrm{GEO}}(x, A)$ similarly — as the intersection of all geodesics from x to A, the geodesic of maximal number of edges, and the union of all geodesics from x to A.

4.1. Existence of finite geodesics and their sizes

The very first step in the study of geodesics is to determine if a geodesic from x to y exists with probability one. Hammersley and Welsh [**108**] showed almost sure existence of finite geodesics if the distribution of the passage times is bounded above and below away from zero. They also conjectured that a.s. existence held for all distributions, with no moment conditions. Smythe and Wierman [**160, 161**] verified this conjecture in two dimensions except for distributions with an atom at zero equal to p_c. In $\mathbb{Z}^2$, the conjecture was proven in full generality by Wierman and Reh [**177**], and it is now the state of the art. It was stated in [**177**, Corollary 1.3] for geodesics between 0 and ne_1, but the same proof applies to any pair of points.

THEOREM 4.2 (Wierman-Reh [**177**]). *For any F, there a.s. exists a geodesic between any two points of $\mathbb{Z}^2$.*

Theorem 4.2 was extended to connected, infinite subsets of $\mathbb{Z}^2$ with connected infinite complement in [**20**]. Existence under no assumptions on the distribution of passage times is currently open in d dimensions if $d > 2$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

QUESTION 4.1.1. Prove that under no assumptions on F, geodesics a.s. exist in the d-dimensional lattice, $d \geq 3$.

We will now explain how, as noted in [**123**, (9.23)], a.s. existence of geodesics follows from standard estimates provided $F(0) < p_c$. In the case $F(0) > p_c$ with $\mathbb{E}\tau_e < \infty$, existence of geodesics was proved in [**189**]. The case $F(0) = p_c$ is still unresolved in general when $d > 2$. However, geodesics are known to exist for distributions F satisfying $F(0) = F(\delta) = p_c$ for some $\delta > 0$ and in sufficiently high dimensions (since it is known that there is no percolation at the critical point in high dimensions).

Recall the definition of the passage time to infinity (3.64), from [**185**]:

$$\rho = \lim_{n \to \infty} [T(0, \partial B(n))],$$

where $B(n) = [-n, n]^d$. By the Kolmogorov 0-1 law, either $\rho = \infty$ a.s., or $\rho < \infty$ a.s..

LEMMA 4.3. *One has*

$$\rho = \inf\{T(\Gamma) : \Gamma \textit{ is an infinite self-avoiding path from } 0\}.$$

PROOF. Let Γ_n be a (self-avoiding) geodesic from 0 to $\partial B(n)$, with Γ a subsequential limit of (Γ_n) along some subsequence (Γ_{n_k}). (Here, we are using the definition of convergence from Definition 4.24.) For a fixed n, if $\hat{\Gamma}_n$ is the initial portion of γ from 0 to $\partial B(n)$, then for all large k, the geodesic from 0 to $\partial B(n_k)$ contains $\hat{\Gamma}_n$ as its first segment, so

$$T(0, \partial B(n_k)) \geq T(\hat{\Gamma}_n) \geq T(0, \partial B(n)).$$

Letting $k \to \infty$, we obtain

$$\rho \geq T(\hat{\Gamma}_n) \geq T(0, \partial B(n)).$$

As $n \to \infty$, the middle term $T(\hat{\Gamma}_n)$ converges to $T(\Gamma)$ and the right term converges to ρ, so $T(\Gamma) = \rho$, showing the inequality $\geq$ in the lemma. The other inequality holds because each infinite self-avoiding path from 0 contains an initial segment from 0 to $\partial B(n)$. □

From Lemma 4.3, we can show that $F(0) < p_c$ implies that geodesics exist a.s..

PROPOSITION 4.4. *The following hold.*

(1) *If $\rho = \infty$ for some configuration (τ_e), then for all $x, y \in \mathbb{Z}^d$, a geodesic exists from x to y.*
(2) *If $F(0) < p_c$, then a.s., $\rho = \infty$.*

PROOF. If $\rho = \infty$ in some passage time configuration (τ_e), pick any x, y, and fix a path Γ between them. Note that since $\rho = \infty$ and

$$|T(x, \partial B(n)) - T(0, \partial B(n))| \leq T(0, x) < \infty,$$

one has $T(x, \partial B(n)) \to \infty$ as so one can choose n large enough so that (a) $x \in B(n)$ and (b) $T(x, \partial B(n)) > T(\Gamma)$. If Γ' is any path from x to y that exits $B(n)$, then

$$T(\Gamma') \geq T(x, \partial B(n)) > T(\Gamma),$$

so the infimum in the definition of $T(x, y)$ is over a finite set of paths, and there must be a minimizer.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

If $F(0) < p_c$, then choose $\delta > 0$ such that $F(\delta) < p_c$. By the definition of p_c, a.s. there is no infinite self-avoiding path of edges e with $\tau_e \leq \delta$, and in fact each such path must have infinitely many edges with weight at least δ. So by the characterization in Lemma 4.3, $\rho = \infty$ a.s.. $\square$

The condition $\rho = \infty$ is not necessary for existence of geodesics, as they exist when $F(0) > p_c$ (and therefore $\rho < \infty$). It is unknown in general dimensions if $\rho = \infty$ for a given F, but in two dimensions, Theorem 3.44 gives necessary and sufficient conditions.

QUESTION 4.1.2. Determine necessary and sufficient conditions on F for dimension $d \geq 3$ so that $\rho = \infty$.

We turn our attention to the size of finite geodesics. As we have seen, the size of a geodesic is easy to estimate in an edge-weight configuration in which all weights are bounded below by a positive constant. For example, if $\tau_e \geq a > 0$ for all e, then if σ is any geodesic from 0 to x, we have

$$a|\sigma| \leq T(\sigma) = T(0,x),$$

so $|\sigma| \leq \frac{1}{a}T(0,x)$. Therefore, in this setting, all moments of the maximal length of a geodesic from 0 to x are bounded by the corresponding moments of $T(0,x)$. In particular, if $\tau_e \leq b$ for all e, then, using the bound $T(0,x) \leq b\|x\|_1$, we obtain $|\sigma| \leq \frac{b}{a}\|x\|_1$. Such linear bounds are more difficult to obtain in the general case, which we now present. However, they are not true in the critical case, where $F(0) = p_c$ (see [**72**] for a superlinear lower bound in dimension two).

We focus on two ways of measuring the size of finite geodesics: the maximal length of a geodesic and the diameter of the union of all geodesics between two points. A fundamental estimate is given by Kesten.

LEMMA 4.5 (Kesten [**123**], Proposition 5.8). *Assume $F(0) < p_c$. There exist $a, C > 0$ such that for all $n \in \mathbb{N}$,*

$$\mathbb{P}\left(\exists \textit{ self-avoiding path } \Gamma \textit{ containing } 0 \textit{ with } |\Gamma| \geq n \textit{ but } T(\Gamma) < an\right) \leq e^{-Cn}.$$

PROOF. This is a simple application of Lemma 3.39. Let $M \geq 1$, to be fixed soon. Let Γ be a self-avoiding path starting at 0 with length at least n. Label its vertices (v_i) in order, with $v_0 = 0$. Set $t(0) = 0, a_i = v_{t(i)} \in \Gamma$ where

$$t(q+1) = \min\{t > t(q) : \|a_q - v_t\|_1 = M\}.$$

We choose these points until we cannot anymore; that is, until we arrive at a Q so that $\|a_Q - v_t\|_1 < M$ for all $t > t(Q)$. As $|\Gamma| \geq n$ it is clear that

$$Q + 1 \geq (2M+1)^{-d} n.$$

We now consider all possible positions taken by the skeleton $(a_i)_{i=1,\ldots,Q}$. The probability in Lemma 4.5 is bounded above by

$$\sum_{Q \geq n(2M+1)^{-d}-1} \sum_{(a_i)} \mathbb{P}\Bigg(\Gamma \text{ passes successively through } (a_i) \\ \text{and } \sum_{i=0}^{Q-1} T(a_i, a_{i+1}) < Cn\Bigg).$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Using Lemma 3.39 this is bounded above by

$$\sum_{Q \geq n(2M+1)^{-d}-1} \sum_{(a_i)} \mathbb{P}\Big(\sum_{i=0}^{Q-1} T'(a_i, a_{i+1}) < Cn \Big),$$

where the $T'(a_i, a_{i+1})$'s are independent copies of the $T(a_i, a_{i+1})$'s. The Markov inequality applied to exponentials and the fact that the sum over (a_i) is restricted to skeletons with $\|a_i - a_{i+1}\|_1 = M$ yield

$$\begin{aligned} \sum_{(a_i)} \mathbb{P}\Big(\sum_{i=0}^{Q-1} T'(a_i, a_{i+1}) < Cn \Big) &\leq e^{\lambda Cn} \textstyle\sum_{(a_i)} \prod_{i=0}^{Q-1} \mathbb{E} e^{-\lambda T(a_i, a_{i+1})} \\ &= e^{\lambda Cn} \Big(\textstyle\sum_{\|a\|_1 = M} \mathbb{E} e^{-\lambda T(0,a)} \Big)^Q. \end{aligned}$$

Finally, choose M such that

$$\sum_{\|a\|_1 = M} \mathbb{P}(T(0,a) = 0) \leq 1/2.$$

This is possible as $F(0) < p_c$: an upper bound for the left side is the expected volume of the intersection of the cluster of zero-weight edges of the origin with the sphere of radius M, which tends to zero as $M \to \infty$ (see [**102**, Theorem 5.75] and [**78**]). Note that for this M, one has

$$\lim_{\lambda \to \infty} \sum_{\|a\|_1 = M} \mathbb{E} e^{-\lambda T(0,a)} = \sum_{\|a\|_1 = M} \mathbb{P}(T(0,a) = 0) \leq 1/2.$$

Once this M is chosen, choose λ so large and $C > 0$ so small that for $n \leq (2M + 1)^d (Q+1)$

$$e^{\lambda Cn} \Big(\sum_{\|a\|_1 = M} \mathbb{E} e^{-\lambda T(0,a)} \Big)^Q \leq e^{\lambda C (2M+1)^d (Q+1)} \left(\frac{3}{4}\right)^Q \leq \left(\frac{7}{8}\right)^Q.$$

Therefore the upper bound for the probability in Lemma 4.5 we obtain is

$$\sum_{Q \geq n(2M+1)^{-d}-1} \left(\frac{7}{8}\right)^Q,$$

which implies the result. □

Using Borel-Cantelli in combination with Lemma 4.5, one has that if $F(0) < p_c$, then

$$\liminf_{\|x\|_1 \to \infty} \frac{T(0,x)}{\|x\|_1} \geq a > 0 \text{ a.s.}. \tag{4.1}$$

We will also use the following consequence of Kesten's lemma.

LEMMA 4.6. *Assume $F(0) < p_c$ and let a be from Lemma 4.5. There exists $D > 0$ such that the variable*

$$Y_x = |\mathrm{GEO}(0,x)| \mathbf{1}_{\{T(0,x) < a|\mathrm{GEO}(0,x)|\}}$$

satisfies $\mathbb{P}(Y_x \geq n) \leq e^{-Dn}$ for all $x \in \mathbb{Z}^d$ and $n \in \mathbb{N}$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Proof. Defining

$$A_m = \{\exists \text{ self-avoiding path } \Gamma \text{ from } 0 \text{ with } |\Gamma| \geq m \text{ but } T(\Gamma) < a|\Gamma|\},$$

and summing the bound of Lemma 4.5 over $n \geq m$, one has, for some $C_1 > 0$,

$$\mathbb{P}(A_m) \leq e^{-C_1 m} \text{ for all } m \in \mathbb{N}. \tag{4.2}$$

For $x \in \mathbb{Z}^d$, assume $Y_x \geq n \geq 1$. Then because $Y_x \neq 0$, one has $|\text{Geo}(0,x)| > (1/a)T(0,x)$, and so

$$T(\text{Geo}(0,x)) = T(0,x) < a|\text{Geo}(0,x)|,$$

with $|\text{Geo}(0,x)| \geq n$. So A_n occurs and (4.2) completes the proof. □

We now give bounds on the moments of $|\text{Geo}(0,x)|$.

Proposition 4.7. *Assume $F(0) < p_c$.*

(1) *For any $\alpha > 0$, there exists a constant C_α such that for all nonzero $x \in \mathbb{Z}^d$, one has*

$$\mathbb{E}|\text{Geo}(0,x)|^\alpha \leq C_\alpha \mathbb{E}T(0,x)^\alpha.$$

(2) *For a set of edges E, let V be the set of endpoints of its edges and define*

$$T_E = \max_{y,z \in V} T(y,z).$$

Further, define

$$|\text{Geo}_E(0,x)| = \max\{|\sigma \cap E| : \sigma \text{ a self-avoiding geodesic from } 0 \text{ to } x\}.$$

For any $\alpha > 0$, there exists $C_\alpha > 0$ such that for all $x \in \mathbb{Z}^d$ and all E, one has

$$\mathbb{E}|\text{Geo}_E(0,x)|^\alpha \leq C_\alpha \mathbb{E}T_E^\alpha.$$

Note that if $\tau_e \leq b < \infty$ a.s. for some b, then the right side of the inequality in item 2 is bounded by $(b \text{ diam } V)^\alpha$. Under the additional condition $\mathbb{E}e^{\beta\tau_e} < \infty$ for some $\beta > 0$, the proof below of item 1 shows a similar bound for exponential moments: there exists $\alpha_1 > 0$ such that

$$\sup_{0 \neq x \in \mathbb{Z}^d} \frac{\log \mathbb{E}e^{\alpha_1 |\text{Geo}(0,x)|}}{\|x\|_\infty} < \infty. \tag{4.3}$$

Proof. To prove item 1, we estimate

$$\mathbb{E}|\text{Geo}(0,x)|^\alpha \leq a^{-\alpha}\mathbb{E}T(0,x)^\alpha + \mathbb{E}Y_x^\alpha,$$

where Y_x is from Lemma 4.6. Since $\mathbb{E}Y_x^\alpha$ is bounded uniformly in x, equation (4.1) (which gives $\mathbb{E}T(0,x)^\alpha \to \infty$ as $\|x\|_1 \to \infty$) shows item 1.

For item 2, we may assume that $|E| < \infty$. If $|E \cap \sigma| \geq \lambda > 0$ for a self-avoiding geodesic σ from 0 to x, then we may find the first and last intersections (say y and z respectively) of σ with V. The portion of σ from y to z is then a geodesic whose edge set has cardinality at least λ. If no self-avoiding path starting from y with at least λ edges has passage time less than $a\lambda$, then we obtain $T(y,z) \geq a\lambda$. Using Lemma 4.5 along with a union bound over all possible values of the starting point y, we obtain

$$\mathbb{P}(|\text{Geo}_E(0,x)| \geq \lambda) \leq |V|e^{-C\lambda} + \mathbb{P}(T_E \geq a\lambda).$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Since $|V| \le c(\text{diam } E)^d$ for a universal c, we have

$$\begin{aligned}
&\mathbb{E}|\textsc{Geo}_E(0,x)|^\alpha \\
&\le (\text{diam } E)^\alpha + \int_{\text{diam } E}^\infty \alpha\lambda^{\alpha-1}\mathbb{P}(|\textsc{Geo}_E(0,x)| \ge \lambda)\ \mathrm{d}\lambda \\
&\le (\text{diam } E)^\alpha + \int_{\text{diam } E}^\infty \alpha\lambda^{\alpha-1}(|V|e^{-C\lambda} + \mathbb{P}(T_E \ge a\lambda))\ \mathrm{d}\lambda \\
&\le (\text{diam } E)^\alpha + \int_{\text{diam } E}^\infty \alpha\lambda^{\alpha-1}c\lambda^d e^{-C\lambda}\ \mathrm{d}\lambda + \int_0^\infty \alpha\lambda^{\alpha-1}\mathbb{P}(T_E \ge a\lambda)\ \mathrm{d}\lambda \\
&\le (\text{diam } E)^\alpha + C_\alpha(1 + \mathbb{E}T_E^\alpha).
\end{aligned}$$

Using (4.1) along with Fatou's lemma, however, we see that $\mathbb{E}T_E^\alpha \ge C'_\alpha(\text{diam } E)^\alpha$ for some $C'_\alpha > 0$ that is independent of E, so we finish with the bound in item 2. □

The diameter of the set $\overline{\textsc{Geo}}(0,x)$ is controlled by the following result that requires no moment condition on the distribution of τ_e.

THEOREM 4.8 (Auffinger-Damron-Hanson [**21**], Theorem 6.2). *Assume $F(0) < p_c$. Then there exist positive constants M, C such that*

$$\mathbb{P}\Big(\text{diam } \overline{\textsc{Geo}}(0,x) \ge M\|x\|_\infty\Big) \le \exp(-C\|x\|_\infty) \quad \textit{for all } x \in \mathbb{Z}^d.$$

Furthermore, we have the following linear bound on the number of edges in a geodesic.

THEOREM 4.9. *Assume $F(0) < p_c$. Then there exist positive constants M, C such that*

$$\mathbb{P}\left(|\textsc{Geo}(0,x)| \ge M\|x\|_\infty\right) \le \exp\left(-C\|x\|_\infty^{1/d}\right) \quad \textit{for all } x \in \mathbb{Z}^d.$$

The above theorems imply other large-deviation inequalities for sizes of geodesics under minimal assumptions. For instance, one can show that if $F(0) < p_c$, there are constants $C_1, C_2 > 0$ such that for all $x \in \mathbb{Z}^d$ and all $\lambda \ge C_1$, one has

$$\mathbb{P}(\text{diam } \overline{\textsc{Geo}}(0,x) \ge \lambda\|x\|_\infty) \le C_1 \exp\left(-C_2\lambda\|x\|_\infty\right). \tag{4.4}$$

To see why, note that if diam $\overline{\textsc{Geo}}(0,x) \ge \lambda|x|$ then we can find $y \in \mathbb{Z}^d$ with $|y| \ge (\lambda/2)|x|$ such that a geodesic Γ from 0 to x contains y. Let v be a vertex that Γ visits after y with

$$(\lambda/(4M))|x| \le |v| \le (\lambda/(2M))|x|,$$

where λ is taken $\ge 4M$, M is from Theorem 4.8, and M is assumed to be > 1. Then

$$\text{diam } \overline{\textsc{Geo}}(0,v) \ge (\lambda/2)|x| \ge M|v| \ge M\|v\|_\infty.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

So by Theorem 4.8, one has

$$\begin{aligned}
&\mathbb{P}\left(\operatorname{diam} \overline{\mathrm{Geo}}(0,x) \geq \lambda\sqrt{d}\|x\|_\infty\right) \\
&\leq \mathbb{P}\left(\operatorname{diam} \overline{\mathrm{Geo}}(0,x) \geq \lambda|x|\right) \\
&\leq \sum_v \exp(-C\|v\|_\infty) \\
&\leq C_3\left((\lambda/(2M))\|x\|_\infty\right)^d \exp\left(-C_4(\lambda/(4M))\|x\|_\infty\right) \\
&\leq C_5 \exp\left(-C_6\lambda\|x\|_\infty\right).
\end{aligned}$$

The proofs of Theorems 4.8 and 4.9 use almost the same method, and since the first has appeared, we give the second.

Proof of Theorem 4.9. For $p \in [0,1]$ let $\mathbb{P}_p$ be the product measure on $\Omega = \{0,1\}^{\mathcal{E}^d}$ with marginal $\mathbb{P}_p(\omega(e) = 1) = p$, where ω is a typical element of Ω. In a configuration ω we write $x \to y$ if there is a path from x to y with edges e satisfying $\omega(e) = 1$. This gives a connectivity equivalence relation and the equivalence classes are called open clusters. It is known that for $p > p_c$, there is a.s. a unique infinite open cluster. Define $B(n) = \{x \in \mathbb{Z}^d : \|x\|_\infty \leq n\}$ and $\partial B(n) = \{x \in \mathbb{Z}^d : \|x\|_\infty = n\}$.

Lemma 4.10. *Let A_n be the event that every path from 0 to $\partial B(n)$ intersects the infinite open cluster. There exists $p_0 \in (p_c, 1)$ such that if $p \in [p_0, 1]$ then for some $C_1 > 0$,*

$$\mathbb{P}_p(A_n) \geq 1 - e^{-C_1 n} \text{ for all } n.$$

Proof. The proof is a slight modification of the result of Kesten [**123**, Theorem 2.24], and is given as [**21**, Lemma 6.3]. □

We now recall the result of Antal-Pisztora and afterward prove Theorem 4.9.

Lemma 4.11 (Antal-Pisztora [**15**], Theorem 1.1). *Let $p > p_c$. Then there exists a constant $\rho = \rho(p,d) \in [1, \infty)$ such that*

$$\limsup_{\|y\|_\infty \to \infty} \frac{1}{\|y\|_\infty} \log \mathbb{P}_p(d_{\mathsf{C}}(0,y) > \rho\|y\|_\infty,\ 0, y \in \mathsf{C}) < 0,$$

where C is the infinite open cluster and d_{C} is the intrinsic distance in C.

We choose $K > 0$ such that $p := \mathbb{P}(\tau_e \leq K) > p_0$, where p_0 is from Lemma 4.10, and define a percolation configuration (η_e) from the weights (τ_e) by

$$\eta_e = \begin{cases} 1 & \text{if } \tau_e \leq K \\ 0 & \text{if } \tau_e > K \end{cases}.$$

For $x \in \mathbb{Z}^d$ and an integer $M > 0$, we first write $n_x = \lfloor \|x\|_\infty^{1/d} \rfloor$ and estimate

$$\mathbb{P}(|\mathrm{Geo}(0,x)| \geq M\|x\|_\infty) \leq 2e^{-C_1 n_x} + \mathbb{P}(|\mathrm{Geo}(0,x)| \geq M\|x\|_\infty,\ A_{n_x},\ B_{n_x}), \tag{4.5}$$

where A_{n_x} is written for the event in Lemma 4.10 for the percolation configuration (η_e) and B_{n_x} is the same event with 0 translated to x. From now on, we take $\|x\|_\infty \geq C'$ with C' chosen so that $n_x \geq 1$ and $4n_x < \|x\|_\infty$. On the event on the right of (4.5), write Γ_1 for the portion of $\mathrm{Geo}(0,x)$ from 0 to its first intersection of $\partial B(n_x)$ and Γ_2 for the portion from its last intersection of $\partial B(x, n_x)$ to x (here

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

$B(x,n)$ is the translate of $B(n)$ centered at x). We can then choose u, v vertices of Γ_1 and Γ_2 respectively such that $\|u\|_\infty \le n_x$, and $\|v-x\|_\infty \le n_x$ and both u, v are in C. By construction, the portion of $\textsc{Geo}(0,x)$ from u to v is a self-avoiding geodesic that has at least

$$M\|x\|_\infty - C_2 n_x^d \ge (M - C_2)\|x\|_\infty \text{ number of edges.}$$

We now apply Lemma 4.11 to find $C_3, C_4 > 0$ such that

$$\mathbb{P}_p(d_{\mathsf{C}}(0,y) > \rho\|y\|_\infty,\ 0, y \in \mathsf{C}) \le C_4 e^{-C_3\|y\|_\infty} \text{ for all } y \in \mathbb{Z}^d.$$

So for $u \in B(n_x)$ and $v \in B(x, n_x)$, as $2\|x\|_\infty \ge \|u - v\|_\infty \ge \|x\|_\infty/4$,

$$\mathbb{P}_p(d_{\mathsf{C}}(u,v) > 2\rho\|x\|_\infty,\ u, v \in \mathsf{C}) \le C_4 e^{-(C_3/4)\|x\|_\infty},$$

and by a union bound,

$$\begin{aligned} &\mathbb{P}_p(d_{\mathsf{C}}(u,v) > 2\rho\|x\|_\infty \text{ for some } u \in B(n_x) \cap \mathsf{C},\ v \in B(x,n_x) \cap \mathsf{C}) \\ &\qquad\qquad \le C_5 e^{-C_6\|x\|_\infty} \end{aligned} \tag{4.6}$$

for some $C_5, C_6 > 0$. On the complement of the event in (4.6), each $u \in B(n_x) \cap \mathsf{C}$ and $v \in B(x, n_x) \cap \mathsf{C}$ have $d_{\mathsf{C}}(u,v) \le 2\rho\|x\|_\infty$ and so $T(u,v) \le 2K\rho\|x\|_\infty$. Use this in the right side of (4.5) to bound it above by

$$C_7 e^{-C_8\|x\|_\infty^{1/d}} + \mathbb{P}\left(\begin{array}{l} \exists\, u \in B(n_x), v \in B(x,n_x) \text{ with } |\textsc{Geo}(u,v)| \text{ at} \\ \text{least } (M - C_2)\|x\|_\infty \text{ but } T(u,v) \le 2K\rho\|x\|_\infty \end{array}\right). \tag{4.7}$$

Last, we appeal to Lemma 4.5 to find constants $a, C_9 > 0$ such that for all $n \ge 1$,

$$\begin{aligned} &\mathbb{P}\Big(\exists \text{ self-avoiding } \Gamma \text{ starting at } 0 \text{ with } |\Gamma| \ge n \text{ but with } T(\Gamma) < an\Big) \\ &\qquad\qquad \le e^{-C_9 n}. \end{aligned} \tag{4.8}$$

By a union bound, for all $n \ge 1$ and $x \in \mathbb{Z}^d$ with $\|x\|_\infty \ge C'$,

$$\begin{aligned} &\mathbb{P}\Big(\exists \text{ self-avoiding } \Gamma \text{ from } B(n_x) \text{ to } B(x,n_x) \text{ with } |\Gamma| \ge n \text{ but } T(\Gamma) < an\Big) \\ &\qquad\qquad \le C_{10} n_x^d e^{-C_9 n}. \end{aligned}$$

So fixing any M with $2K\rho/a + C_2 < M$, the expression in (4.7) is bounded above by

$$C_7 e^{-C_8\|x\|_\infty^{1/d}} + C_{10} n_x^d e^{-C_9(M-C_2)\|x\|_\infty},$$

and this is bounded above by $C_{11} e^{-C_{12}\|x\|_\infty^{1/d}}$ for some $C_{11}, C_{12} > 0$. Increasing M if necessary, we obtain $e^{-C_{13}\|x\|_\infty^{1/d}}$, completing the proof. $\square$

We now show a similar bound for the cardinality of $\overline{\textsc{Geo}}$. It turns out that, similarly to the bound in Theorem 4.9, the union of geodesics to a fixed x generally consists of order $\|x\|_\infty$ number of edges.

THEOREM 4.12. *Let I be the infimum of the support of F, and suppose $F(I) < p_c$, with $\mathbb{E}\min\{t_1, \ldots, t_{2d}\} < \infty$, where the t_i's are independent copies of τ_e. There exists a constant C such that, for all $x \in \mathbb{Z}^d$ with $|x| > 0$, we have*

$$\mathbb{E}\left|\overline{\textsc{Geo}}(0,x)\right| \le C\|x\|_\infty.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

The main idea of the proof is to show that many edges of $\overline{\text{GEO}}$ are either elements of $\underline{\text{GEO}}$ or would become elements of $\underline{\text{GEO}}$ when their edge weights are reduced. The structure of this reduction or modification argument is why we obtain an upper bound in expectation and not control of higher moments as in Theorem 4.9.

QUESTION 4.1.3. Do there exist constants C, M such that

$$\mathbb{P}(\overline{\text{GEO}}(0,x) \geq M\|x\|_\infty) \leq e^{-C\|x\|_\infty}?$$

Similarly, can one get upper bounds for $\mathbb{E}\overline{\text{GEO}}(0,x)^\alpha$ of the order $\|x\|_\infty^\alpha$ for positive $\alpha \neq 1$?

PROOF. We first show that a positive density of edges in $\overline{\text{GEO}}(0,x)$ must take large values. Fix a value $\delta > 0$ such that $F(I + 2\delta) < p_c$. Define the set N_x of high-weight edges in $\overline{\text{GEO}}(0,x)$:

$$N_x := \{e \in \overline{\text{GEO}}(0,x) : \tau_e > I + 2\delta\}.$$

We first claim that there exists some $c > 0$ such that, for all x, we have

$$\mathbb{P}\left(|N_x| \leq c\left|\overline{\text{GEO}}(0,x)\right|\right) \leq \frac{1}{c}\exp(-c\|x\|_\infty). \tag{4.9}$$

Equation (4.9) is in spirit similar to Kesten's estimate in Lemma 4.5 above and is in fact nearly identical to [**138**, Theorem 5], where the corresponding statement is made in the case where weights are assigned to vertices instead of edges. Indeed, the proof of the statement from [**138**] is easily adapted to show (4.9), so we will not give a detailed argument here.

Using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}|\overline{\text{GEO}}(0,x)| &\leq \frac{1}{c}\mathbb{E}|N_x| + \mathbb{E}|\overline{\text{GEO}}(0,x)|\mathbf{1}_{\{|\overline{\text{GEO}}(0,x)| \geq \frac{1}{c}|N_x|\}} \\
&\leq \frac{\mathbb{E}|N_x|}{c} + \left[C\mathbb{E}\text{diam}(\overline{\text{GEO}}(0,x))^{2d}\mathbb{P}\left(|N_x| \leq c|\overline{\text{GEO}}(0,x)|\right)\right]^{\frac{1}{2}}.
\end{aligned}$$

Using the geodesic diameter bound (4.4), it follows from (4.9) that there exists a uniform $c' > 0$ such that

$$\mathbb{E}|N_x| \geq c'\mathbb{E}|\,\overline{\text{GEO}}(0,x)|. \tag{4.10}$$

In other words, a typically not too small fraction of edges in $\overline{\text{GEO}}(0,x)$ have weights above $I + 2\delta$.

We next show that if $\tau_e > I + 2\delta$, we can "modify" τ_e to take a value below $I + \delta$ with constant probability, and that $e \in \underline{\text{GEO}}(0,x)$ in the resulting configuration. Consider the event

$$\overline{G}_e = \{e \in \overline{\text{GEO}}(0,x),\, \tau_e > I + 2\delta\}.$$

Define $\widehat{G}_e$ to be the event (measurable with respect to $\{\tau_f\}_{f \neq e}$) that e will be in $\overline{\text{GEO}}(0,x)$ for some particular value of τ_e greater than $I + 2\delta$. We note that $\overline{G}_e \subseteq \widehat{G}_e \cap \{\tau_e > I + 2\delta\}$, so by independence we have

$$\begin{aligned}
\mathbb{E}|N_x| = \sum_e \mathbb{P}(\overline{G}_e) &\leq \sum_e \mathbb{P}(\widehat{G}_e,\, \tau_e > I + 2\delta) \\
&= [1 - F(I + 2\delta)]\sum_e \mathbb{P}(\widehat{G}_e).
\end{aligned} \tag{4.11}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

We note the following "monotonicity" property of geodesics:

- Suppose (τ_f) is a configuration of edge-weights in which $e \in \overline{\text{Geo}}(0,x)$, and suppose (τ_f') is a configuration such that $\tau_f = \tau_f'$ for all $f \neq e$; suppose that $\tau_e > \tau_e'$. Then $e \in \underline{\text{Geo}}(0,x)$ in the configuration (τ_f').

An immediate consequence of this observation is the fact that

$$\left(\widehat{G}_e \cap \{\tau_e \leq I + \delta\}\right) \subseteq \{e \in \underline{\text{Geo}}(0,x)\}. \tag{4.12}$$

We express (4.12) as an inequality of indicator functions, take expectations, and sum over e to find

$$F(I+\delta) \sum_e \mathbb{P}\left(\widehat{G}_e\right) \leq \mathbb{E}|\,\underline{\text{Geo}}(0,x)|.$$

Combining the last display with (4.11) and (4.10) gives

$$\mathbb{E}|\,\underline{\text{Geo}}(0,x)| \geq c' F(I+\delta)\mathbb{E}|\,\overline{\text{Geo}}(0,x)|.$$

On the other hand, by Lemma 2.3 and Proposition 4.7, we have under our moment assumption that $\mathbb{E}|\,\underline{\text{Geo}}(0,x)| \leq C_4 \|x\|_\infty$, which completes the proof. $\square$

Given the last three results, which give linear bounds for the size of geodesics, one may ask if the length of the geodesic from 0 to x, when divided by $\|x\|$ converges. This appears to be a difficult problem and is only solved in some special cases.

Question 4.1.4. Show that for continuous F, one has for fixed nonzero $x \in \mathbb{Z}^d$

$$\lim_{n\to\infty} \frac{|\text{Geo}(0,nx)|}{n} \quad \text{exists a.s.}$$

In [**190**], Zhang-Zhang considered edge-weight distributions in two dimensions satisfying $F(0) > p_c$ (so that zero-weight edges form infinite clusters) and defined

$$N_{0,x} = \min\{|\Gamma| : \Gamma \text{ is a geodesic from } 0 \text{ to } x\}.$$

Their main result is that $N_{0,ne_1}/n$ converges a.s.. This theorem was extended by Zhang [**182**, Theorem 2] to all dimensions:

Theorem 4.13 (Zhang-Zhang [**190**], Zhang [**182**]). *Suppose that $F(0) > p_c$ and $d \geq 1$. Then*

$$\lim_{n\to\infty} \frac{1}{n} N_{0,ne_1} \quad \textit{exists a.s. and in } L^1.$$

Although the result is stated for geodesics (and point-to-plane geodesics in [**182, 190**]) in direction e_1, the argument can also be used to show convergence of $N_{0,nx}/n$ a.s. and in L^1 for any $x \in \mathbb{Z}^d$. The proof of the above theorem uses the subadditive ergodic theorem but also the condition $F(0) > p_c$ in an important way. It is an open problem to extend the theorem to distributions satisfying $F(0) < p_c$. (As remarked above Theorem 4.8, it does not hold for $F(0) = p_c$.) Such a result would have implications for "shift differentiability" of the time constant. See [**164**] for more on this connection.

On several occasions we will be interested in the case where $\overline{\text{Geo}}(0,x)$ is a single path from 0 to x.

Definition 4.14. A metric space is called uniquely geodesic if between any two points there exists a unique geodesic.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

A necessary and sufficient condition for our space to be uniquely geodesic a.s. is that the edge-weights have a continuous distribution. Given the existence of finite geodesics, one can define the geodesic graph as follows. For each $x \in \mathbb{Z}^d$ we define the (directed) graph $\mathcal{T}(x)$ with vertex set $\mathbb{Z}^d$ and whose edge set is the collection of all directed edges e that belong to some finite geodesic between x and another point y, oriented as they are traversed moving from x to y. This graph has the following properties a.s.:

(1) $\mathcal{T}(x)$ is connected.
(2) Every finite directed path in $\mathcal{T}(x)$ is a finite geodesic.
(3) If the space is uniquely geodesic, then $\mathcal{T}(x)$ is a spanning tree of $\mathbb{Z}^d$. In this case, $\mathcal{T}(x)$ is called the geodesic tree or the tree of infection of x.

Many of the geodesic questions discussed in this book can be rephrased in the language of geodesic trees. Many of the "large scale" questions — for instance, the number of ends of the tree $\mathcal{T}(x)$ — will require the study of infinite geodesics, which we will define shortly. Before discussing these large scale questions, let us focus on a natural, more local question: how straight is the geodesic between x and y?

4.2. The wandering exponent ξ

A natural asymptotic question for geodesics is: what does $\overline{\text{Geo}}(0,x)$ "typically" look like for $|x|$ large? Since the first-passage metric is in some sense a random perturbation of the ℓ^1 metric, one could perhaps imagine that $\overline{\text{Geo}}(0,x)$ often stays order $|x|$ distance from the midpoint $x/2$. On the other hand, as discussed in Chapter 2, it is reasonably well-accepted that the limit shape should have uniformly positive curvature, suggesting that $\overline{\text{Geo}}(0,x)$ should perhaps look more like the straight line between 0 and x, as in the case of the ℓ^2 metric.

Indeed, the ℓ^2 intuition is believed to be the correct one, and $\overline{\text{Geo}}(0,x)$ is believed to lie close to the line

$$L_x = \{mx : m \in \mathbb{R}\}.$$

It is easy to see very weak versions of this claim under the assumption that the limit shape is strictly convex. For instance, if $x = x(n) = (n, n, \ldots, n)$, and we denote by Γ_n the path which first takes n steps down the e_1 axis to connect 0 to ne_1, then moves n steps parallel to the e_2 axis, etc., we see by the shape theorem that

$$\lim_{n\to\infty} T(\Gamma_n)/n = d\mu(e_1).$$

On the other hand, $\lim_{n\to\infty} T(0,x(n))/n = \mu(e_1+\ldots+e_d)$, which is strictly smaller than $d\mu(e_1)$ if the limit shape is strictly convex. In particular, Γ_n is not only not the geodesic to $x(n)$, but actually pays order $|x|$ more in passage time than the geodesic does. This simple argument only rules out a single path, however, and stronger arguments have been developed to simultaneously control all candidate geodesics.

Set $D(0,x)$ to be the maximal distance of any point in $\overline{\text{Geo}}(0,x)$ to L_x. An early argument showing that $\overline{\text{Geo}}(0,x)$ lies close to L_x (under assumptions on the limit shape) appeared in [**123**, Lemma 9.10], where it was shown that, assuming that the limit shape has an exposed point in the e_1 direction,

$$\lim_{\epsilon\to 0} \liminf_{n\to\infty} \mathbb{P}\left(D(0,ne_1) < \epsilon n\right) = 1. \tag{4.13}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

To discuss strengthening (4.13), we will introduce the "wandering exponent" ξ. There is no improvement to (4.13) that does not make unproven assumptions on the limit shape, so we pose the following.

QUESTION 4.2.1. Show that for continuous F, one has

$$\lim_{n\to\infty} \frac{\mathbb{E}D(0,ne_1)}{n} = 0.$$

Roughly speaking, ξ is the number such that the maximal distance between $\overline{\text{GEO}}(0,x)$ and L_x is typically of order $|x|^\xi$. A simple attempt at giving a concrete definition could be to set $\epsilon = n^{\alpha-1}$ in the left-hand side of (4.13) and define ξ to be the smallest number α such that (4.13) still holds; however, there is no general agreement on the "correct" definition, and we will discuss several possible alternatives. The exponent ξ is believed to be universal in the sense that, for all reasonable choices of continuous distributions F, one should find the same value of ξ, and ξ should govern geodesic fluctuations in all directions. In general, however, ξ should depend on the dimension d. It is widely believed that, for $d = 2$, the value of ξ should be 2/3 [**116**,**117**,**120**,**121**]. For distributions in $\mathcal{M}_p$ with $p > \vec{p}_c$, defined in Section 2.5, infinite oriented paths of weight-one edges give $\xi = 1$ — and $\chi = 0$ — for directions in the percolation cone.

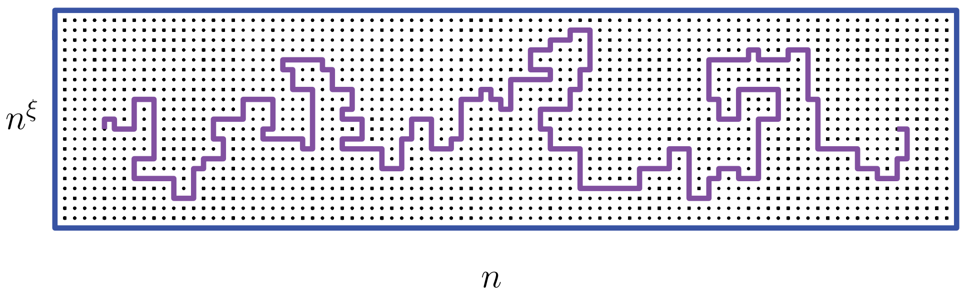


FIGURE 8. Illustration of geodesic wandering behavior on scale n^ξ. The deviation of a geodesic from 0 to ne_1 from the straight line that connects these points is of order n^ξ.

4.2.1. Upper bounds on the wandering exponent ξ. We now discuss the most robust current upper bounds, from [**147**,**149**].

Letting $\mathcal{C}(x,m)$ denote the cylinder with axis $L_x = \{mx : m \in \mathbb{R}\}$ and radius m; that is, using dist for the Euclidean distance,

$$\mathcal{C}(x,\, m) = \{z \in \mathbb{R}^d : \operatorname{dist}(z,\, L_x) \le m\}, \tag{4.14}$$

then one expects $\mathcal{C}(x,|x|^\xi)$ is the "right scale" of cylinder to contain $\overline{\text{GEO}}(0,x)$. To be more precise, one expects

$$\lim_{|x|\to\infty} \mathbb{P}\left(\overline{\text{GEO}}(0,x) \subseteq \mathcal{C}(x,|x|^{\xi+\epsilon})\right) = 1, \tag{4.15}$$

$$\lim_{|x|\to\infty} \mathbb{P}\left(\overline{\text{GEO}}(0,x) \subseteq \mathcal{C}(x,|x|^{\xi-\epsilon})\right) = 0, \tag{4.16}$$

for any $\epsilon > 0$. (See Figure 8.)

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Note that (4.15) and (4.16) carry an implicit isotropy; if these equations are true, then the order of geodesic wandering must be the same in all directions. More should be true: if we replace $\overline{\text{GEO}}(0,x)$ by, for instance, the union of all geodesics to a hyperplane at distance n, then this union should asymptotically be contained in cylinders of radius $n^{\xi+\epsilon}$ (and not in cylinders of radius $n^{\xi-\epsilon}$). Lastly, "near-geodesics" should have the same behavior: if $\overline{\text{GEO}}(x,y;M)$ is the union of all paths Γ connecting x to y and satisfying $T(\Gamma) \leq T(x,y)+M$, then $\overline{\text{GEO}}(x,y;M)$ should satisfy relations like (4.15) and (4.16).

The article [**149**], in conjunction with its studies of χ, provides a technique for bounding ξ in terms of χ. (A more detailed discussion of the relation between the exponents can be found in the next section.) For a unit vector x, let

$$\xi_x = \inf\left\{\alpha > 0 : \liminf_n \mathbb{P}\big(\overline{\text{GEO}}(0,nx) \subset \mathcal{C}(x,n^\alpha)\big) > 0\right\}.$$

We also define the fluctuation exponent

$$\chi' = \inf\left\{\kappa :\ (t-t^\kappa)\mathcal{B} \subseteq B(t) \subseteq (t+t^\kappa)\mathcal{B} \text{ for large } t, \text{ a.s.}\right\}.$$

Note that χ' is measuring not only random but also nonrandom fluctuations; that is, it measures the deviation of T from μ, not from $\mathbb{E}T$. The following inequality is a version of one of the inequalities of the scaling relation $\chi = 2\xi - 1$, discussed in Section 4.3 below. In their paper, Newman-Piza explain that their argument is essentially a rigorous version of the argument of Krug-Spohn [**133**] for this scaling relation.

We say that the unit vector x is a direction of curvature for $\mathcal{B}$ if the lower bound in (2.28) holds in direction x with $\kappa = 2$; that is, if there is a hyperplane H_0 through the origin such that $H_0 + x/\mu(x)$ is a supporting hyperplane for $\mathcal{B}$ and there are constants $\epsilon, c > 0$ such that

$$\mu(x+z) - \mu(x) \geq c|z|^2$$

whenever $z \in H_0$ satisfies $|z| < \epsilon$. Note that, for any limit shape $\mathcal{B}$, there is at least one direction of curvature. Take a ball of radius r large enough that contains $\mathcal{B}$ and decrease r to the first time it intersects $\partial\mathcal{B}$. The directions of contact of the ball with $\mathcal{B}$ are then directions of curvature for $\mathcal{B}$. This argument appeared in [**149**].

THEOREM 4.15 (Newman-Piza [**149**], Theorem 6). *Assume $d \geq 2$ with $\mathbb{E}\tau_e^2 < \infty$ and $F(0) < p_c$. Assume that x is a direction of curvature for the limit shape; then*

$$\xi_x \leq \frac{1+\chi'}{2}.$$

A sketch of the argument that leads to Theorem 4.15 is given below, after Theorem 4.16. In the same paper, under the assumptions of Theorem 3.8, the authors also obtained the following bound, which they refer to as the Wehr-Aizenman bound, after the latter's work on phase transitions in disordered systems [**174**]:

$$\chi_x \geq \frac{1-(d-1)\xi_x}{2}, \tag{4.17}$$

where

$$\chi_x = \sup\left\{\alpha > 0 :\ \text{for some } C > 0,\ \operatorname{Var} T(0,nx) \geq Cn^{2\alpha} \text{ for all } n \geq 1\right\}.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

In dimension two, Theorem 4.15 can be combined with the previous bound $\chi' \leq 1/2$, derived by Kesten [**125**] and Alexander [**7**] (see Theorem 3.33). One still assumes that x is a direction of curvature of the limit shape. The result is the following.

THEOREM 4.16 (Newman-Piza [**149**], Theorem 7). *Assume $F(0) < p_c$, and that $\mathbb{E}e^{\alpha\tau_e} < \infty$ for some $\alpha > 0$. Then, if x is a direction of curvature for the limit shape $\mathcal{B}$,*

$$\xi_x \leq 3/4 \quad \text{and} \quad \chi_x \geq 1/8,$$

the latter bound holding when $d = 2$.

Without the curvature assumption on the limit shape the best current bounds are still $0 \leq \xi \leq 1$ (see Question 4.2.1), the upper bound following directly from the shape theorem.

We now give a sketch of the argument for Theorem 4.15. Rather than prove this inequality, we show a more intuitive Newman-Piza type argument under stronger assumptions, like those made in Chatterjee's work [**51**]. The argument also explains the inequality $\xi_x \leq 3/4$ from Theorem 4.16. Suppose that χ is any number for which there is $c > 0$ such that

$$\mathbb{P}(|T(0,z) - \mu(z)| \geq \lambda |z|^{\chi}) \leq e^{-c\lambda} \text{ for all } z \in \mathbb{Z}^d \text{ and } \lambda \geq 0. \tag{4.18}$$

(In fact, it is only necessary here to make a similar assumption for deviations of $T(0,z)$ about its mean, not $\mu(z)$, due to Remark 3.21). We aim now to show that if x is a direction of curvature, then

$$\xi_x \leq \frac{1+\chi}{2}. \tag{4.19}$$

Since (4.18) is known to hold for $\chi = 1/2 + \delta$ for any $\delta > 0$ (see Theorem 3.10 and Remark 3.21), one obtains $\xi_x \leq 3/4$.

SKETCH OF PROOF OF (4.19). The version of the Newman-Piza argument we present here is from Chatterjee [**51**]. For simplicity, take $x = e_1$. Let $\epsilon > 0$ and $\hat{\xi} = \frac{1+\chi}{2} + \epsilon$. Note that since χ can be chosen < 1, one has $\hat{\xi} < 1$. Let L be the e_1-axis, and for $n \geq 1$, let $L' = L'(n)$ be the segment connecting 0 and ne_1. Now define

$$V = \{v \in \mathbb{Z}^d : \exists\ w \in L' \text{ such that } |w - v| \in [n^{\hat{\xi}}, 2n^{\hat{\xi}}]\}.$$

Then we must show that the probability that a geodesic from 0 to ne_1 enters V goes to 0 with n. This will show that $\xi_x \leq \hat{\xi}$ and, taking $\epsilon \to 0$, we obtain the result.

By a union bound, the probability that a geodesic from 0 to ne_1 enters V is bounded by

$$\begin{aligned} &\sum_{v \in V} \mathbb{P}(v \text{ in a geodesic from } 0 \text{ to } ne_1) \\ = &\sum_{v \in V} \mathbb{P}(T(0, ne_1) = T(0,v) + T(v, ne_1)). \end{aligned}$$

The goal now is to show that the sum on the right is small as $n \to \infty$. We claim there is $c > 0$ such that for all n and all $v \in V$,

$$\mathbb{P}(T(0,ne_1) = T(0,v) + T(v,ne_1)) \leq (1/c) \exp\left(-cn^c\right). \tag{4.20}$$

After summing over v and taking $n \to \infty$, we complete the proof.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

The main step is to show that, due to the curvature assumption, there is C such that for all $v \in V$,

$$\mu(v) + \mu(ne_1 - v) - \mu(ne_1) \geq Cn^{2\hat{\xi}-1} = Cn^{\chi+2\epsilon}. \tag{4.21}$$

Note that this inequality implies that if we replace T by μ in the event $\{T(0,v) + T(v,ne_1) = T(0,ne_1)\}$, then the event cannot occur and, furthermore, gives us a large lower bound on the difference of both sides. Assuming this inequality for the moment, on the event $\{T(0,v) + T(v,ne_1) = T(0,ne_1)\}$, one has

$$0 \geq [T(0,v) - \mu(v)] + [T(v,ne_1) - \mu(ne_1 - v)] - [T(0,ne_1) - \mu(ne_1)] + Cn^{\chi+2\epsilon},$$

and so at least one of the first three terms on the right has absolute value at least $(C/3)n^{\chi+2\epsilon}$. Applying (4.18) would give (4.20) and complete the proof.

Inequality (4.21) is straightforward in the case that $v = (n/2)e_1 + n^{\hat{\xi}}e_2$: for illustration, in this case, we use curvature and symmetry for

$$\begin{aligned}
\mu(v) + \mu(ne_1 - v) - \mu(ne_1) &= [\mu(v) - \mu(n/2e_1)] + [\mu(ne_1 - v) - \mu(n/2e_1)] \\
&= 2[\mu(v) - \mu(n/2e_1)] \\
&= n\left[\mu\left(e_1 + \frac{2n^{\hat{\xi}}}{n}e_2\right) - \mu(e_1)\right] \\
&\geq Cn\left(\frac{n^{\hat{\xi}}}{n}\right)^2 \\
&= Cn^{\chi+2\epsilon}.
\end{aligned}$$

To modify this argument for other v's, we split into cases. Let H be the hyperplane taken in the definition of curvature (which in the case $x = e_1$ is just orthogonal to e_1 by symmetry), and set w to be the projection of v onto L along H. Note that convexity of the limit shape implies that $\mu(v) \geq \mu(w)$ and $\mu(ne_1 - v) \geq \mu(ne_1 - w)$. Furthermore, at least one of the two vectors v or $ne_1 - v$ has Euclidean norm at least $n/2$.

(1) Case 1: $w \in L'$. In this case, we may assume that v is the vector with norm at least $n/2$, and use curvature as above:

$$\begin{aligned}
\mu(v) - \mu(w) = |w|\left(\mu\left(\frac{w}{|w|} + \frac{v-w}{|w|}\right) - \mu\left(\frac{w}{|w|}\right)\right) &\geq C|w|\left|\frac{v-w}{|w|}\right|^2 \\
&\geq Cn^{2\hat{\xi}-1}.
\end{aligned}$$

Therefore, since $w \in L'$, one has $\mu(ne_1 - w) + \mu(w) = \mu(ne_1)$, and so

$$\begin{aligned}
\mu(v) + \mu(ne_1 - v) - \mu(ne_1) &= [\mu(v) - \mu(w)] + [\mu(ne_1 - v) - \mu(ne_1 - w)] \\
&\geq \mu(v) - \mu(w) \\
&\geq Cn^{2\hat{\xi}-1}.
\end{aligned}$$

(2) Case 2: $w \in L \setminus L'$, on the side closer to ne_1. In this case we still have as above

$$\mu(v) - \mu(w) \geq C|w|\left|\frac{v-w}{|w|}\right|^2.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Now

$$n^{2\hat{\xi}} \le |v - ne_1|^2 = (|w| - n)^2 + |v - w|^2 = (|w| - n)^2 + |w|^2 \left| \frac{v - w}{|w|} \right|^2,$$

and $n \le |w| \le 3n$. Therefore either $|w|^2 \left|\frac{v-w}{|w|}\right|^2 > n^{2\hat{\xi}}/2$, or $|w - ne_1| = |w| - n \ge n^{\hat{\xi}}/\sqrt{2}$. In the first case,

$$\begin{aligned} \mu(v) + \mu(ne_1 - v) - \mu(ne_1) \ge \mu(v) - \mu(w) &\ge C|w|^2 \left| \frac{v - w}{|w|} \right|^2 |w|^{-1} \\ &\ge Cn^{2\hat{\xi}-1}, \end{aligned}$$

and in the second case, since $\hat{\xi} < 1$,

$$\mu(v) + \mu(ne_1 - v) - \mu(ne_1) \ge \mu(w) - \mu(ne_1) \ge C|w - ne_1| \ge Cn^{\hat{\xi}}$$

is still at least $Cn^{2\hat{\xi}-1}$.

(3) Case 3: $w \in L \setminus L'$, on the side closer to 0. This case is similar to Case 2.

These three cases together prove (4.21) and complete the proof. □

4.2.2. Licea-Newman-Piza lower bounds on the wandering exponent ξ. We turn to the question of lower bounds for the wandering exponent ξ, taken up in [**140**]. The lower bounds proved in [**140**] all depend on the particular definition of ξ given. We note a basic guideline for the possible lower bounds expected. As discussed earlier, one expects $\xi = 2/3$ when $d = 2$, and, in view of the scaling relation $\chi = 2\xi - 1$ discussed in Section 4.3 below, one expects $\xi \ge 1/2$ for all d.

We are now ready to present the results of [**140**]. They consider four different definitions of ξ (in their words, proving these definitions are equivalent is "one of the open foundational problems of the subject"). Their first definition, $\xi^{(0)}$, is the only point-to-point exponent considered:

$$\xi^{(0)} := \sup \left\{ \alpha \ge 0 : \limsup_{|x| \to \infty} \mathbb{P}\left(\overline{\text{GEO}}(0, x) \subseteq \mathcal{C}(x, |x|^\alpha)\right) < 1 \right\},$$

where $\mathcal{C}$ is as in (4.14); compare (4.16). The first result of [**140**] is

THEOREM 4.17 (Licea-Newman-Piza [**140**], Theorem 1). *If $d \ge 2$, then $\xi^{(0)} \ge 1/(d+1)$.*

The lower bound in Theorem 4.17 is extended in [**140**, Theorem 1] to $\xi^{(1)}$, the second version of ξ defined in that paper; we omit it here for brevity.

All the remaining bounds in [**140**] are for point-to-hyperplane versions of ξ, so we will need notation for hyperplanes. For $\hat{\theta}$ a unit vector of $\mathbb{R}^d$, let

$$\Lambda(\hat{\theta}, L) := \{y \in \mathbb{Z}^d : y \cdot \hat{\theta} < L\}$$

be the corresponding family of half-spaces. Let $\partial\Lambda(\hat{\theta}, L)$ be the set of $y \in \mathbb{Z}^d \setminus \Lambda(\hat{\theta}, L)$ adjacent to some vertex of $\Lambda(\hat{\theta}, L)$; the $\partial\Lambda$ are the families of hyperplanes to which we will consider geodesics.

For the next definition of ξ, let $\overline{\text{GEO}}(x, A; M)$ be the union of "near-geodesics" from x to a set A as below (4.16). Let end$[\overline{\text{GEO}}(x, A; M)]$ be the set of endpoints

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

in A of these near-geodesics.

$$\begin{aligned}\xi^{(2)}(M) = \sup \Big\{ \alpha \geq 0 : \exists (\hat{\theta}_n)_n,\ (J_n)_n \text{ with } J_n \to \infty \text{ such that} \\ \text{there is no deterministic } A_n \text{ with } \operatorname{diam}(A_n) \leq J_n^\alpha \\ \text{such that } \mathbb{P}\left(\operatorname{end}[\overline{\text{GEO}}(0, \partial\Lambda(\hat{\theta}_n, J_n); M)] \subseteq A_n\right) \to 1 \Big\}.\end{aligned}$$

Note that in some respects, this definition is considerably weaker than the definition of $\xi^{(0)}$: it concerns only near-geodesics, and it concedes a great deal of uniformity over direction and "scale" (J_n). On the other hand, it is stronger in the sense that we actually know where the wandering occurs (at the endpoints; in the defintion of $\xi^{(0)}$ it is not specified). The second theorem of [**140**] is, recalling that I is the infimum of the distribution of τ_e,

Theorem 4.18 (Licea-Newman-Piza [**140**], Theorem 2). *Let $M > I$ and $d \geq 2$. Then $\xi^{(2)}(M) \geq 1/2$.*

The last definition of ξ considered has in some ways all the weaknesses of both $\xi^{(0)}$ and $\xi^{(2)}$:

$$\begin{aligned}\xi^{(3)}(M) := \sup \Big\{ \alpha \geq 0 : \exists (\hat{\theta}_n)_n,\ (J_n)_n \text{ with } J_n \to \infty \text{ such that} \\ \text{there is no deterministic } x_n \text{ such that} \\ \mathbb{P}\left(\overline{\text{GEO}}(0, \partial\Lambda(\hat{\theta}_n, J_n); M) \subseteq \mathcal{C}(x_n,\ J_n^\alpha)\right) \to 1 \Big\}.\end{aligned}$$

Theorem 4.19 (Licea-Newman-Piza [**140**], Theorem 3). *Let $M > I$ and $d = 2$. Then $\xi^{(3)}(M) \geq 3/5$.*

We will attempt to justify the above results without providing complete proofs. In some cases, unessential details or complications will be elided for presentation; for complete arguments, see [**140**].

Proof of Theorem 4.17. The idea is an extension of the variance estimate of Newman-Piza [**149**] discussed in Section 3.3. This idea was later used in Wüthrich [**179**], Johansson [**119**], and Auffinger-Damron [**18**]. Let $\alpha > \xi^{(0)}$ and choose some sequence of vertices x_n with $|x_n| \to \infty$ such that

$$\mathbb{P}(\overline{\text{GEO}}(0, x_n) \subseteq \mathcal{C}(x_n,\ |x_n|^\alpha)) \to 1. \tag{4.22}$$

We will need to define a shifted version of x_n; let $x_n^\perp$ be a unit vector perpendicular to $x_n/|x_n|$ and define $0'$ and x_n' to be the closest vertices to $0 + C|x_n|^\alpha x_n^\perp$ and $x_n + C|x_n|^\alpha x_n^\perp$. where C is some large constant independent of n.

Define $\Delta T := T(0, x_n) - T(0', x_n')$; the proof is by analysis of ΔT. Note that ΔT cannot be too large in magnitude. Indeed, there are paths Γ_1 and Γ_2 connecting 0 to $0'$ and x_n to x_n' respectively, such that $|\Gamma_i| \leq C|x_n|^\alpha$. Therefore,

$$|\Delta T| \leq T(\Gamma_1) + T(\Gamma_2) \lesssim C|x_n|^\alpha \tag{4.23}$$

where the precise strength of the bound is a function of the tail behavior of τ_e (if τ_e is bounded, then the left-hand side is really $\leq C|x_n|^\alpha$).

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

On the other hand, we can get a lower bound on $\operatorname{Var}\Delta T$ using the machinery of Newman-Piza. Defining as before the sigma-algebra Σ_j generated by the first j edge weights in some enumeration $(e(j))$ of the edges, we have

$$\operatorname{Var}\Delta T = \sum_{j=1}^{\infty} \mathbb{E}\left[\mathbb{E}\left[\Delta T \mid \Sigma_j\right] - \mathbb{E}\left[\Delta T \mid \Sigma_{j-1}\right]\right]^2. \tag{4.24}$$

As in (3.17), one can provide a lower-bound for (4.24) in terms of influences of individual edges. There is an extra complication: when decreasing an edge weight $\tau_{e(i)}$, we are only guaranteed to decrease ΔT if e_i simultaneously satisfies

$$e(i) \in \overline{\mathrm{GEO}}(0, x_n) \text{ and } e(i) \notin \overline{\mathrm{GEO}}(0', x_n'). \tag{4.25}$$

Letting $\overline{F}_i$ be the event described in (4.25), we have similarly to before

$$\operatorname{Var}(\Delta T) \geq C \sum_{j=1}^{\infty} \mathbb{P}(\overline{F}_j)^2. \tag{4.26}$$

For a further lower bound, we can restrict the sum in (4.26) to j such that $e(j) \in \mathcal{C}(x_n, |x_n|^\alpha)$. For this restricted set $\mathfrak{J}$ of indices, we use the choice of α for

$$\begin{aligned}\sum_{j\in\mathfrak{J}} \mathbb{P}(\overline{F}_j) &\geq \mathbb{E}\left|\overline{\mathrm{GEO}}(0, x_n) \cap \mathcal{C}(x_n, |x_n|^\alpha)\right| - \mathbb{E}\left|\overline{\mathrm{GEO}}(0', x_n') \cap \mathcal{C}(x_n, |x_n|^\alpha)\right| \\ &\geq c|x_n|\end{aligned}$$

for some $c > 0$ and all n large (the last inequality holds by (4.22)). As in (3.18), using Jensen's inequality on (4.26), we see

$$\begin{aligned}\operatorname{Var}\Delta T \geq \sum_{j\in\mathfrak{J}} \mathbb{P}(\overline{F}_j)^2 &\geq \frac{1}{|\mathcal{C}(x_n, |x_n|^\alpha)|}\left(\sum_{j\in\mathfrak{J}} \mathbb{P}(F_j)\right)^2 \\ &\geq C|x_n|^{1-\alpha(d-1)}.\end{aligned} \tag{4.27}$$

Comparing this inequality with (4.23) (which gives a variance upper bound of order $|x|^{2\alpha}$) gives $1 - \alpha(d-1) \leq 2\alpha$ for any $\alpha > \xi^{(0)}$. Thus $\alpha \geq 1/(d+1)$, and taking $\alpha \downarrow \xi^{(0)}$ completes the proof. $\square$

Proof of Theorem 4.18. The proof of this theorem relies on a geometrical construction. Rather than trying to exactly give the parameters of this construction, we will outline the main idea; we leave it to the reader to fill in the details (and in particular to understand why $\xi^{(2)}$ is defined as it is). We first outline the argument in case $d = 2$, then explain the necessary changes for $d > 2$.

Assume $\xi^{(2)} < 1/2$, and choose $1/2 > \alpha > \xi^{(2)} + \epsilon$. We choose a finite set of unit vectors $\{\hat{\theta}_i\}_{i=0}^N$ and, for each L, construct a polygon P_L (having the symmetries of $\mathbb{Z}^2$) whose sides are segments of lines perpendicular to $\hat{\theta}_i$:

$$P_L = \bigcap_i \Lambda(\hat{\theta}_i, L_i), \text{ where } L_i \sim L.$$

(See Figure 9.) Let S_i be the side of P_L corresponding to $\hat{\theta}_i$ — that is, the side of P_L which is a segment of the boundary of $\Lambda(\hat{\theta}_i, L_i)$. We will choose $\hat{\theta}_i$ and L_i such that the lengths $|S_i| \sim L^\alpha$. We will also ensure our choice of $\hat{\theta}_i$ guarantees geodesics to the lines defining P_L are localized on sides of P_L.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

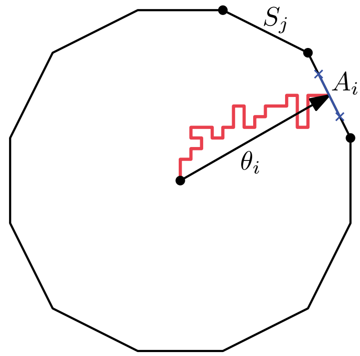


FIGURE 9. The construction of the polygon P_L. The segments A_i are close to the midpoints of the sides S_i and, with high probability, contain the endpoints of the geodesics from $\overline{\text{GEO}}(0, S_i)$.

Let A_i be a deterministic subset of $\partial\Lambda(\hat{\theta}_i, L)$ of diameter smaller than $L^{\xi^{(2)}+\epsilon}$ in which end $[\overline{\text{GEO}}(0, \partial\Lambda(\hat{\theta}_i, L))]$ is localized with high probability; such a subset exists by the definition of $\xi^{(2)}$. Our main restriction in choosing $\hat{\theta}_i$ is that A_i should be close to the midpoint of S_i, and in particular be at least distance $L^{\xi^{(2)}+\epsilon}$ from S_j, $j \neq i$. This is done inductively.

To start the induction, we can choose $\hat{\theta}_0 = e_1$ and $L_0 = L$, with $|S_0| = L^\alpha$. Then by the symmetries of the lattice, we can choose A_0 symmetric about the e_1-axis (and in particular localized near the midpoint of S_0). Our last vector $\hat{\theta}_N$ will be the "45 degree vector" $(\sqrt{2})^{-1}(e_1 + e_2)$; again, by the symmetries of the lattice, A_N will be chosen symmetric about $\hat{\theta}_N$. The remaining vectors will only be chosen to fill out the polygon in the arc of angles $[0, \pi/4]$; once this is done, we fill in the rest by symmetry.

Assuming $\hat{\theta}_0, \ldots, \hat{\theta}_{j-1}$ and the corresponding $\{L_i\}$ have been chosen correctly, let z_{j-1} be the counterclockwise endpoint of S_{j-1}. If z_{j-1} is within distance $\leq CL^\alpha$ from a point on the line $\{a\hat{\theta}_N : a \in \mathbb{R}\}$, then choose $\hat{\theta}_j = \hat{\theta}_N$ and L_j is chosen so that $z_{j-1} \in \partial\Lambda(\hat{\theta}_j, L_j)$. Otherwise, we will choose $|S_j| = L^\alpha$, and we need to find a direction $\hat{\theta}_j$ with $\arg\hat{\theta}_j \in (\arg\hat{\theta}_{j-1}, \pi/4)$ such that A_j is appropriately localized on S_j (once we choose $\hat{\theta}_j$, then L_j is fixed by the necessity of connecting to S_{j-1}).

The fact that such a choice is possible is somewhat subtle. If we were to try to choose $\hat{\theta}_j$ extremely close to $\hat{\theta}_{j-1}$, then we would fail. Indeed, A_{j-1} is close to the midpoint of S_{j-1}, and as $\hat{\theta}_j \to \hat{\theta}_{j-1}$, the side S_j is essentially a subsegment of the boundary of $\Lambda(\hat{\theta}_{j-1}, L_{j-1})$ and so A_j should be identical to A_{j-1}, which lies order L^α distance *below* the midpoint of S_j. Similarly, taking $\hat{\theta}_j \to \hat{\theta}_N$ would give an A_j lying too far *above* the midpoint. By a continuity argument, we can therefore find a $\hat{\theta}_j$ that meets our requirements.

After this somewhat lengthy construction, we can quickly conclude. Note that, because $|S_i| \leq CL^\alpha$, we have

$$\arg(\hat{\theta}_i) - \arg(\hat{\theta}_{i-1}) \leq CL^{\alpha-1}.$$

In particular, $\partial\Lambda(\hat{\theta}_{i-1}, L_{i-1})$ comes within distance $L^{2\alpha-1} = o(1)$ of the midpoint of S_i. But this means it takes only $O(1)$ edges to extend a time-minimizing path for

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

$T(0, \partial\Lambda(\hat{\theta}_i, L_i))$ from S_i to $\partial\Lambda(\hat{\theta}_{i-1}, L_{i-1})$ — since this extension costs very little in the way of passage time, there is a "near-geodesic" to $\partial\Lambda(\hat{\theta}_{i-1}, L_{i-1})$ lying on S_i, contradicting the placement of A_{i-1}. We conclude that α cannot be chosen $< 1/2$; that is, $\xi^{(2)} \geq 1/2$. This argument assumes that $T(0, S_i) < T(0, S_{i-1})$, which, without loss of generality, holds with positive probability.

The case $d \geq 3$ is similar. Now instead of using polygons P_L, we consider polygonal "barrels" of the form $P_L \times \mathbb{R}^{d-2}$, where P_L is constructed similarly to before. The arguments above may be adapted to these barrels without substantially different ideas.

□

Proof of Theorem 4.19. The proof is in some ways heavily motivated by the proofs of both of the preceding theorems. Consider $\alpha > \xi^{(3)}$. We construct a polygon P_L with sides $\{S_j\}$ similarly to the arguments in the proof of Theorem 4.18. For fixed j, we consider

$$\Delta T = T(0, \partial\Lambda(\hat{\theta}_{j+1}, L_{j+1})) - T(0, \partial\Lambda(\hat{\theta}_j, L_j)).$$

As before, $\partial\Lambda(\hat{\theta}_j, L_j)$ comes within distance $\sim L^{2\alpha-1}$ of every point on S_{j+1}, so geodesics to $\partial\Lambda(\hat{\theta}_j, L_j)$ may be extended to $\partial\Lambda(\hat{\theta}_{j+1}, L_{j+1})$ at an increase in passage time $\lesssim L^{2\alpha-1}$. This (combined with a symmetric argument in the other direction) gives an a priori bound

$$\mathrm{Var}(\Delta T) \lesssim L^{4\alpha-2}. \tag{4.28}$$

As in the proof of Theorem 4.17, we find an α-dependent lower bound which contradicts this when α is too small. The geodesics to both hyperplanes are confined with high probability in strips with radius L^α. These strips are largely disjoint for L large, and so there are typically order L many edges which are in $\overline{\mathrm{Geo}}(0, \partial\Lambda(\hat{\theta}_j, L_j))$ but not in $\overline{\mathrm{Geo}}(0, \partial\Lambda(\hat{\theta}_{j+1}, L_{j+1}))$.

With this in hand, we lower bound $\mathrm{Var}(\Delta T)$ with a martingale expansion as in Theorem 4.17; we ultimately arrive at an analogue of (4.27). As before, this yields

$$\mathrm{Var}(\Delta T) \geq CL^{1-\alpha}.$$

Comparing this bound with (4.28) completes the contradiction. □

4.3. The scaling relation $\chi = 2\xi - 1$

Consider a geodesic from the origin to a point x with passage time $T(0, x)$. One of the central questions in FPP (and in related models) is to establish the following statement. For dimension $d \geq 2$, there exists an intrinsic relation between the magnitude of deviation of $T(0, x)$ from its mean and the magnitude of deviation of the geodesic for $T(0, x)$ from a straight line joining 0 and x. This relation is universal; that is, it is independent of the dimension and of the law of the weights (as long they satisfy certain assumptions).

As discussed in Section 3.1, the fluctuations of the passage time $T(0, x)$ about $\mathbb{E}T(0, x)$ should be of order $|x|^\chi$, where χ is the fluctuation exponent. Analogously, the transversal fluctuation or wandering exponent ξ, studied in Section 4.2, measures the maximal Euclidean distance of a geodesic from 0 to x from the straight line that joins 0 to x. The intrinsic relation described above should be given as

$$\chi = 2\xi - 1. \tag{4.29}$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

As previously discussed, the existence and the "correct" definition of these exponents is part of the problem of establishing (4.29). Before we discuss the mathematical history behind (4.29) and current progress, let us stress that, as of today, there is not a single distribution of passage times for which it has been rigorously proved. Therefore, proving equation (4.29) is still an open question. (As we will see in the paragraphs below, one will probably need to solve Question 2.8.1 first.)

QUESTION 4.3.1. Find a distribution of passage times for which (4.29) holds with $d \geq 2$.

As shown in Theorem 4.15, a version of the inequality $\chi \geq 2\xi - 1$ was proved in the 1995 work of Newman-Piza [**149**]. A proof of the other inequality appeared first in a model of a Brownian particle in a Poissonian potential [**179**]; see also [**119**]. In FPP, Chatterjee [**51**] proposed a stronger definition of the exponents that allows a proof of (4.29). Unlike the definitions of ξ and χ given in the previous sections, Chatterjee's exponents are not known to exist for any distribution of passage times, but they do allow significant progress toward (4.29). The proof in [**51**] relies on a construction similar to that in [**52**]. One first breaks a geodesic into smaller segments and then uses an approximation scheme to compare the passage time to a sum of nearly i.i.d. random variables. The proof is then a tradeoff between minimizing the error and maximizing the variance of the passage time. Assuming that the distribution is "nearly Gamma" (see Section 3.2), the optimization can be achieved by choosing different parameters in the approximation.

The scaling relation should be valid only under the unproven assumption of uniform positive curvature of the limit shape. This is one of the main reasons for the introduction of a strong definition of exponents. It is reasonable to believe that given a different curvature exponent κ as defined in (2.38), one would have the alternate scaling relation $\chi = \kappa\xi - (\kappa - 1)$; see [**18**, Section 3] for more details. The inequality $\leq$ comes from (4.32) below, and the case $\kappa = 2$ reduces to $\chi = 2\xi - 1$.

We will now sketch how one could derive (4.29) assuming positive curvature. Here we follow the proof of (4.29) given in [**18**]. It starts with a standard fact: if X' is an independent copy of a random variable X then

$$\operatorname{Var} X = \frac{1}{2}\mathbb{E}(X - X')^2.$$

Thus, if we want to estimate χ, it suffices to compare the difference of two independent copies of $T(0, ne_1)$ as

$$n^{2\chi} \sim \mathbb{E}(T(0, ne_1) - T'(0, ne_1))^2. \tag{4.30}$$

The exponent ξ tells us exactly how to build such an independent copy T'. As in the proof of Theorem 4.17, one just needs to consider passage times from starting points 0 and $n^\xi e_2$, as the geodesics from there to ne_1 and $n^\xi e_2 + ne_1$ will with high probability be contained in two disjoint cylinders. Therefore, one should expect that

$$n^{2\chi} \lesssim \mathbb{E}(T'(0, ne_1) - T(0, ne_1))^2 \sim \mathbb{E}(T(n^\xi e_2, n^\xi e_2 + ne_1) - T(0, ne_1))^2. \tag{4.31}$$

Now, recall the definition of the curvature exponent given in Definition 2.38. If we are allowed to bound the passage time by its asymptotic value (as $T(0, x) \sim \mu(x)$), we would have

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

$$
\begin{aligned}
T(n^\xi e_2, ne_1) - T(0, ne_1) &\le \mu(ne_1 - n^\xi e_2) - \mu(ne_1) + \text{ error} \\
&= n\Big(\mu(e_1 - n^{\xi-1}e_2) - \mu(e_1)\Big) + \text{ error} \\
&\lesssim n n^{\kappa(e_1)(\xi-1)} + \text{ error} \\
&= n^{2\xi-1} + \text{ error},
\end{aligned} \tag{4.32}
$$

if $\kappa(e_1) = 2$. In the first equality above, we used the fact that μ is a norm. The above sequence of approximations tells us $2\xi - 1$ comes from the curvature assumption.

Now we compare the left side of (4.32) with the right side of (4.31). The difference in passage times in both equations is almost the same except that the ending points are not equal. However, a crossing trick from [**18**] and an argument from [**51**] using assumed exponential concentration on scale n^χ from the strong definition of χ allows us to take the same ending points, and thus we can replace (4.32) in (4.31) to get, from (4.30), $n^{2\chi} \le n^{2\xi-1}+$ error, or $\chi \le 2\xi - 1$.

The above argument is made rigorous in [**18**]. The assumptions on the exponents ξ and χ are used to justify the approximations in (4.32).

4.4. Infinite geodesics

In this section we investigate infinite geodesics. We start with the following definitions.

Definition 4.20. We will say that an infinite self-avoiding nearest-neighbor path Γ is an infinite geodesic (for a given edge-weight realization) if every finite subpath of Γ is a finite geodesic.

Infinite geodesics come in two varieties:

(1) Indexed by $\mathbb{N}$; that is, Γ's edges in order are $e_1, e_2, \ldots$. These are called unigeodesics, singly infinite geodesics, geodesic rays or (when it does not cause confusion) simply geodesics.
(2) Indexed by $\mathbb{Z}$. These are called bigeodesics or geodesic lines.

Much work on geodesics in the model has focused on determining the number of geodesics of either type which exist, as well as their properties and relationships.

4.4.1. Existence of geodesic rays. It is straightforward to check that, given any $x \in \mathbb{Z}^d$ there exists a.s. at least one geodesic ray originating from x. The argument proceeds as follows. Suppose without loss of generality that x is the origin 0. Consider the sequence of geodesics $\Gamma_n = \mathrm{Geo}(0, ne_1)$. The first vertex of each Γ_n is 0. There are $2d$ possible choices for the second vertex of each Γ_n. Therefore, there must be some edge $e(1)$ incident to the origin such that infinitely many Γ_n have $e(1)$ as their first edge. Take that edge to be the first edge of the infinite geodesic. Now there are $2d - 1$ possible choices for the third vertex of the subsequence of Γ_n that goes through $e(1)$. By the same reasoning as before, there must be some edge $e(2)$ incident to $e(1)$ such that infinitely many Γ_n have $e(2)$ as their second edge. Repeating this argument on subsequences yields a singly infinite path which is seen to be a geodesic.

Definition 4.21. Two geodesic rays are distinct if they share at most finitely many edges and vertices.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

A question that follows is whether there exist multiple distinct geodesic rays with positive probability. A natural way to construct distinct geodesics would be to consider the subsequential limit obtained in the last construction by taking a different sequence of endpoints, such as $-ne_1$. Given the observations on the wandering exponent given in Section 4.2, it is reasonable to believe that one can construct many infinite geodesics this way. However, this intuition is difficult to make rigorous.

In the two-dimensional case, with exponentially distributed edge-weights, Häggström-Pemantle [**106**] showed that there are at least two distinct geodesic rays with positive probability. This result was extended to a wide range of first-passage distributions by Garet-Marchand [**91**] and Hoffman [**111**].

An important further advance was made in a paper of Hoffman [**112**]. He demonstrated the existence of more than two geodesic rays directly using properties of the limit shape. We write sides($\mathcal{B}$) for the number of sides of the limit shape $\mathcal{B}$ (that is, sides($\mathcal{B}$) is the number of extreme points of $\mathcal{B}$ and is finite if and only if $\mathcal{B}$ is a polygon).

THEOREM 4.22 (Hoffman [**112**], Theorem 1.2). *Assume $d = 2$ and F is continuous with $\mathbb{E}\tau_e^{2+\alpha} < \infty$ for some $\alpha > 0$. Define $G(x_1, \ldots, x_k)$ to be the event that there exist distinct geodesic rays beginning at vertices $x_1, \ldots, x_k$. Then for any $k \leq$ sides($\mathcal{B}$) and any $\epsilon > 0$, there exist $x_1, \ldots, x_k \in \mathbb{Z}^2$ such that*

$$\mathbb{P}(G(x_1, \ldots, x_k)) > 1 - \epsilon.$$

This result also holds for a wide range of non-i.i.d. passage time distributions and was used by Hoffman to establish existence of four distinct geodesic rays. We will discuss this theorem (and give some idea of its proof) in another guise below as Theorem 6.8. In light of Hoffman's results, the following question could be seen as a weak version of the uniform curvature conjecture, since uniform curvature implies sides($\mathcal{B}$) $= \infty$.

QUESTION 4.4.1. Show that for continuous distributions and $d \geq 2$, there are infinitely many distinct geodesic rays.

It is expected that sides($\mathcal{B}$) $= \infty$ for any distribution that has more than one point in its support. As we saw in Section 2.5, this fact is only known for distributions in $\mathcal{M}_p$ with $p \geq \vec{p}_c$.

QUESTION 4.4.2. Find a distribution that is not in $\mathcal{M}_p$ that has an infinite number of distinct geodesic rays.

The last observation combined with Hoffman's result allows us to conclude:

THEOREM 4.23 (Auffinger-Damron [**19**], Theorem 2.3). *Assume $d = 2$. For any measure $\nu \in \mathcal{M}_p$, $p \geq \vec{p}_c$, a.s., there exist infinitely many distinct geodesic rays.*

It is important to note that when $p > \vec{p}_c$, for $\nu \in \mathcal{M}_p$, one can trivially find infinitely many distinct geodesic rays by choosing them to be oriented up-right and to contain only weight-one edges. In the above result, the geodesics constructed can be chosen to contain a positive fraction of edges e with $\tau_e > 1$. Each of these rays will be directed in a sector that is disjoint from the flat edge of the limit shape.

The argument that we provided at the beginning of this section for the existence of a geodesic ray leads to another important question. For its statement, we need a definition.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

DEFINITION 4.24. If Γ_n is a sequence of finite geodesics, we say that $\Gamma_n \to \Gamma$ for some path Γ if for any $N > 0$, the first N steps of Γ_n equal those of Γ if n is sufficiently large.

QUESTION 4.4.3. Is it true that the sequence of geodesics $(\mathrm{GEO}(0, ne_1))$ converges to some Γ?

By the definition of convergence, such Γ as in Question 4.4.3 is a geodesic ray. The question was solved in the two-dimensional upper half-plane or any other infinite connected subgraph of $\mathbb{Z}^2$ with infinite connected complement in [**20**] (in the latter case, ne_1 is replaced by a sequence of points on the boundary of the domain). It was also solved in the full plane $\mathbb{Z}^2$ if one assumes that $\partial\mathcal{B}$ is differentiable both in direction e_1, and at θ_1, θ_2, where the θ_i's are the endpoints of the sector of angles of contact of the unique tangent line to $\mathcal{B}$ in direction e_1. See Theorem 5.21 and Remark 5.26. Further progress on these questions under weaker differentiability conditions has been made by Ahlberg-Hoffman in [**5**].

4.4.2. Directions and coalescence. In this section, we investigate whether there exist geodesic rays that have asymptotic directions, and if the geodesic with a given direction is unique. Much of the work described here was done in the 1990s by Newman and co-authors under curvature assumptions on $\mathcal{B}$; for a complete pedagogical account of these results we refer to the book of Newman [**148**]. (Also see Chapter 5 on Busemann functions, where geodesics are analyzed without a curvature assumption.) For $\theta \in \mathbb{R}^d$ with $|\theta| = 1$ (identified with $[0, 2\pi)$ when $d = 2$), we will say that a self-avoiding singly infinite path $(e_1, e_2, \ldots)$ has direction θ if $\arg v_n := v_n/|v_n| \to \theta$ as $n \to \infty$, where $e_n = \{v_{n-1}, v_n\}$. Under the assumption of uniformly positive curvature, the following theorem shows that infinite geodesics cannot wind around the origin or wander through sectors infinitely often; they must have directions.

THEOREM 4.25 (Newman [**147**], Theorem 2.1). *For $d \geq 2$, assume that $F(0) < p_c$, that there exists $\alpha > 0$ such that $\mathbb{E}e^{\alpha\tau_e} < \infty$ and that $\mathcal{B}$ is uniformly curved (see Definition 2.39 or (2.29)). Then with probability one,*

(1) *Every geodesic ray has a direction.*
(2) *For every θ, there exists a geodesic ray with direction θ.*

Newman's theorem is stated for continuous distributions, in which both items above are deduced not for geodesic rays, but for infinite branches in $\mathcal{T}(0)$, but the same proofs yield the above extension to τ_e's that do not necessarily have continuous distribution.

PROOF. We will need to control geodesics from all points at once, and we will control them by ensuring that they exit cones along certain parts of their boundaries. For this, we need definitions of cone sectors: for $x \in \mathbb{Z}^d$ and $\epsilon > 0$, define the set

$$C_x = \{z \in \mathbb{Z}^d : |z| \in [|x|/2, 2|x|],\ |\arg x - \arg z| \leq |x|^{-1/4+\epsilon}\},$$

its "forward boundary"

$$\partial_f C_x = \{z \in \mathbb{Z}^d : \exists\ w \in C_x \text{ such that } |z - w| = 1 \text{ and } |z| > 2|x|\},$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

and the rest of the boundary is $\partial' C_x = \{z \in \mathbb{Z}^d \setminus C_x : \exists\, w \in C_x \text{ such that } |z-w| = 1\} \setminus \partial_f C_x$. Define the "out tree" of x as

$$\text{out}(x) = \{z \in \mathbb{Z}^d : T(0,z) = T(0,x) + T(x,z)\}$$

(the set of vertices z such that $x \in \overline{\text{GEO}}(0,z)$) and the event

$$G_x = \{\text{out}(x) \cap \partial' C_x \neq \emptyset\}.$$

A union bound gives

$$\mathbb{P}(G_x) \leq \sum_{z \in \partial' C_x} \mathbb{P}(T(0,z) = T(0,x) + T(x,z)).$$

Nearly the same argument as in the Newman-Piza bound (summing a similar bound to that given in (4.20)) shows that there is a $c > 0$ such that for all $x \in \mathbb{Z}^d$,

$$\mathbb{P}(G_x) \leq (1/c)e^{-c|x|^c}.$$

(Here it is important that the aperture of the cone C_x be at least $|x|^{-1/4+\epsilon}$, so that the distance between x and the side boundaries of C_x be of larger order than $|x|^{3/4}$. The exponent $3/4$ is the bound for the wandering exponent from Theorem 4.16.)

By the Borel-Cantelli lemma, we can a.s. find a random M such that for $|x| \geq M$, the event G_x^c occurs. Now fix any such configuration; we claim that the following geodesic-direction estimate holds. Consider a geodesic Γ starting from 0 with vertices $0 = x_0, x_1, x_2, \ldots, x_r$. So long as $n \in [0,r]$ is such that $|x_n| \geq M$, then

$$|\arg x_r - \arg x_n| \leq C|x_n|^{-1/4+\epsilon}, \tag{4.33}$$

where $C = \sum_{i=1}^{\infty} 2^{-i(1/4-\epsilon)}$. To prove this inequality, define a sequence of points $y_1, \ldots$ inductively as follows. Let $y_1 = x_n$ and for $i \geq 2$, define y_i to be the first point on Γ after y_{i-1} and after the first exit from $C_{y_{i-1}}$ (as long as there is one). Then there is $\mathcal{I}$ such that $x_r \in C_{y_\mathcal{I}}$, and we can bound

$$\begin{aligned}
|\arg x_r - \arg x_n| &\leq \sum_{i=1}^{\mathcal{I}-1} |\arg y_i - \arg y_{i+1}| + |\arg y_\mathcal{I} - \arg x_r| \\
&\leq \sum_{i=1}^{\mathcal{I}} |y_i|^{-1/4+\epsilon} \\
&\leq \sum_{i=1}^{\infty} (2^i |y_1|)^{-1/4+\epsilon} \\
&= C|x_n|^{-1/4+\epsilon}.
\end{aligned}$$

In the second to last inequality, we have used that $G_{y_i}^c$ occurs, and thus the geodesic exits C_{y_i} along the forward boundary. This proves (4.33).

Now we can argue the two points of Newman's theorem. For any passage time configuration as above, let Γ be an infinite geodesic. We may assume it starts from 0 and label its vertices $x_0, x_1, \ldots$. Pick n such that $|x_n| \geq M$ and note now that by (4.33), one has $|\arg x_n - \arg x_m| \leq |x_n|^{-1/4+\epsilon}$ for all $m \geq n$. This means $(\arg x_n)$ is a Cauchy sequence and thus converges, proving that Γ has a direction.

For the second point, for our given passage time configuration, pick any direction θ and let (y_n) be any sequence of vertices with $\arg y_n \to \theta$. Let Γ_n be a self-avoiding geodesic from 0 to y_n and let (Γ_{n_k}) be a subsequence that converges

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

to some Γ. We claim that Γ, which is an infinite geodesic, has direction θ. Indeed, if its vertices are $x_0, x_1, x_2, \ldots$, then for any n, Γ_{n_k} has as its initial segment the portion of Γ from 0 to x_n, so long as k is sufficiently large. Thus if $|x_n| \geq M$, by (4.33),

$$|\arg y_{n_k} - \arg x_n| \leq |x_n|^{-1/4+\epsilon}.$$

Taking $k \to \infty$, we obtain $|\theta - \arg x_n| \leq |x_n|^{-1/4+\epsilon}$, and this means $\arg x_n \to \theta$. $\square$

Under the unproven assumption of uniformly positive curvature, one could now ask how many geodesic rays with direction θ exist. Is it possible that there is more than one such ray, and does the answer change if we insist that they share the same initial vertex?

In this direction, the following two results are due to Licea-Newman [**139**]. The first one roughly says that geodesic rays cannot bifurcate in the same deterministic direction. We note, however, that birfucations are expected in random directions, and indeed such a statement has been proved in a related last-passage model [**58**]. See Remark 4.27 below.

THEOREM 4.26 (Licea-Newman [**139**], Theorem 0; Newman [**147**], Theorem 2.2). *Let $d = 2$ and assume that F is continuous. Fix $x \in \mathbb{Z}^2$. There exists some set $D \subseteq [0, 2\pi)$ of full Lebesgue measure such that if $\theta \in D$, then there is zero probability that there exist distinct geodesic rays starting at x with direction θ.*

The proof of Theorem 4.26 is surprisingly simple, and we will provide a sketch of the main idea below. Let us first mention that not much is known about the set D, although it is expected that $D = [0, 2\pi)$. Answering the question below would be a step towards one of the main problems (absence of geodesic lines) discussed in Section 4.5.

QUESTION 4.4.4. Show that $D = [0, 2\pi)$.

In [**148**, Theorem 1.5], it was shown by Zerner that D has at most a countably infinite complement. But we do not even know that a particular angle — for instance, 0 — belongs to D.

QUESTION 4.4.5. Show that $0 \in D$.

These two questions have been recently solved by Damron-Hanson under additional assumptions on the limit shape boundary [**69**]. Namely, the first is solved if one assumes that the limit shape boundary is differentiable everywhere. For the second, the condition is differentiability of the boundary in the direction $\theta = 0$ and in directions θ_1 and θ_2, where the θ_i's are the endpoints of the sector of angles at which the unique tangent line to the limit shape in direction $\theta = 0$ contacts the limit shape. See Theorem 5.21.

PROOF OF THEOREM 4.26. Suppose there are two different geodesic rays r_1, r_2 starting at some x with the same direction θ. As they are different, they will have to bifurcate at a vertex $u \in \mathbb{Z}^2$, one taking the directed edge (u, v_1), the other the edge (u, v_2), with $v_2 \neq v_1$. As the space is uniquely geodesic, r_1 and r_2 do not intersect after u. The construction that follows buids a geodesic between r_1 and r_2 by exploring the geodesic tree $\mathcal{T}(x)$ between these two rays. By planarity, any other geodesic ray that sits between these two geodesics must also have direction θ.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

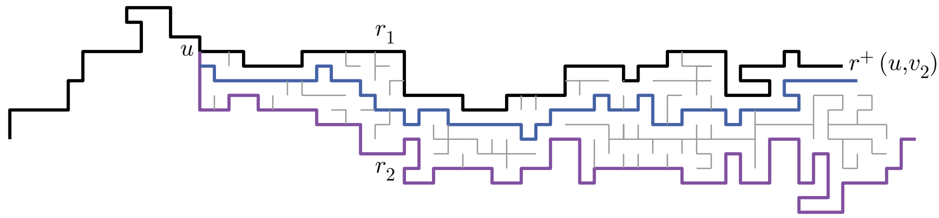


FIGURE 10. Construction of the path $r^+(u, v_2)$. The path r_1 is the uppermost, r_2 is the bottommost and $r^+(u, v_2)$ is in between.

Assume without loss of generality that r_1 is asymptotically counter-clockwise to r_2. Now construct the geodesic ray $r^+(u, v_2)$ that starts with the edge (u, v_2) and always chooses the most counter-clockwise edge possible in $\mathcal{T}(x)$ (in particular this edge must belong to at least one geodesic ray). Roughly speaking, the ray $r^+(u, v_2)$ is the "closest" geodesic ray to r_1 that takes the edge (u, v_2) in the region delimited by r_1 and r_2. (See Figure 10.) As r_2 is a geodesic ray, $r^+(u, v_2)$ sits between r_2 and r_1 and must have direction θ. Through this construction, one see that to any direction θ with multiple geodesic rays from x corresponds at least one edge e such that $r^+(e)$ has direction θ. Let $G(e, \theta)$ be the event that $r^+(e)$ has direction θ. We thus get the estimate

$$\mathbb{P}\Big(\text{There exist at least two geodesic rays from } x \text{ with direction } \theta\Big) \leq \sum_e \mathbb{P}(G(e, \theta)).$$

Now, for each edge e, $r^+(e)$ cannot be a geodesic ray with direction θ for more than one θ; thus Fubini's theorem implies

$$\int \mathbb{P}(G(e, \theta))d\theta = \int\int \mathbf{1}_{G(e,\theta)}(\omega)d\theta d\mathbb{P} = 0.$$

(One must check that $\mathbf{1}_{G(e,\theta)}(\omega)$ is jointly measurable in (θ, ω), but we omit this detail.) Therefore $\mathbb{P}(G(e, \theta)) = 0$ for Lebesgue almost all θ. Combining the last two displays ends the proof of the theorem. □

REMARK 4.27. It is expected (and, as mentioned above, proved under the assumption of differentiability of $\partial\mathcal{B}$) that for any fixed θ, one cannot find disjoint geodesic rays with direction θ. However, there should exist (and in fact, it is provable if one assumes existence of geodesics in any deterministic direction) a random, dense, countably infinite set of θ's for which there are disjoint geodesic rays with direction θ. Such a statement is proved in the related exactly solvable last-passage percolation model by Coupier [**58**], along with the statement that a.s., there exist no θ such that three geodesic rays from 0 have direction θ. Although this last result does not yet have a counterpart in non-exactly solvable models, progress toward the first under limit shape assumptions was made in [**97**].

The second theorem in this direction establishes that for all $\theta \in D$ any two geodesic rays with direction θ (starting from possibly different points) must coalesce.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

THEOREM 4.28 (Licea-Newman [**139**], Theorem 1; Newman [**147**], Theorem 2.3). *For $d = 2$ and $\theta \in D$ fixed with F continuous, there is zero probability that there are disjoint geodesic rays in the direction θ.*

The proof of the above theorem is two-dimensional. It is unclear if the result holds for $d > 2$.

QUESTION 4.4.6. Decide whether the claim of Theorem 4.28 holds in arbitrary dimension.

Ending this section, we quote an apparently still open question raised in the end of the original paper of Hammersley-Welsh.

QUESTION 4.4.7. Let $\Gamma(v)$ be a geodesic from 0 to v. Consider the collection of all edges that belong to $\Gamma(v)$ for infinitely many v and let $f(r)$ be the number of such edges which intersect the circle $x^2 + y^2 = r^2$. Does $f(r) \to \infty$ as $r \to \infty$; and, if so, how fast?

SKETCH OF THE PROOF OF THEOREM 4.28. We sketch the proof in the case where the weights are unbounded. The bounded-weight case requires a little modification of the arguments, and the details can be found in the original paper [**139**].

Consider all geodesic rays that have asymptotic direction $\theta \in D$. Construct the geodesic graph G composed of the union of all these geodesic rays. By Theorem 4.26, if two such rays meet, they must coalesce. Thus, the geodesic graph must be a forest with $N \geq 1$ distinct trees. The proof in [**139**] has 3 steps.

STEP 1. *If* $\mathbb{P}(N \geq 2) > 0$ *then* $\mathbb{P}(N \geq 3) > 0$.

To see this, without loss of generality, assume that the direction θ has positive e_1-coordinate. Also suppose that the two different geodesics touch the e_2-axis at two points $y_1 \neq y_2$ and are contained in the half-plane $\{z : z \cdot e_1 \geq 0\}$. By vertical translation invariance, it is possible to find with positive probability two different points y_3, y_4 that also have two such disjoint geodesic rays. By planarity, at least three of these rays must be distinct (see Figure 11). In fact, if $\mathbb{P}(N \geq 2) > 0$ then there exist $y_1 < y_2 < y_3$ in $\mathbb{Z}$ such that $\mathbb{P}(E(y_1, y_2, y_3)) > 0$, where $E(y_1, y_2, y_3)$ is the event that there are three disjoint geodesic rays with direction θ starting from $(0, y_i)$ and contained in the half-plane $\{(x, y) \in \mathbb{Z}^2 : x > 0\}$ except their initial points.

Let F be the subevent of $E(y_1, y_2, y_3)$ on which the tree in G containing the "middle" point y_2 does not intersect the half-plane $\{z : z \cdot e_1 < 0\}$. The second step of the proof is the following.

STEP 2. If $\mathbb{P}(N \geq 3) > 0$ then $\mathbb{P}(F) > 0$.

By the arguments in STEP 1, with positive probability there are three disjoint geodesic rays r_1, r_2, r_3 starting respectively from the points $(0, y_1), (0, y_2), (0, y_3)$ on the e_2-axis. We now set $y^* = \max\{|y_1|, |y_3|\}$ and increase the values of the edge-weights on edges between $u = (-1, y)$ and $v = (0, y)$ for $-y^* \leq y \leq y^*$. If we increase this finite collection of edge-weights enough, we create a barrier through which no geodesic can pass. (Specifically, if $\mathcal{F}$ is this set of edges, one needs that for all $e \in \mathcal{F}$, τ_e is greater than $T_e(x, y)$, where x, y are the endpoints of e, and T_e is the passage time among all paths which do not use e.) Note that, by our construction, this change also does not alter the $r_i's$. Furthermore, as we have unbounded edge-weights, one can construct such a barrier with positive probability. Now, since

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

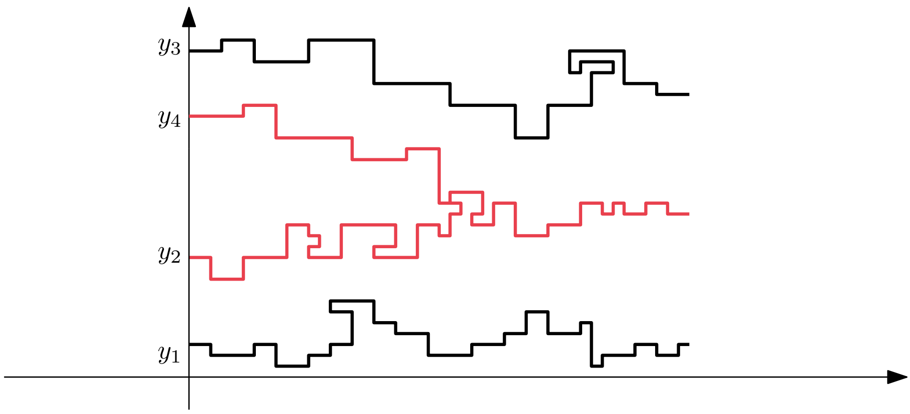


FIGURE 11. Construction of the three disjoint geodesics starting from y_1, y_2, y_3. The geodesic starting from y_4 may intersect the geodesic emanating from y_2 but all others are disjoint.

there is zero probability that two geodesic rays with direction θ meet and do not coalesce, this barrier attached to L, the segment of the e_2-axis between $-y^* e_2$ and $y^* e_2$, together with r_1 and r_3 "shields" r_2 from other geodesics in G and implies that the tree that contains r_2 is the desired tree.

STEP 3. $\mathbb{P}(F) > 0$ is impossible.

Consider all possible translates L^u of L by points of the form $3uy^*$ for $u \in \mathbb{Z}^2$ and the corresponding translated events F^u of F. If F^u and F^v both occur for distinct u, v, then, by definition, the corresponding "middle" geodesics must be disjoint, as they are blocked from each other by the "outer" geodesics constructed in the events F^u and F^v. Now, let C_M be the number of L^u's contained in the box $[0, M] \times [0, M]$ and N_M be the random number of the corresponding F^u's which occur. By translation invariance, and the assumption that $\mathbb{P}(F) > 0$,

$$\mathbb{E}N_M = C_M \mathbb{P}(F) \geq cM^2, \tag{4.34}$$

for some $c > 0$. But N_M is bounded above by the number of disjoint trees that touch $[0, M] \times [0, M]$. Since each tree is infinite, this number cannot exceed the number of boundary points in that box, which is at most $c'M$ for some c'. As $c'M < cM^2$ for M large, this contradicts (4.34), showing that $\mathbb{P}(F) = 0$. $\square$

4.5. Geodesic lines or bigeodesics

A final (and one of the most) important question in the study of infinite geodesics is the following:

QUESTION 4.5.1. Do geodesic lines exist?

4.5.1. Heuristic argument. We start with a heuristic argument that there should a.s. exist no geodesic lines (bigeodesics), at least in dimensions d where $\xi > 1/2$. In particular, the argument gives a plausible reason to disbelieve in bigeodesics in two dimensions, where ξ is expected to be $2/3$. It is important to

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

note that this argument is far from rigorous even under usual unproven assumptions (e.g. curvature) on the model.

Indeed, a main feature of the argument is that it uses various meanings of ξ interchangably. For instance, it is assumed that the geodesics $\text{Geo}(0, ne_1)$ and $\text{Geo}(m^\xi, ne_1)$ ($m \ll n$) merge after a distance of order m, in addition to $\text{Geo}(0, ne_1)$ having transversal wandering typically of order n^ξ.

This argument was presented by C. M. Newman at the workshop "First-passage percolation and related models" at the American Institute of Mathematics in Summer, 2015.

Setup. Consider the box $[-N, N]^d$ for N large; we denote the boundary of this box by the symbol ∂_N. Assume that there were positive probability that a bigeodesic existed (and hence that this event had probability one). Then this bigeodesic would have to pass through the origin, at least with positive probability.

Such a bigeodesic must pass through a vertex of ∂_N before passing through 0 and another vertex of ∂_N afterwards. In particular, the event that a geodesic between a pair of vertices of ∂_N passes through 0 must have uniformly (in N) positive probability. To rule out bigeodesics, it therefore suffices to show

$$\lim_{N\to\infty} \mathbb{P}\big(\exists\, x, y \in \partial_N \text{ such that } 0 \in \text{Geo}(x, y)\big) = 0. \tag{4.35}$$

The key idea to prove (4.35) is to use the wandering property of individual geodesics to show that they are likely to avoid 0.

Unfortunately, using a union bound on the events $\{0 \in \text{Geo}(x, y)\}_{x,y\in\partial_N}$ is unlikely to give (4.35), since there are many such pairs (x, y). To overcome this difficulty, we will group geodesics together, tiling ∂_N with $(d-1)$-dimensional cubes B_j of side length $\sim N^\xi$. For a particular cube B_j, let B_{-j} denote the corresponding cube on the opposite side of ∂_N.

Reducing to opposite blocks. We will need to make assumptions on the structure of geodesics between vertices in blocks B_i, B_j. The first is that we can neglect the contribution of the probability

$$\mathbb{P}\left(\exists\, x \in B_i, y \in B_j \text{ for some } i \neq -j \text{ such that } 0 \in \text{Geo}(x, y)\right)$$

to (4.35).

This is a not unreasonable assumption, since the straight line segment between such x and y will typically lie distance $\gg |x-y|^\xi$ from 0 (perhaps requiring the diameter of B_j to be at least $N^{\xi+\epsilon}$). This of course is likely to necessitate a strong concentration bound on the variable

$$D(0, ne_1)/n^\xi,$$

where D was defined above (4.13). In any case, once such a result is established, it allows us to reduce the problem of showing (4.35) to showing that as $N \to \infty$,

$$\mathbb{P}\big(\exists j,\ x \in B_j,\ y \in B_{-j} \text{ such that } 0 \in \text{Geo}(x, y)\big) \tag{4.36}$$

tends to zero.

Reducing within blocks. We have (at least assuming $\xi < 1$) succeeded in reducing the number of pairs of geodesics under consideration by a power by reducing to opposite pairs of blocks. Since geodesics emanating from vertices within the same B_j are likely to be highly correlated, we can reduce further by grouping together

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

such geodesics. In particular, we will assume that all geodesics originating in one block B_j merge before coming close to the origin.

For each block B_j, choose a particular vertex x_j near the center of B_j. As discussed earlier, it is likely that there is "geodesic merging on scale ξ." In particular, for any $y \in B_j$, $y' \in B_{-j}$, it is plausible that the event $\{0 \in \mathrm{Geo}(y, y')\}$ has extremely small symmetric difference with the event $\{0 \in \mathrm{Geo}(x_j, x_{-j})\}$. Thus, for fixed j we can approximate

$$\mathbb{P}\big(\exists\, x \in B_j,\, y \in B_{-j} \text{ such that } 0 \in \mathrm{Geo}(x, y)\big) \approx \mathbb{P}\big(0 \in \mathrm{Geo}(x_j, x_{-j})\big). \tag{4.37}$$

With this approximation established, we have reduced the problem sufficiently to take a union bound:

$$\begin{aligned} (4.36) &\le \mathbb{P}\big(\exists\, j \text{ such that } 0 \in \mathrm{Geo}(x_j, x_{-j})\big) \\ &\lesssim \#(\text{boxes } B_j)\ \mathbb{P}\big(0 \in \mathrm{Geo}(x_1, x_{-1})\big). \end{aligned} \tag{4.38}$$

Individual geodesic wandering. We now need to control the probability $\mathbb{P}(0 \in \mathrm{Geo}(x_1, x_{-1}))$, which will require a further assumption on geodesic wandering. In some sense, what is required here is a lower bound on fluctuations, that $\mathrm{Geo}(x_1, x_{-1})$ is "spread out" on scale N^ξ far from its endpoints.

To be more precise, if L is the line between x_1 and x_{-1} and H denotes the hyperplane passing through 0 to which L is normal, we assume that $\mathrm{Geo}(x_1, x_{-1})$ passes through order one number of vertices of H and every vertex in a block of diameter N^ξ is equally likely to be passed through. In particular,

$$\mathbb{P}(0 \in \mathrm{Geo}(x_1, x_{-1})) \sim N^{-\xi(d-1)}.$$

With this information, we can conclude by bounding the right-hand side of (4.38). Note that the number of boxes B_i is of the order of

$$\frac{N^{d-1}}{N^{\xi(d-1)}} = N^{(1-\xi)(d-1)},$$

giving

$$(4.38) \lesssim \frac{N^{(1-\xi)(d-1)}}{N^{\xi(d-1)}} = \big(N^{1-2\xi}\big)^{d-1}.$$

In particular, if $\xi > 1/2$, then (4.35) holds, giving a.s. absence of bigeodesics.

4.5.2. Rigorous results. In this section, we move to rigorous results, outlining the work of Licea-Newman and Wehr from the 1990s. In Section 5.6, we give the most recent results, ruling out bigeodesics in any deterministic direction under assumptions on the limit shape boundary.

The first partial answer to the bigeodesic question comes from Theorem 4.29 below. We say that a geodesic line has asymptotic directions (θ_1, θ_2) if its two rays have directions θ_1 and θ_2.

THEOREM 4.29 (Licea-Newman [**139**], Theorem 2). *For $d = 2$, $\theta_1,\ \theta_2 \in D$, and F continuous, there is zero probability that there exists a geodesic line that has directions (θ_1, θ_2).*

The proof of Theorem 4.29 is not difficult once we have Theorem 4.28.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Proof of Theorem 4.29. If there are two distinct geodesic lines with directions (θ_1, θ_2), then by Theorem 4.28 they must a.s. coalesce in both directions. This cannot happen, as we assumed the model is uniquely geodesic. Thus, there must be at most one geodesic line with directions (θ_1, θ_2). Let E be the event that there exists a unique geodesic line with directions (θ_1, θ_2). By ergodicity, the probability of the event E is either 0 or 1.

Suppose for a contradiction that $\mathbb{P}(E) = 1$. Then it is possible to find L large enough such that the geodesic line intersects the box $B_L := [0, L] \times [0, L]$ with probability at least 2/3. Choosing a sequence of points $z_n \in \mathbb{Z}^2$ with $z_n/|z_n| \to \alpha \neq \theta_1$ or θ_2, the geodesic line passes through only finitely many shifted boxes $z_n + B_L$. Letting E_S be the event that the geodesic line intersects the set $S \subseteq \mathbb{Z}^2$, this implies

$$\mathbb{P}(E_{B_L} \cap E_{z_n+B_L}) \to 0 \text{ as } n \to \infty.$$

On the other hand, by translation invariance,

$$\mathbb{P}(E_{B_L} \cap E_{z_n+B_L}) \geq 1 - \mathbb{P}(E^c_{B_L}) - \mathbb{P}(E^c_{z_n+B_L}) \geq 1/3,$$

a contradiction. □

We note that it is possible to replace this proof with one that proceeds along the lines of Wehr's proof that the number of bigeodesics in all of $\mathbb{Z}^d$ must be zero or infinity:

Theorem 4.30 (Wehr [**173**]). *Assume F is continuous, $d \geq 2$, and $\mathbb{E}\tau_e < \infty$. The number $\mathcal{N}$ of bigeodesics in a configuration (τ_e) is a.s. constant, and $\mathcal{N} \in \{0, \infty\}$.*

The idea of the proof is that if there are only finitely many bigeodesics, then by translation invariance, in a large box, at least one of the bigeodesics must take a positive density of edges. So the passage time along this bigeodesic from its first entry to its last exit of the box must be of the order of the volume of the box. This contradicts the fact that the Euclidean distance between these two points is of the order of the length of the box, and so the passage time between them is also of this same order, with high probability. This theorem has been extended to some non-i.i.d. times by Boivin-Derrien [**37**]. In that paper, they also construct non-i.i.d. but translation invariant distributions on (τ_e) for which there exist bigeodesics.

A positive answer to Question 4.4.4, combined with the above result, would exclude the possibility of geodesic lines with deterministic asymptotic directions. It would not solve Question 4.5.1, as one could still have geodesic lines in random directions.

Question 4.5.1 was answered negatively in the upper half-plane $\mathcal{H}$ of $\mathbb{Z}^2$ with edge-weights that have finite mean and continuous distribution by Wehr and Woo [**175**].

Theorem 4.31 (Wehr-Woo [**175**], Theorem 1). *Assume F is continuous and $\mathbb{E}\tau_e < \infty$. A.s., there exist no geodesic lines in $\mathcal{H}$.*

We outline the proof of Theorem 4.31 at the end of this section. The assumption of finite mean was relaxed in [**20**], where one just requires uniqueness of finite geodesics. That work also included a range of non-i.i.d. distributions.

Before sketching Wehr-Woo's argument, let us stress that the importance of determining non-existence of geodesic lines goes beyond the field of FPP. Apparently, this problem was first posed by H. Furstenberg in a different context. As far

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

as we know, the first mathematical reference is in [**123**, p. 258]. Question 4.5.1 is equivalent to the problem of existence of nontrivial (that is, non-constant) ground states in the two-dimensional Ising ferromagnet with random exchange constants. Ground states of higher-dimensional random ferromagnets are similarly related to hypersurfaces with minimal random weights [**124**, **173**].

More precisely, consider the lattice dual to $\mathbb{Z}^2$, defined by

$$(\mathbb{Z}^2_*, \mathcal{E}^2_*) = (\mathbb{Z}^2, \mathcal{E}^2) + \frac{1}{2}(e_1 + e_2),$$

where $\mathcal{E}^2$ is the set of nearest-neighbor edges of $\mathbb{Z}^2$, and define a spin configuration as an element σ of $\{-1, +1\}^{\mathbb{Z}^2_*}$. Let $(J_{x,y})_{(x,y)\in\mathcal{E}^2_*}$ be a collection of i.i.d. positive random variables. For any configuration and any finite set $S \subset \mathbb{Z}^2_*$ define the (random) energy

$$H_S(\sigma) = - \sum_{\substack{\{x,y\}\in\mathcal{E}^2_* \\ x\in S}} J_{x,y}\sigma_x\sigma_y.$$

We will call σ a ground-state for $(J_{x,y})$ if, for each configuration ρ such that $\rho_x = \sigma_x$ for all x outside some finite set, we have

$$H_S(\sigma) \le H_S(\rho) \text{ for all finite } S \subset \mathbb{Z}^2_*.$$

It is an open problem to describe the set of ground states for this ferromagnetic model. In particular, it is not known how many ground states there are for a given $(J_{x,y})$, although it is conjectured (see, for instance, the discussion in [**148**, Chapter 1]) that, if the law of J is continuous, there should be only two a.s..

These two are the constant configurations $\sigma \equiv \pm 1$, which are clearly a.s. ground states. If any nonconstant ground states σ exist, they cannot have finite regions of disagreement; that is, there can be no finite S such that $\sigma_x = +1$ for all $x \in S$ and $\sigma_y = -1$ for all $y \in \partial S$, or vice-versa. Therefore, any nonconstant ground state must have a two-sided and circuitless infinite (original lattice) path of edges dual to edges $\{x, y\}$ such that $\sigma_x = -\sigma_y$.

Now consider FPP with $\tau_e = J_{x,y}$ if $\{x, y\}$ is the dual of the edge e. If such a first-passage configuration had a geodesic line, then the configuration σ which takes the value $+1$ on one side of the geodesic line and -1 on the other would be a nonconstant ground state for the associated spin model. Therefore in $d = 2$, based on the analogy above with the Ising ferromagnet, it is believed that one should have a negative answer for Question 4.5.1.

Partial results in this direction were obtained by Newman under the assumption of uniformly positive curvature. In [**147**], it was shown that there cannot exist geodesic lines other that those that have directions on both ends. As we have seen also, Theorem 4.29 ruled out almost all possible pairs of deterministic directions and Theorem 4.30 showed that the number of geodesic lines is 0 or infinity. There are also further arguments in the physics literature [**89**] against the existence of nonconstant ground states.

Let us end this subsection by presenting the argument of Wehr-Woo for absence of geodesic lines in the upper half-plane $\mathcal{H}$. For simplicity, we assume that $\mathbb{E}\tau_e < \infty$ and F is continuous, so that finite geodesics are unique a.s.. Let

$$E = \{\text{there exists a geodesic line in } \mathcal{H}\}.$$

By horizontal translation ergodicity, $\mathbb{P}(E)$ is zero or one; so let us assume for a contradiction that $\mathbb{P}(E) = 1$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

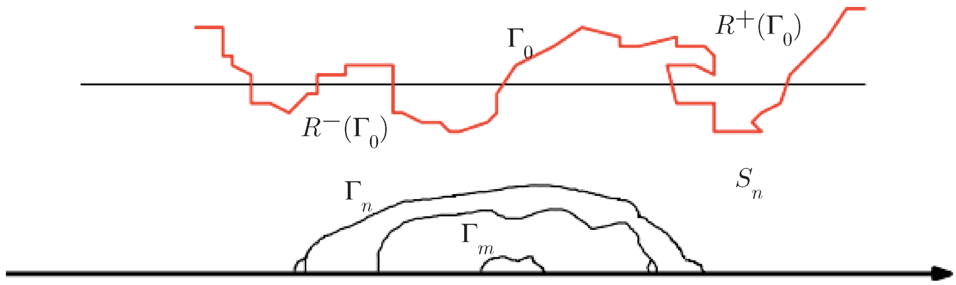


FIGURE 12. The geodesic line Γ_0 (above) and the regions $R^+(\Gamma_0)$ and $R^-(\Gamma_0)$. The intersection of Γ_0 and the strip S_n is a union of bounded sets. The geodesics Γ_n monotonically converge to Γ_0.

Any geodesic line Γ divides $\mathbb{R}^2 \setminus \Gamma$ into two components, say $R^+ = R^+(\Gamma)$ and $R^- = R^-(\Gamma)$; that is,

$$R^+(\Gamma) \cap R^-(\Gamma) = \emptyset,$$
$$R^+(\Gamma) \cup R^-(\Gamma) = \mathbb{R}^2 \setminus \Gamma,$$
$$\partial R^+ = \partial R^- = \Gamma,$$

where R^- is a region that contains $(0, -1)$ and where ∂A denotes the usual boundary of a set $A \subset \mathbb{R}^2$. Hence, by uniqueness of passage times, for any points $x, y \in R^-(\Gamma)$, no edge belonging to the finite geodesic $\text{GEO}(x, y)$ can intersect $R^+(\Gamma)$.

STEP 1. The first step of the proof is to establish the following fact [**175**, Proposition 4]. If a geodesic line exists with probability one, then there is a lowest.

PROPOSITION 4.32. *Assume that* $\mathbb{P}(E) = 1$. *Consider the sequence of geodesics* (Γ_n) *from* $(-n, 0)$ *to* $(n, 0)$ *for* $n \in \mathbb{N}$. *With probability 1, this sequence has a limit:*

$$\Gamma_0 = \lim_{n\to\infty} \Gamma_n.$$

Moreover, Γ_0 *is a geodesic line, and for any geodesic line* Γ,

$$\Gamma_0 \subset \left[R^-(\Gamma) \cup \Gamma\right].$$

PROOF. Suppose that Γ is a geodesic line. By uniqueness of passage times, for every n, Γ_n must lie in $R^-(\Gamma) \cup \Gamma$. Now, compactness (that is, the same type of argument used to establish existence of geodesic rays from the beginning of Section 4.4.1) implies that (passing through a subsequence) (Γ_{n_k}) must converge. For instance, there is only a finite number of possible edges for the intersection of Γ_n with the connected segment of the e_2-axis containing 0 and touching Γ_0 only once. Furthermore, this intersection is nonempty for every $n \geq 1$. However, the geodesics Γ_n are increasing in the sense that for $n > m$, Γ_n must lie above Γ_m. Thus, they converge. □

STEP 2. If Γ_0 intersects the strip $S_n = \{(x, y) \in \mathbb{Z}^2 : 0 \leq y < n\}$ then $R^+(\Gamma_0) \cap S_n$ is a non-empty union of bounded connected sets. See Figure 12.

This step is the main tool to prove:

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

STEP 3. Uniqueness of geodesics implies that Γ_0 must intersect any large box with probability bounded below uniformly in the position of the box.

Precisely, for $l \in \mathbb{N}$, write $B = B(l) = [-l, l] \times [0, 2l]$ and let K be the event that at least one geodesic line intersects B. Define, for $L \in \mathbb{N}$, translations of B by

$$B_{i,j} = B_{i,j}(l, L) = B + (iL, jL) \text{ for } (i,j) \in \mathbb{Z}^2 \text{ and } j \geq 0.$$

For $L > 2l$, the $B_{i,j}$'s are mutually disjoint.

PROPOSITION 4.33. *Let $\delta = 1 - \mathbb{P}(K)$. Then for all i, j with $j \geq 0$,*

$$\mathbb{P}\Big(B_{i,j} \subset R^+(\Gamma_0)\Big) \leq \delta, \qquad \mathbb{P}\Big(B_{i,j} \subset R^-(\Gamma_0)\Big) \leq \delta \quad \textit{and} \tag{4.39}$$

$$\mathbb{P}\Big(\Gamma_0 \cap B_{i,j} \neq \emptyset \textit{ for all } 0 \leq i, j < k\Big) \geq 1 - 2k^2\delta. \tag{4.40}$$

To prove (4.39), one uses STEP 1 and STEP 2. Consider the reflected half-planes

$$\mathcal{H}_j^r = \big\{(x, y) \in \mathbb{R}^2 : y \leq jL + 2l\big\}$$

and define the event

$$K_{i,j} := \big\{\text{at least one } \mathcal{H}_j^r\text{-geodesic line intersects } B_{i,j}\big\}.$$

Since the model is invariant under translations and under rotations by 180 degrees,

$$\mathbb{P}(K) = \mathbb{P}(K_{i,j}). \tag{4.41}$$

If $B_{i,j} \subset R^+(\Gamma_0)$ then, by STEP 2, $B_{i,j}$ is contained in exactly one of the bounded components of $R^+(\Gamma_0) \cap \mathcal{H}_j^r$. Hence, there is no $\mathcal{H}_j^r$-geodesic line intersecting $B_{i,j}$, since otherwise two different finite $\mathcal{H}_j^r \cap \mathcal{H}$-geodesics would meet more than once. Thus the event $\{B_{i,j} \subset R^+(\Gamma_0)\}$ is contained in the event $K_{i,j}^c$, which combined with (4.41) leads to the first part of (4.39).

The second equation in (4.39) uses STEP 1 and the fact that the box $B_{i,j}$ would be trapped for n large by the union of the geodesic Γ_n and the boundary of the upper half-plane. Therefore no geodesic line in the shifted half-plane $\mathcal{H}_j = \{(x, y) \in \mathbb{R}^2 : y \geq jL\}$ can intersect $B_{i,j}$. To get (4.40), one uses (4.39) to obtain

$$\mathbb{P}(\Gamma_0 \cap B_{i,j} \neq \emptyset) \geq 1 - 2\delta$$

for any i, j. Thus,

$$\mathbb{P}\Big(\bigcap_{0 \leq i,j < k} \{\Gamma_0 \cap B_{i,j} \neq \emptyset\}\Big) \geq 1 - 2k^2\delta.$$

STEP 4. Equation (4.40) contradicts $\mathbb{P}(E) = 1$. Thus, $\mathbb{P}(E) = 0$.
To see this, assume $\mathbb{P}(E) = 1$ and for any $k \geq 1$, choose $l = l(k)$ large enough so that $B = B(l)$ satisfies

$$\mathbb{P}(\Gamma_0 \cap B \neq \emptyset) = \mathbb{P}(K) > 1 - 1/4k^2.$$

By (4.40), this leads to

$$\mathbb{P}\Big(\bigcap_{0 \leq i,j < k} \{\Gamma_0 \cap B_{i,j} \neq \emptyset\}\Big) \geq 1/2. \tag{4.42}$$

The proof now consists of showing that (4.42) cannot hold for k large if L, the distance of the shift, is large enough.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

To see this, choose $L > 2l$ so all shifted boxes $B_{i,j}$ are disjoint. If the event in (4.42) occurs, it is possible to find k^2 different points in the geodesic line Γ_0, each one inside a different box $B_{i,j}$. Let $x_m, 1 \leq m \leq k^2$ be these points, ordered as they appear in Γ_0. As the ℓ^1-diameter of the union of the $B_{i,j}$ is at most $2Lk$, one has

$$L - 2l \leq \|x_1 - x_{k^2}\|_1 \leq 2Lk \ \text{ and } |\textsc{Geo}(x_1, x_{k^2})| \geq k^2 L/4.$$

There at most $4l^2k^2$ choices for both x_1 and x_{k^2} so a union bound leads to

$$\mathbb{P}\Bigg(\bigcap_{0 \leq i,j < k} \{\Gamma_0 \cap B_{i,j} \neq \emptyset\} \Bigg) \leq 4l^2k^4 \sup_{\|x-y\|_1 \leq 2Lk} \mathbb{P}\Bigg(|\textsc{Geo}(x,y)| \geq k^2 L/4 \Bigg).$$

If k (and therefore l) is fixed large enough then, by Theorem 4.9, the supremum on the right side is exponentially small in L. This contradicts (4.42) for L large, ending the proof of Theorem 4.31.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/05

CHAPTER 5

Busemann functions

The modern approach to infinite geodesics involves Busemann functions. They were first exploited in FPP in the important papers of Hoffman [**111**, **112**], and they allow one to obtain some information about geodesics without the unproven assumptions (for example, positive curvature of the limiting shape) of Newman.

5.1. Basics of Busemann functions

Long before FPP existed, H. Busemann invented various tools to study the geometry of geodesics in certain metric spaces [**42**]. Busemann functions grew out of the attempt to understand parallelism between geodesics. We can give some basic properties in the setting of FPP. Consider a geodesic ray R, and list its vertices in order as

$$V(R) = \{x_0, x_1, \ldots\}.$$

Here, x_0 is taken as the initial vertex of the ray.

Definition 5.1. The Busemann function associated to the ray R is $f_R : \mathbb{Z}^d \to \mathbb{R}$, given by

$$f_R(x) = \lim_n \left[T(x, x_n) - T(x_0, x_n)\right].$$

This limit exists since the terms are bounded ($|T(x, x_n) - T(x_0, x_n)| \leq T(x_0, x)$) and monotone:

$$\begin{aligned} T(x, x_{n+1}) - T(x_0, x_{n+1}) &= T(x, x_{n+1}) - T(x_n, x_{n+1}) - T(x_0, x_n) \\ &\leq T(x, x_n) - T(x_0, x_n). \end{aligned}$$

The Busemann function for R measures asymptotically how far behind x_0 is relative to x when both points attempt to travel down the ray R. The next lemma collects a few properties of this function.

Lemma 5.2. *The following hold.*

(*a*) *For $m < n$, one has $f_R(x_m) - f_R(x_n) = T(x_m, x_n)$.*
(*b*) *For all x, one has $|f_R(x)| \leq T(x, x_0)$.*
(*c*) *If R_1 and R_2 are geodesic rays that coalesce (they have finite symmetric difference) and have initial points x_0 and y_0, then*

$$f_{R_1}(x) = f_{R_2}(x) - f_{R_2}(x_0) \text{ for all } x \in \mathbb{Z}^d.$$

Proof. Proofs of items (a) and (b) follow directly from the definition. As for (c), write the vertices of R_1 as $x_0, x_1, \ldots$ and the vertices of R_2 as $y_0, y_1, \ldots$. By coalescence, we can find $N \in \mathbb{N}$ such that for n large, one has $x_n = y_{n+N}$. For such

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

n,

$$\begin{aligned} T(x, x_n) - T(x_0, x_n) &= T(x, y_{n+N}) - T(y_0, y_{n+N}) \\ &\qquad - T(x_0, y_{n+N}) + T(y_0, y_{n+N}) \\ &\to f_{R_2}(x) - f_{R_2}(x_0). \end{aligned}$$

□

We will often use a Busemann function that does not depend on the initial point. If R is a geodesic ray, then set

$$f_R(x, y) = f_R(x) - f_R(y).$$

We note the following properties of this Busemann function:

(1) f_R is anti-symmetric: $f_R(x, y) = -f_R(y, x)$.
(2) f_R is additive:

$$f_R(x, y) = f_R(x, z) + f_R(z, y) \text{ for } x, y, z \in \mathbb{Z}^d.$$

(3) If R_1 and R_2 are geodesic rays that coalesce, then $f_{R_1}(x, y) = f_{R_2}(x, y)$ for all $x, y \in \mathbb{Z}^d$.

PROOF. Using item (c) in Lemma 5.2 and writing x_0 for the initial point of R_1, one has

$$\begin{aligned} f_{R_1}(x) - f_{R_1}(y) &= (f_{R_2}(x) - f_{R_2}(x_0)) - (f_{R_2}(y) - f_{R_2}(x_0)) \\ &= f_{R_2}(x) - f_{R_2}(y). \end{aligned}$$

□

(4) (Translation covariance) Let θ be a translation of the lattice by an integer vector. Then

$$f_R(x, y)(\omega) = f_{\theta R}(\theta(x), \theta(y))(\theta(\omega)).$$

Here, the translated weight-configuration $\theta(\omega)$ is defined as $\tau_e(\theta(\omega)) = \tau_{\theta^{-1}e}(\omega)$ for any edge e. Furthermore θR is the translated geodesic ray.

5.2. Hoffman's argument for multiple geodesics

In this section we present Hoffman's [**111**] argument that there exist at least two infinite geodesics a.s.. This result was also proved by Garet-Marchand [**91**], but Hoffman's techniques, involving Busemann functions, led to many other results. We will use the assumptions of Hoffman, which do not require the τ_e's to be independent:

(1) $\mathbb{P}$ is ergodic under lattice translations.
(2) $\mathbb{P}$ has unique passage times: a.s., distinct paths have distinct passage times.
(3) $\mathbb{E}\tau_e^{2+\delta} < \infty$ for some $\delta > 0$.
(4) The limit shape $\mathcal{B}$ for $\mathbb{P}$ is bounded.

Recall from Definition 4.21 that two geodesic rays R_1 and R_2 are distinct if they share at most finitely many edges and vertices. Let $\mathcal{N} = \mathcal{N}(\omega)$ be the maximal number of distinct geodesic rays in the edge-weight configuration for outcome ω. It can be shown that $\mathcal{N}$ is a measurable function of the edge-weights and is invariant under lattice translations. So by ergodicity, it is a.s. constant.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

THEOREM 5.3 (Hoffman [**112**], Theorem 2; Garet-Marchand [**91**], Theorem 3.2). *The number $\mathcal{N}$ is at least two.*

PROOF. By taking a subsequential limit of the geodesics from 0 to ne_1 as in Section 4.4.1, we see that $\mathcal{N} > 0$. Assume for a contradiction that $\mathcal{N} = 1$. Let $R = R(\omega)$ be any geodesic ray and define $f : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}$ by

$$f(x,y) = f_R(x,y) = f_{R(\omega)}(x,y)(\omega).$$

On the probability 1 set on which $\mathcal{N} = 1$, this function is well-defined and is independent of the choice of R. Indeed, if $R' = R'(\omega)$ is another geodesic ray, then it cannot be disjoint from R. Due to uniqueness of passage times, R' must coalesce with R and therefore $f_R = f_{R'}$.

One can also check that f is measurable. Due to translation covariance of f_R, the function f is also translation covariant:

$$f(x,y)(\omega) = f\big(\theta(x), \theta(y)\big)(\theta(\omega)) \text{ a.s.}$$

for any lattice translation θ. This, combined with additivity and the ergodic theorem, implies that for each $x \in \mathbb{Z}^d$,

$$\frac{1}{n} f(0, nx) = \frac{1}{n} \sum_{k=1}^{n} f((k-1)x, kx) \to \mathbb{E} f(0,x) \quad \text{a.s. and in } L^1. \tag{5.1}$$

To evaluate this limit, we use symmetry: for $k = 1, \ldots d$,

$$0 = \mathbb{E} f(0, e_k) + \mathbb{E} f(e_k, 0) = \mathbb{E} f(0, e_k) + \mathbb{E} f(-e_k, 0)$$

and by translation invariance, this equals

$$\mathbb{E} f(0, e_k) + \mathbb{E} f(0, e_k),$$

so $\mathbb{E} f(0, e_k) = 0$. We further obtain

$$\mathbb{E} f(0, me_k) = m \mathbb{E} f(0, e_k) = 0$$

for all integers m and then by additivity,

$$\mathbb{E} f(0, x) = 0 \quad \text{for all } x \in \mathbb{Z}^d.$$

Combining with (5.1), one has

$$\frac{1}{n} f(0, nx) \to 0 \quad \text{a.s. and in } L^1.$$

To upgrade this convergence to an omni-directional statement, one can use the condition $\mathbb{E}\tau_e^{2+\delta} < \infty$ along with the bound $|f(x,y)| \leq T(x,y)$ to prove a type of shape theorem for the function f. That is, for each $\epsilon > 0$,

$$\mathbb{P}\left(|f(0,x)| < \epsilon \|x\|_1 \text{ for all } x \in \mathbb{Z}^d \text{ with } \|x\|_1 \text{ large enough}\right) = 1. \tag{5.2}$$

(The proof follows the same method as in Section 2.3.) In particular, if we define R to be the (a.s. unique) limit of geodesics from 0 to ne_1, and we write its vertices in order as $x_0, x_1, \ldots$, then one has

$$f(0, x_n)/\|x_n\|_1 \to 0 \quad \text{a.s..}$$

On the other hand, by property 1 of f_R, one has $f(0, x_n) = T(0, x_n)$. Therefore

$$T(0, x_n)/\|x_n\|_1 \to 0 \quad \text{a.s..}$$

This contradicts the fact that the limit shape is bounded and therefore shows $\mathcal{N} > 1$. □

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

5.3. Directions of geodesics via Busemann functions

In [**147**], Newman introduced the following limits, in an attempt to study the local geometry of the boundary of the growing ball $B(t)$:

$$\lim_n [T(x, x_n) - T(y, x_n)] \quad \text{for } x, y \in \mathbb{Z}^2,$$

where (x_n) is a deterministic sequence diverging to infinity in a direction θ. He was able to show that under the uniformly positive curvature assumption on $\mathcal{B}$ (Definition 2.39 or (2.29)), the limit exists a.s. for Lebesgue-almost every direction θ.

In 2008, Hoffman [**112**] defined Busemann-type functions similar to the above, and introduced the idea of studying geodesics through their Busemann functions. These ideas were continued by Damron-Hanson in 2013 [**68**]. In the next two sections, we will describe some of the ideas therein.

Let H_n be the hyperplane $\{x \in \mathbb{R}^d : x \cdot e_1 = n\}$ and define the function

$$B_n(x, y) = T(x, H_n) - T(y, H_n) \quad \text{for } x, y \in \mathbb{Z}^d. \tag{5.3}$$

For purposes of illustration, we will in this section assume that

$$B(x, y) := \lim_n B_n(x, y) \quad \text{exists a.s.} \tag{5.4}$$

for $x, y \in \mathbb{Z}^d$. This assumption should be compared to assumption (2.18). It is not known how to show that limits like (5.4) exist, and this is one of the first and main obstacles when dealing with Busemann functions. We will further discuss this issue in Section 5.4, where we can get around it by considering types of weak limits. The results presented in the rest of this section are valid under assumption (5.4).

QUESTION 5.3.1. Prove that under some conditions on the edge-weights the limit (5.4) exists.

REMARK 5.4. For the generalized Busemann functions introduced in Section 2.6, the question above corresponds to Question 2.6.1. One may also want to consider Busemann limits to points rather than lines; that is, limits of the form $\lim_n [T(x, ne_1) - T(y, ne_1)]$.

LEMMA 5.5. *Assume* (5.4) *and Hoffman's conditions 1–4. One has*

$$\mathbb{E}B(x, y) = (x - y) \cdot \rho \textit{ for } x, y \in \mathbb{Z}^d,$$

where $\rho = e_1 \mu(e_1)$.

PROOF. By translation invariance,

$$\mathbb{E}B(x, y) = \mathbb{E}B(\theta(x), \theta(y)) \tag{5.5}$$

for any lattice translation θ. As B is additive and satisfies

$$\mathbb{E}B(0, -x) = -\mathbb{E}B(0, x),$$

we find that $\mathbb{E}B(x, y)$ is a linear function of $x - y$. To find this function, it suffices to find $\mathbb{E}B(0, e_i)$ for $i = 1, \ldots, d$. By symmetry, one has

$$\mathbb{E}B(0, e_i) = 0$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

for $i > 1$. To find $\mathbb{E}B(0, e_1)$, we use an averaging trick introduced by Garet-Marchand and Hoffman (and which also appears in some form in Kingman [**128**, eq. (26)]). By translating by e_1, we obtain

$$\begin{aligned}\mathbb{E}T(0, H_n) &= \sum_{k=1}^{n} \mathbb{E}[T((k-1)e_1, H_n) - T(ke_1, H_n)] \\ &= \sum_{k=1}^{n} \mathbb{E}[T(0, H_{n-k+1}) - T(e_1, H_{n-k+1})] \\ &= \sum_{k=1}^{n} \mathbb{E}B_k(0, e_1).\end{aligned}$$

Using the bound $|B_k(x, y)| \le T(x, y)$ and the dominated convergence theorem, one has

$$\mathbb{E}B(0, e_1) = \lim_n \frac{1}{n}\mathbb{E}T(0, H_n) = \mu(e_1).$$

The last equality can be shown using the shape theorem. (See, for example, [**68**, Lemma 3.6].) □

Using Lemma 5.5, along with the ergodic theorem, one has a.s. and in L^1,

$$\frac{1}{n}B(0, nx) = \frac{1}{n}\sum_{k=1}^{n} B((k-1)x, kx) \to x \cdot \rho \quad \text{for } x \in \mathbb{Z}^d.$$

As in (5.2), it is not difficult to upgrade this to the following sort of shape theorem for the Busemann function. See, for example, Damron-Hanson [**68**, Section 4], where this is done for a "reconstructed Busemann function."

LEMMA 5.6. *Assume* (5.4) *and Hoffman's conditions 1–4. For each* $\epsilon > 0$,

$$\mathbb{P}\left(|B(0, x) - x \cdot \rho| > \epsilon\|x\|_1 \text{ for only finitely many } x \in \mathbb{Z}^d\right) = 1,$$

where $\rho = e_1\mu(e_1)$.

We now show how assumption (5.4), along with an "exposed point" assumption, implies existence of geodesic rays with asymptotic direction e_1. Let S be the set

$$S = \partial\mathcal{B} \cap \{w : \rho \cdot w = 1\}.$$

Because the rightmost set is a supporting hyperplane for the limit shape at the point $e_1/\mu(e_1)$, the set S is a portion of the boundary $\partial\mathcal{B}$ containing $e_1/\mu(e_1)$. We will say that a geodesic ray with vertices $x_0, x_1, \ldots$ is asymptotically directed in S if each limit point of the sequence $(x_n/\mu(x_n))$ is contained in S. If S contains only one point, we then say that this ray has an asymptotic direction. If the point $e_1/\mu(e_1)$ is exposed (there is a supporting hyperplane for $\mathcal{B}$ that touches $\mathcal{B}$ only at $e_1/\mu(e_1)$), then by symmetry, $S = \{e_1/\mu(e_1)\}$. In that case, the following theorem implies that a.s., there is an infinite geodesic starting from 0 that has asymptotic direction e_1. It is an open problem to show without assumption (5.4) that with positive probability, there is a geodesic ray with an asymptotic direction:

QUESTION 5.3.2. Show that under general assumptions on the edge-weights (τ_e) and for some $d \ge 2$, there a.s. exists an infinite geodesic with vertices $0 = x_0, x_1, \ldots$ that has an asymptotic direction; that is, the sequence $(x_n/\mu(x_n))$ converges.

The following theorem is similar to [**68**, Theorem 5.3].

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

THEOREM 5.7. *Assume* (5.4) *and Hoffman's conditions 1–4. A.s., every subsequential limit of geodesics from 0 to H_n is asymptotically directed in S.*

PROOF. Let Γ be any subsequential limit of geodesics from 0 to H_n (which must be self-avoiding by our assumption of uniqueness of passage times) and label its vertices $0 = x_0, x_1, x_2, \ldots$. First note that for $m \geq 0$, one has

$$B(0, x_n) = T(0, x_n). \tag{5.6}$$

Indeed, choose (n_k) to be a subsequence such that some geodesics Γ_{n_k} from 0 to H_{n_k} converge to Γ as $k \to \infty$. For k large, the first n vertices of Γ_{n_k} coincide with those of Γ, so

$$\begin{aligned} B(0, x_n) &= \lim_{N \to \infty} [T(0, H_N) - T(x_n, H_N)] \\ &= \lim_{k \to \infty} [T(0, H_{n_k}) - T(x_n, H_{n_k})] = T(0, x_n). \end{aligned}$$

Choose z to be any limit point of the sequence $(x_n/\mu(x_n))$ so that for some subsequence (n_j), one has $x_{n_j}/\mu(x_{n_j}) \to z$. Now by Lemma 5.6, a.s. for any choice of Γ and (n_j),

$$\lim_{j \to \infty} \frac{B(0, x_{n_j})}{\mu(x_{n_j})} = \lim_{j \to \infty} \rho \cdot (x_{n_j}/\mu(x_{n_j})) = \rho \cdot z.$$

On the other hand, by the shape theorem and (5.6), this equals

$$\lim_{j \to \infty} \frac{T(0, x_{n_j})}{\mu(x_{n_j})} = 1,$$

and so $\rho \cdot z = 1$. Since $z \in \partial \mathcal{B}$, this means $z \in S$. □

COROLLARY 5.8. *Assume* (5.4)*, Hoffman's conditions 1–4, and that $e_1/\mu(e_1)$ is an exposed point of $\mathcal{B}$. A.s., every subsequential limit of geodesics from 0 to H_n has asymptotic direction e_1.*

5.4. Busemann increment distributions and geodesic graphs

Here we will explain one method of getting around Question 5.3.1 to obtain results about directions of geodesic rays. This method shares some similarities with the plan of attack used in the original paper of Kingman, described in Section 2.6. (We make this relation precise below in Remark 5.16.) Instead of trying to establish existence of the limit (5.4), we will focus on a weak limit by considering translational averages. We follow Damron-Hanson [**68**]. Their results were stated only for two dimensions, but the existence of geodesics directed in sectors (Theorem 5.9) also holds for general dimensions, so we outline an argument for this expanded statement.

We will focus on i.i.d. distributions, although the theorems apply to a wide class of translation-ergodic measures (similar to Hoffman's conditions from Section 5.2). Consider the following conditions on the common distribution function F for i.i.d. edge-weights:

I. $\mathbb{E}Y^d < \infty$, where Y is the minimum of $2d$ i.i.d. edge-weights. Furthermore, $F(0) < p_c$.
II. F is continuous.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Recall that a supporting hyperplane H for the limit shape $\mathcal{B}$ at a point $z_0 \in \partial\mathcal{B}$ is a hyperplane that contains z_0 and such that $\mathcal{B}$ does not intersect both components of H^c. If there is only one such hyperplane, we say that $\partial\mathcal{B}$ is differentiable at z_0. In this case, we write H_{z_0} for this hyperplane. Write

$$S_{z_0} = H_{z_0} \cap \partial\mathcal{B}.$$

THEOREM 5.9 (Damron-Hanson [**68**], Theorem 1.1). *Assume I. If $\partial\mathcal{B}$ is differentiable at z_0, then a.s., there is a geodesic ray Γ containing the origin which is asymptotically directed in S_{z_0}. This means every limit point of the set $\{x/\mu(x) : x \in \Gamma\}$ is contained in S_{z_0}.*

Note that if z_0 is an exposed point of differentiability, then there is a geodesic ray with asymptotic direction z_0, since in that case, $S_{z_0} = \{z_0\}$. This is an improvement on Newman's theorem (Theorem 4.25) because we have replaced the global curvature condition with a local, directional condition.

The next result is related to coalescence of geodesic rays.

THEOREM 5.10 (Damron-Hanson [**68**], Theorem 1.11). *Take dimension $d = 2$, assume I and II and that $\partial\mathcal{B}$ is differentiable at z_0. With probability one, there exists a collection of geodesic rays $(\Gamma_x : x \in \mathbb{Z}^2)$ satisfying the following properties:*

(1) *Each $x \in \mathbb{Z}^2$ is a vertex of Γ_x.*
(2) *(Directedness) Each Γ_x is asymptotically directed in S_{z_0}.*
(3) *(Coalescence) For all $x, y \in \mathbb{Z}^2$, the paths Γ_x and Γ_y coalesce.*
(4) *(Finiteness of backward paths) Each $x \in \mathbb{Z}^2$ is on Γ_y for only finitely many $y \in \mathbb{Z}^2$.*

As discussed in Section 4.4.2, Licea-Newman [**139**] (along with an improvement by Zerner [**148**]) have shown under the assumption of uniformly positive curvature for $\mathcal{B}$ that there exists a deterministic set

$$D \subset [0, 2\pi)$$

with countable complement such that, for each $\theta \in D$, the following holds a.s..

(1) (Existence) There exists a collection of infinite geodesics $(\Gamma_x : x \in \mathbb{Z}^2)$ such that each Γ_x starts from x, has asymptotic direction θ, and each Γ_x and Γ_y coalesce.
(2) (Uniqueness) For each x, Γ_x above is the only infinite geodesic starting from x with asymptotic direction θ.

The above theorem shows that one can take $D = [0, 2\pi)$ for existence, still under global conditions on the limit shape, but does not address uniqueness, which will be the focus of Section 5.6.

PROOF OUTLINE. This outline differs from what appears in [**68**] in that some arguments are simplified and extended to $\mathbb{Z}^d$. The idea is to work with subsequential Busemann limits in distribution. First we set $\Omega_1 = [0, \infty)^{\mathcal{E}^d}$ to be a copy of our edge-weight space with our i.i.d. joint edge-weight distribution $\mathbb{P}$ and let $\Omega_2 = \mathbb{R}^{\mathbb{Z}^d \times \mathbb{Z}^d}$ be our space for recording Busemann increments. Last, we have a space $\Omega_3 = \{0,1\}^{\vec{\mathcal{E}}^d}$, where $\vec{\mathcal{E}}^d$ is the set of oriented nearest-neighbor edges of $\mathbb{Z}^d$, which will record geodesic graphs. Put $\widetilde{\Omega} = \prod_{i=1}^3 \Omega_i$.

Let H be any supporting hyperplane for $\mathcal{B}$ at z_0 and let ρ be the vector with

$$H = \{w : w \cdot \rho = 1\}.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Define

$$H_\alpha = \{w : w \cdot \rho = \alpha\}$$

for $\alpha \in \mathbb{R}$. We now define Busemann increments and geodesic graphs toward H_α. For an outcome $\omega \in \Omega_1$ (written also as an edge-weight configuration (τ_e)), set

$$B_\alpha(\omega) = \big(B_\alpha(x,y) : x, y \in \mathbb{Z}^d\big) \in \Omega_2,$$

where

$$B_\alpha(x,y) = T(x, H_\alpha) - T(y, H_\alpha).$$

Furthermore define the geodesic graph configuration $\eta_\alpha(\omega)$ by

$$\eta_\alpha(\omega)((x,y)) = \begin{cases} 1 & \text{if } \{x,y\} \in \overline{\mathrm{GEO}}(z, H_\alpha) \text{ for some } z \text{ and } B_\alpha(x,y) \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Last, define $\Phi_\alpha : \Omega_1 \to \widetilde{\Omega}$ by

$$\Phi_\alpha(\omega) = (\omega, B_\alpha(\omega), \eta_\alpha(\omega))$$

and μ_α to be the push-forward of μ through Φ_α. The measure μ_α is a Borel probability measure on $\widetilde{\Omega}$.

A generic element of $\widetilde{\Omega}$ we write as

$$\widetilde{\omega} = ((\tau_e), B, \eta).$$

The coordinate B is called the reconstructed Busemann function. We would like to take α to infinity and sample a configuration $\widetilde{\omega}$ from the limit of (μ_α). Such a measure would be the distribution of Busemann increments and geodesic graphs "at infinity" in a direction related to H. However, we do not know how to show that this sequence of measures converges, so we settle for subsequential limits of averages of μ_α over α. The averaging is done to ensure that any limit is translation-invariant. That is, set

$$\mu_k^* = \frac{1}{k}\int_0^k \mu_\alpha \, \mathrm{d}\alpha$$

and let μ^* be any subsequential limit of (μ_k^*). (Tightness is not difficult to verify, due to the inequality $|B_\alpha(x,y)| \leq T(x,y)$.) Sampling from the measure μ_k^* can be thought of as representing first sampling a hyperplane H_α uniformly at random for $\alpha \in [0,k]$ and then sampling the configuration $\Phi_\alpha(\omega)$.

The idea is then to approximate the proofs of Section 5.3 to (a) find asymptotics of the function $B(0,x)$ as $|x| \to \infty$ and (b) use them to control geodesics. First, we give the mean of B, writing $\mathbb{E}_{\mu^*}$ for expectation relative to μ^*. Recall that ρ is the vector with $H_\alpha = \{w : w \cdot \rho = \alpha\}$.

LEMMA 5.11. *Assume I. The mean of the reconstructed Busemann function is*

$$\mathbb{E}_{\mu^*} B(x,y) = \rho \cdot (y - x) \quad \textit{for } x, y \in \mathbb{Z}^d.$$

The idea of this lemma is that the Busemann function on average measures distances traveled perpendicular to H, but from a sort of stationary state, where the velocity is given by the time constant in direction z_0 (which is the same time constant corresponding to the hyperplane H). The reader should compare to Lemma 5.5 in the previous section, where the mean of the Busemann function is similarly given by a projection along the vertical hyperplane H used to create the function.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

PROOF. The proof uses a version of the averaging trick (inspired by Gouéré [**100**]) from last section. For $k \geq 1$ we can use stationarity and the fact that $H_\alpha + x = H_{\alpha+\rho\cdot x}$ to write

$$
\begin{aligned}
\mathbb{E}_{\mu_k^*} B(-x,0) &= \frac{1}{k}\left[\int_0^k \mathbb{E}T(-x,H_\alpha)\,\mathrm{d}\alpha - \int_0^k \mathbb{E}T(0,H_\alpha)\,\mathrm{d}\alpha\right] \\
&= \frac{1}{k}\left[\int_0^k \mathbb{E}T(0,H_{\alpha+\rho\cdot x})\,\mathrm{d}\alpha - \int_0^k \mathbb{E}T(0,H_\alpha)\,\mathrm{d}\alpha\right] \\
&= \frac{1}{k}\left[\int_k^{k+\rho\cdot x} \mathbb{E}T(0,H_\alpha)\,\mathrm{d}\alpha - \int_0^{\rho\cdot x} \mathbb{E}T(0,H_\alpha)\,\mathrm{d}\alpha\right].
\end{aligned}
$$

Choosing k_j so that $\mu_{k_j}^* \to \mu^*$, one can show using translation invariance and the bound

$$\mathbb{E}_{\mu_{k_j}^*} B(-x,0)^2 \leq \mathbb{E}T(-x,0)^2$$

that the left side $\mathbb{E}_{\mu_k^*} B(-x,0)$ converges to $\mathbb{E}_{\mu^*} B(0,x)$. Noting that the last term vanishes in the limit when divided by k, we can change variables in the first term to obtain the formula

$$\mathbb{E}_{\mu^*} B(0,x) = \lim_{j\to\infty} \int_0^{\rho\cdot x} \frac{\mathbb{E}T(0,H_{\alpha+k_j})}{k_j}\,\mathrm{d}\alpha.$$

One can show using the shape theorem (see for example [**68**, Lemma 3.6]) that the integrand converges to 1, so by the dominated convergence theorem, we obtain a limit of $\rho \cdot x$. By translation invariance, then, of μ^*, one has

$$\mathbb{E}_{\mu^*} B(x,y) = \mathbb{E}_{\mu^*} B(0,y-x) = \rho \cdot (y-x).$$

$\square$

As usual, we can prove asymptotics for the reconstructed Busemann function in the form of a shape theorem. However, due to the extra randomness introduced through α, we may have lost ergodicity for μ^* under translations, and therefore obtain a random "shape."

LEMMA 5.12 (Shape theorem for B). *Assume I. There exists a random vector $\varrho \in \mathbb{R}^2$ such that for any $\epsilon > 0$,*

$$\mu^*\left(|B(0,x) - x\cdot\varrho| > \epsilon\|x\|_1 \text{ for infinitely many } x \in \mathbb{Z}^d\right) = 0.$$

The vector ϱ satisfies the following conditions:

(1) *μ^*-a.s., the hyperplane*

$$H_\varrho := \{w \in \mathbb{R}^d : w\cdot\varrho = 1\}$$

is a supporting hyperplane for $\mathcal{B}$ at z_0.

(2) *The mean of ϱ under μ^* is ρ.*

The proof of the above result establishes first radial limits of the form $\lim_{n\to\infty} B(0,nx)/n$ and patches them together using shape theorem arguments. A key tool is that the vector ϱ is invariant under translating the edge-weights, Busemann increments, and geodesic graphs. The mean of ϱ follows directly from the Lemma 5.11, whereas the fact that H_ϱ is a supporting line for $\mathcal{B}$ follows from

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

the statements (a) $\mu^*(x \cdot \varrho \leq 1) = 1$ for each $x \in \mathcal{B}$ and (b) $\mu^*(z_0 \cdot \varrho = 1) = 1$. The claim (a) is shown by noting that for $x \in \mathcal{B}$, one has

$$x \cdot \varrho = \lim_{n\to\infty} \frac{B(0, nx)}{n} \leq \lim_{n\to\infty} \frac{T(0, nx)}{n} = \mu(x) \leq 1.$$

In particular,

$$\mu^*(\rho \cdot z_0 \leq 1) = 1. \tag{5.7}$$

On the other hand, using the mean of ϱ, one has

$$\mathbb{E}_{\mu^*}(\varrho \cdot z_0) = \rho \cdot z_0 = 1.$$

Combining this with (5.7) gives $\varrho \cdot z_0 = 1$ with μ^*-probability one.

At this point, we may identify the asymptotics of the reconstructed Busemann function if $\partial\mathcal{B}$ is differentiable at z_0. In this case, H is the unique supporting hyperplane H_{z_0} for $\mathcal{B}$ at z_0, so this must be H_ϱ. In other words, one has

COROLLARY 5.13. *Assume I. If $\partial\mathcal{B}$ is differentiable at z_0, then*

$$\mu^*(\varrho = \rho) = 1.$$

In the presence of differentiability, the asymptotics of the Busemann function are given exactly by projection onto the hyperplane H used to create the function. This result gives us a major piece needed to establish directional properties of geodesic rays sampled from μ^*. To complete the proofs of the directional results, we simply need to argue as in the proof of Theorem 5.7, combining asymptotics of B with the shape theorem for T.

To do this, we consider the configuration $\eta \in \Omega_3$ sampled from μ^* and build a directed graph from it. Specifically, we set $\mathbb{G} = \mathbb{G}(\eta)$ to be the directed graph with vertex set equal to $\mathbb{Z}^d$ and with edge set equal to $\{(x, y) : \eta((x, y)) = 1\}$. Simple properties of geodesics carry over to the weak limit. For example: with μ^*-probability one, for all $x, y \in \mathbb{Z}^d$,

- Each directed path in $\mathbb{G}$ is a geodesic and from each x there is an infinite self-avoiding directed path.
- If $x \to y$ in $\mathbb{G}$ (there is a directed path from x to y in $\mathbb{G}$), then $B(x, y) = T(x, y)$.
- Under assumption II, viewed as an undirected graph, $\mathbb{G}$ has no circuits.
- Each vertex x has out-degree at least 1 in $\mathbb{G}$. Under assumption II, the out-degree is exactly 1, and thus emanating from each x is a unique infinite directed path Γ_x.

Given these properties, the proof of the next result is analogous to that of Theorem 5.7, so we omit it.

THEOREM 5.14 (Damron-Hanson [**68**], Theorem 5.3). *Assume I. With μ^*-probability 1, each directed infinite path starting from 0 in $\mathbb{G}$ is asymptotically directed in $S_\varrho = \partial\mathcal{B} \cap H_\varrho$.*

In the case that $\partial\mathcal{B}$ is differentiable at z_0, the hyperplane H_ϱ is μ^*-a.s. equal to H_{z_0} and therefore $S_\varrho = S_{z_0}$. Therefore Theorem 5.14 implies a version of Theorem 5.9 on the space $\widetilde{\Omega}$. To deduce the result on the original space, one applies a pull-back argument using the regular conditional measure of μ^* given the edge-weights.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

For Theorem 5.10, item 1 is a restatement of Theorem 5.9 using translation invariance of μ^*. The proof of the coalescence statement follows the argument of Licea-Newman outlined in Theorem 4.28. The main difference is that one must be careful to construct barrier events on the original space Ω_1 and port them over to $\widetilde{\Omega}$, and this creates a considerable headache. We refer the reader to [**68**, Section 6]. As for absence of infinite backward paths, we have

THEOREM 5.15 (Damron-Hanson [**68**], Theorem 6.9). *Assume I and II in dimension $d = 2$. With μ^*-probability 1, the set $\{y : 0 \in \Gamma_y\}$ is finite.*

PROOF. The idea is a modification of that of Burton-Keane [**46**], from their proof of uniqueness of the infinite occupied cluster in Bernoulli percolation. For $x \in \mathbb{Z}^2$, define the event A_x that in $\mathbb{G}$, x has an infinite forward path and two disjoint infinite backward paths. On A_x, the vertex x is, in the language of [**46**], an "encounter point." A counting argument (see [**46**, p. 504]) shows that in the box $B(n) = [-n, n]^2$, one has deterministically

$$|\{x \in B(n) : A_x \text{ occurs}\}| \leq Cn,$$

as the number of points for which A_x occurs is bounded above by the number of points on the boundary of $B(n)$. On the other hand, if we show that $\mathbb{P}(A_x) > 0$, then translation invariance of μ^* gives

$$\mathbb{E}_{\mu^*}|\{x \in B(n) : A_x \text{ occurs}\}| = |B(n)|\mu^*(A_x) = (2n+1)^2\mu^*(A_x),$$

and this would be a contradiction for large n.

Assume then that with positive probability, the set $\{y : 0 \in \Gamma_y\}$ is infinite; we will show that this implies $\mu^*(A_x) > 0$. By directedness of paths in $\mathbb{G}$, we may find a line L intersecting $\mathbb{Z}^2$ (which we will for illustration take to be the e_2-axis) such that each Γ_y eventually stays on one fixed side of L, say $\{z : z \cdot e_1 > 0\}$. Choose y on the other side $\{z : z \cdot e_1 < 0\}$ of L and follow Γ_y on the event that $\{z : y \in \Gamma_z\}$ is infinite until its last intersection with L. Such a point has an infinite forward path in $\mathbb{G}$ that does not intersect L except at its initial point, and has an infinite backward path. By the ergodic theorem, we can then find $x_1, x_2 \in L$ such that with positive μ^*-probability, each of x_1 and x_2 has an infinite forward path that does not touch L except at the initial point and has an infinite backward path. Because all forward paths in $\mathbb{G}$ coalesce, we may choose x to be the coalescence point of Γ_{x_1} and Γ_{x_2} and so A_x occurs. Therefore there is an x such that $\mu^*(A_x) > 0$. ☐

Combining the above results with an argument to pull back to Ω_1 ends the proof of Theorem 5.10. ☐

REMARK 5.16. Here we explain the relation between the above method and Kingman's approach. The analogue of Kingman's variables defined in (2.20) and (2.21) would be, respectively (using the notation from (5.3))

$$f_k = \frac{1}{k}\sum_{j=1}^{k}(T(0, H_j) - T(e_1, H_j)) = \frac{1}{k}\sum_{j=1}^{k} B_j(0, e_1)$$

and the replacement for

$$f_k + Tf_k + \ldots + T^{n-1}f_k$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

would be $\frac{1}{k}\sum_{j=1}^{k}\left[B_j(0,e_1)+\cdots+B_{j+n-1}((n-1)e_1,ne_1)\right]$, but it is more natural for us to choose

$$\frac{1}{k}\sum_{j=1}^{k} B_j(0,ne_1).$$

Note that f_k is similar to the Busemann field coordinate B under the averaged measure μ_k^*, but the averaging is done one level lower (at the level of random variables rather than measures). To make this precise, suppose that f_k converges a.s. to some random variable f and that μ_k^* converges to a measure μ. Then one would have

$$f=\mathbb{E}_\mu\left[B(0,e_1)\mid(\tau_e)\right],$$

where we have conditioned on the edge-weight configuration (τ_e). In other words, f would be an average of the reconstructed Busemann function $B(0,e_1)$ over the additional randomness we introduced into μ (the uniformly random hyperplane H_α).

In fact, exactly this approach is taken in Liggett's proof of his improved subadditive ergodic theorem [**142**]. In that result, Kingman's assumptions are weakened, and the question of building a decomposition into an additive and nonnegative subadditive process in Theorem 2.30 is handled using weak limits, instead of employing weak-$*$ compactness to find a limit point for the sequence (f_k) in $(L^1)^{**}$, as Kingman did. Precisely, for a subadditive ergodic sequence $(X_{m,n})$, Liggett defined an independent uniform $\{1,\ldots,n\}$ random variable U_n (similar to the index of our uniformly random hyperplane H_α) and set

$$Y_i^{(n)}=X_{0,i+U_n}-X_{0,i+U_n-1}.$$

The distribution of $Y_i^{(n)}$ is similar to that of our $B(0,e_1)$ under μ_n^*, but is reflected. Then he set $(Y_1,Y_2,\ldots)$ to be any subsequential limit in distribution of the sequences $(Y_1^{(n)},Y_2^{(n)},\ldots)$ and used the distributional monotonicity

$$(Y_1,Y_1+Y_2,Y_1+Y_2+Y_3\ldots)\leq_{st}(X_{0,1},X_{0,2},X_{0,3}\ldots)$$

in place of Kingman's $f\leq X_{0,1}$.

One could ask whether the approach used in this section could work by only considering quantities at the level of the random variable f_k rather than at the level of μ_k^*. In other words, instead of averaging the distributions of the quantities

$$(T(x,H_\alpha)-T(y,H_\alpha))_{x,y}$$

over α, could we simply average the random variables as

$$\left(\frac{1}{n}\int_0^n (T(x,H_\alpha)-T(y,H_\alpha))\,\mathrm{d}\alpha\right)_{x,y},$$

which would be more similar to Kingman's original approach? This may not be possible in general translation-ergodic environments. Suppose that geodesics from 0 to H_n and e_1 to H_n have exactly two pairs of subsequential limiting geodesics $\Gamma_0^{(1)},\Gamma_0^{(2)}$ and $\Gamma_{e_1}^{(1)},\Gamma_{e_1}^{(2)}$. Then if $\Gamma_0^{(i)}$ and $\Gamma_{e_1}^{(i)}$ coalesce for each fixed i, one would expect the Busemann functions constructed from these limiting geodesics to be different. However it is still possible that the limit f of f_k exists, assuming the subsequences on which we took limits are regular enough. Then f is simply an average of these two Busemann functions corresponding to the different geodesics and for this reason, it will no longer necessarily have property (a) from Lemma 5.2.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

That is, if only one of the two geodesics passes over the edge from 0 to e_1, then it need not be true that $f = \tau_{\{0,e_1\}}$. This property is essential for deriving directional properties of geodesics from their Busemann functions.

5.5. Busemann functions along boundaries in $\mathbb{Z}^2$

In one case, point-to-point Busemann limits similar to (5.4) can be shown to exist deterministically, and this implies that limits exist a.s. for certain sequences of finite geodesics. We explain in this section the results of Auffinger-Damron-Hanson [**20**], where these statements are proved when Busemann limits are taken to points on boundaries of subsets in $\mathbb{Z}^2$. The existence of a boundary will allow for a "paths-crossing" trick due to Alm-Wierman [**13**] (first introduced in [**11**] in a special case), and existence of Busemann limits follows from this.

For simplicity, the results below will be stated for the half plane with vertices

$$V_H = \{(x,y) \in \mathbb{Z}^2 : y \geq 0\}$$

and edges

$$E_H = \{\{x,y\} : x, y \in V_H, |x-y| = 1\},$$

although we will remark about extending them to general subsets. For $x, y \in V_H$, let $T_H(x,y)$ be the minimum passage time among all paths with all vertices in V_H from x to y.

THEOREM 5.17 (Auffinger-Damron-Hanson [**20**], Theorem 1.1). *Let $x_n = ne_1$ and let (τ_e) be any edge-weight configuration in $[0,\infty)^{E_H}$. For all $x, y \in V_H$, the Busemann limit to x_n exists:*

$$B_H(x,y) = \lim_{n\to\infty} \left[T_H(x,x_n) - T_H(y,x_n)\right]. \tag{5.8}$$

As usual, existence of Busemann limits gives us quite a bit of control on geodesics. Using Theorem 5.17, we can prove existence of limiting geodesic graphs. Formally, we represent geodesic graphs as elements of a directed graph space, as in Section 5.4. Let $\vec{E}_H$ be the set of directed edges of V_H

$$\vec{E}_H = \{(x,y) : x, y \in V_H, |x-y| = 1\}$$

and write η for an arbitrary element of $\{0,1\}^{\vec{E}_H}$. Build the graph $\mathbb{G} = \mathbb{G}(\eta)$ as before: the vertices are all the vertices of V_H and an edge $e \in \vec{E}_H$ is present in the graph if and only if $\eta(e) = 1$. For the sequence of vertices $(x_n) = (ne_1)$, we let $\eta_n(e) = 1$ if $e = (x,y)$ is in a geodesic from some point to x_n and

$$T_H(x,x_n) \geq T_H(y,x_n);$$

we then set $\mathbb{G}_n = \mathbb{G}(\eta_n)$. The graphs $\mathbb{G}_n$ converge to a graph $\mathbb{G} = \mathbb{G}(\eta)$ if for each $e \in \vec{E}_H$, one has $\eta_n(e) \to \eta(e)$.

THEOREM 5.18 (Auffinger-Damron-Hanson [**20**], Theorem 1.3). *Let $\mathbb{P}$ be a probability measure on $[0,\infty)^{E_H}$ such that*

$$\mathbb{P}(\exists \textit{ geodesic between } x,y \textit{ for all } x,y \in V_H) = 1. \tag{5.9}$$

Then with a.s., $(\mathbb{G}_n)$ converges to a graph $\mathbb{G}$. Each directed path in $\mathbb{G}$ is a geodesic.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

REMARK 5.19. The above two theorems are valid in the following more general setting. Let V be any subset of $\mathbb{Z}^2$ that is connected and infinite and has infinite connected complement. (The reader can think of slit planes, sectors, etc.) There is a unique doubly infinite path in the dual lattice $\mathbb{Z}^2 + (1/2, 1/2)$ which separates V and V^c. Enumerate the vertices in V that are adjacent to this dual path as

$$\dots, x_{-1}, x_0, x_1, \dots.$$

If we denote passage times between vertices x, y in V as $T_V(x, y)$, the minimum passage time among all paths with all vertices in V connecting x and y, then the Busemann limits

$$B_V(x, y) = \lim_{n\to\infty} [T_V(x, x_n) - T_V(y, x_n)]$$

exist for all edge-weight configurations in $[0, \infty)^{E_V}$, and limiting geodesic graphs exist a.s. for any measure on $[0, \infty)^{E_V}$ that has the geodesic property (5.9). Here, E_V is the set of nearest-neighbor edges with both endpoints in V.

In the case of the half-plane, one can say more about the structure of the limiting graph $\mathbb{G}$. We will only state the theorem (with a restricted class of edge-weight distributions), and refer to [**20**] for the complete proof.

THEOREM 5.20 (Auffinger-Damron-Hanson [**20**], Theorem 1.5). *Let $\mathbb{P}$ be an i.i.d. product measure on $[0, \infty)^{E_H}$ with continuous marginals. The limiting geodesic graph $\mathbb{G}$ from Theorem 5.18 satisfies the following a.s..*

(1) *Each vertex in V_H has out-degree 1. Therefore from each $x \in V_H$ emanates a unique infinite directed path Γ_x.*
(2) *Viewed as an undirected graph, $\mathbb{G}$ has no circuits.*
(3) *(Coalescence) For all $x, y \in V_H$, Γ_x and Γ_y coalesce. That is, their edge symmetric difference is finite.*
(4) *(Finiteness of backward paths) For each $x \in V_H$, the backward cluster $\{y \in V_H : x \in \Gamma_y\}$ is finite.*

PROOF. We give the ideas of the proofs of Theorems 5.17 and 5.18. For Theorem 5.17, existence of Busemann limits as in (5.8) are proved as a consequence of the "paths-crossing" trick of Alm and Wierman [**13**]. We begin by showing the limit for $x = m_1e_1, y = m_2e_1$ with $m_1 < m_2 \in \mathbb{Z}$. For simplicity, let us assume that for each $w, z \in V_H$, there is a geodesic between w and z. If this is not the case, then the geodesics can be replaced by paths that have passage time within ϵ of the infimum.

Let $n_2 > n_1 > m_2$ and let σ_1 be a geodesic from x to n_1e_1, with σ_2 a geodesic from y to n_2e_1. Note that by planarity, the paths σ_1 and σ_2 must share a vertex z. Define the path $\hat{\sigma}_1$ by traversing σ_1 from x to z and then σ_2 from z to n_2e_1. Define $\hat{\sigma}_2$ by traversing σ_2 from y to z and then σ_1 from z to n_1e_1. Then (see Figure 13),

$$\begin{aligned} T_H(x, n_1e_1) + T_H(y, n_2e_1) = T_H(\sigma_1) + T_H(\sigma_2) &= T_H(\hat{\sigma}_1) + T_H(\hat{\sigma}_2) \\ &\geq T_H(x, n_2e_1) + T_H(y, n_1e_1). \end{aligned}$$

Rearranging this inequality, we obtain

$$T_H(x, n_1e_1) - T_H(y, n_1e_1) \geq T_H(x, n_2e_1) - T_H(y, n_2e_1).$$

Therefore the sequence in (5.8) is monotone (and bounded by subadditivity) and the limit exists.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

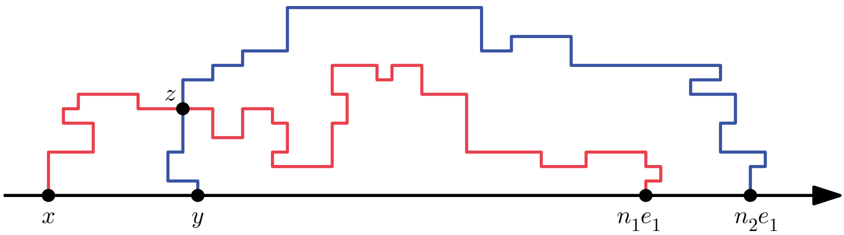


FIGURE 13. "Paths-crossing" trick: the path σ_1 from x to n_1e_1 and the path σ_2 from y to n_2e_1 must intersect at a point z.

To generalize this argument to points off the axis, we take $x \in V_H$ to be of the form (x_1, x_2) with $x_2 > 0$ and $x' \in V_H$. Let B be the box $[-k, k] \times [0, k]$ for some k large enough so that B contains both x and x'. The set $V' := V_H \setminus B$ is connected and infinite with complement that is also connected and infinite. So we define for $y, z \in V'$ the restricted passage time $T'(y, z)$ to be the minimum passage time of all paths from y to z which have only vertices in V'. One may then repeat the above "paths-crossing" trick to see that if y, z are in $\partial B \cap V'$ (they are each in V' but have a $\mathbb{Z}^2$-neighbor which is in B), then the limit

$$B'(y, z) = \lim_{n\to\infty} [T'(y, x_n) - T'(z, x_n)]$$

exists. Furthermore, for all such y, z, the sequence defining $B'(y, z)$ is monotone.

Now the idea is to use existence of Busemann limits $B'(y, z)$ for $y, z \in \partial B \cap V'$ to prove existence for x and x'. The crucial point is that for large n,

$$T_H(x, x_n) = \min\{T_H(x, y) + T'(y, x_n) : y \in \partial B \cap V'\}.$$

For our point x, another fixed point $z \in \partial B \cap V'$, and $y \in \partial B \cap V'$ variable, we set

$$\psi_n(y) = T_H(x, y) + T'(y, x_n) - T'(z, x_n).$$

We find then that $\psi_n(y)$ has a finite limit $\psi(y) = T_H(x, y) + B'(y, z)$ for each y, and furthermore

$$\lim_{n\to\infty} [T_H(x, x_n) - T'(z, x_n)] = \min\{\psi(y) : y \in \partial B \cap V'\}.$$

Reversing the roles of x and x', we find that $\lim_{n\to\infty} [T_H(x, x_n) - T_H(x', x_n)]$ exists.

To prove existence of limiting geodesic graphs, we take $e = (x, y) \in \vec{E}_H$ and again let B be a box of the form $[-k, k] \times [0, k]$ which is large enough to contain x and y. If N is large enough so that $x_n \notin B$ for all $n \geq N$, then if we again put $V' = V_H \setminus B$ one has for all $y \in \partial B \cap V'$,

$$y \text{ is on a geodesic from } x \text{ to } x_n \Leftrightarrow \psi_n(y) = \min\{\psi_n(u) : u \in \partial B \cap V'\},$$

where $\psi_n(y)$ was defined as above as $T_H(x, y) + T'(y, x_n) - T'(z, x_n)$ for the fixed $z \in \partial B \cap V'$. However, one can show using monotonicity of the differences $T'(y, x_n) - T'(z, x_n)$ in n for $y \in \partial B \cap V'$ that the set of minimizers $\mathfrak{m}_n$ of ψ_n is eventually constant for n large. Thus, the set of points in $\partial B \cap V'$ which are in geodesics from x to x_n is eventually constant in n.

The above work readily implies that $(\mathbb{G}_n)$ converges. Indeed, one can check that for our $e = (x, y)$, one has $\eta_n(e) = 1$ if and only if e is in a geodesic from x to a vertex of $\mathfrak{m}_n$. Since $\mathfrak{m}_n$ is eventually constant, the value of $\eta_n(e)$ must also be eventually constant. Since e was arbitrary, this completes the proof. $\square$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

5.6. Nonexistence of bigeodesics in fixed directions

Recently, Busemann functions have been used in general last-passage percolation (LPP) models by Georgiou, Rassoul-Agha, and Seppäläinen to prove analogues (and improved versions) of the directional results from [**68**] in FPP. This is a big advance because most work on LPP has assumed exponential or geometric weights, where random matrix and queueing theory analysis can be used.

A main tool in the general two-dimensional LPP models is the directedness of paths, which allows one to use various forms of the "paths-crossing" argument of Alm and Wierman. Applied to Busemann functions, this gives a certain monotonicity of directional Busemann functions. In FPP, paths are not directed, and this creates a fundamental obstacle to obtaining similar improvements.

In this section, we describe recent work of Damron-Hanson [**69**] which shows that the LPP results are also valid in FPP. These theorems address the "uniqueness" issue as described below Theorem 5.10, and consequently rule out bigeodesics in fixed directions. In that paper, the following theorem is proved for general edge-weight distributions considered in the first work of Damron-Hanson [**68**], but we will focus again on i.i.d. weights, assuming I and II from Section 5.4.

We will again make some assumptions on the limiting shape: let $\theta \in [0, 2\pi)$ and let v_θ be the point on $\partial\mathcal{B}$ in direction θ. When v_θ is a point of differentiability of $\partial\mathcal{B}$, let L_θ be the unique tangent line and let S_θ be the sector of angles of contact of L_θ with $\partial\mathcal{B}$. Let θ_1 and θ_2 be the endpoints of S_θ. We say that a sequence (x_n) of distinct points of $\mathbb{Z}^2$ is directed in S_θ if all limit points of the sequence of arguments $(\arg x_n)$ are contained in S_θ. As before, we say an infinite geodesic with vertices (x_n) is directed in S_θ if the sequence (x_n) is.

Theorem 5.21 (Damron-Hanson [**69**], Theorem 1). *Assume I, II, and that $\partial\mathcal{B}$ is differentiable at θ, θ_1, and θ_2. The following hold a.s..*

(1) *(Uniqueness) For each $x \in \mathbb{Z}^2$, there is exactly one infinite geodesic Γ_x that is directed in S_θ such that for any (random) sequence (x_n) of distinct points directed in S_θ,*
$$\Gamma_x = \lim_{n\to\infty} \mathrm{Geo}(x, x_n).$$
(2) *(Coalescence) For each $x, y \in \mathbb{Z}^2$, the geodesics Γ_x and Γ_y coalesce.*
(3) *(Finiteness of backward paths) There are no bigeodesics with one end directed in S_θ.*

Notice that this result shows Licea-Newman's set D (see Theorem 4.26) of directions to be equal to $[0, 2\pi)$. Furthermore, their result rules out bigeodesics with both ends directed in D, whereas the above theorem only requires directedness of one end (in a sector). The reader should also consult the recent preprint of Ahlberg and Hoffman [**5**] for more work on these questions.

From this result we obtain the following progress on the bigeodesic conjecture.

Corollary 5.22. *Assume I, II and that $\partial\mathcal{B}$ is differentiable. Then for each θ,*
$$\mathbb{P}(\textit{there is a bigeodesic with one end in direction } \theta) = 0.$$

Remark 5.23. Since geodesic lines with fixed directions cannot exist, one should ask if infinite geodesics are even required to have directions. One can show using planarity and the results of [**68**] that if $\partial\mathcal{B}$ is differentiable and I and II hold, then the following statements are true a.s.:

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

(1) for all θ, there is an infinite geodesic starting from 0 directed in S_θ and
(2) every infinite geodesic is directed in S_θ for some θ.

As a consequence of Theorem 5.21 and arguments like those presented elsewhere in this section (for instance in Section 5.3), one has the following result.

THEOREM 5.24 (Damron-Hanson [**69**], Theorem 2). *Assume I, II, and that $\partial\mathcal{B}$ is differentiable at θ, θ_1, and θ_2. A.s., for each $x, y \in \mathbb{Z}^2$ and (random) sequence (x_n) of distinct points directed in S_θ, the limit*

$$B(x,y) = \lim_n [T(x,x_n) - T(y,x_n)]$$

exists. Furthermore, letting ρ be the unique vector in $\mathbb{R}^2$ such that $\{r \in \mathbb{R}^2 : r\cdot\rho = 1\}$ is the tangent line to $\mathcal{B}$ in direction θ, one has:

(1) $\mathbb{E}B(0,x) = \rho \cdot x$ *for all* $x \in \mathbb{Z}^2$.
(2) *For each $\epsilon > 0$, the set of $x \in \mathbb{Z}^2$ such that $|B(0,x) - \rho \cdot x| > \epsilon \|x\|_1$ is finite.*

REMARK 5.25. The conclusions of Theorems 5.21 and 5.24 are also valid under a weaker condition on $\partial\mathcal{B}$. Namely one must only assume that θ_i is a limit of extreme points of $\mathcal{B}$ for each $i = 1, 2$.

In these theorems, one may ask the degree to which the differentiability assumption is necessary. The main obstacle to removing differentiability is the result of Häggstrom-Meester (see Remark 2.25 and Theorem 7.6), which we recall states that given any compact, convex subset $\mathcal{C}$ of $\mathbb{R}^d$ that is symmetric ($x \in \mathcal{C} \Rightarrow -x \in \mathcal{C}$), there exists a translation-ergodic distribution of edge-weights whose limit shape is $\mathcal{C}$. In particular, there are models of FPP whose limit shapes are polygons. In these situations, it is reasonable to believe that one can construct models in which the only infinite geodesics are ones which do not have asymptotic directions — they wander across the sectors corresponding to the sides of $\mathcal{B}$. This leads one to consider directedness in sectors. However, it is reasonable to expect existence of some translation-ergodic edge-weight distributions which have polygonal limit shapes, and for which there are infinite geodesics directed toward the corners. In this case, uniqueness of infinite geodesics in sectors corresponding to sides abutting such corners cannot hold. These situations are prevented by making a differentiability assumption on $\partial\mathcal{B}$.

REMARK 5.26. The above results give progress toward solving the "BKS midpoint problem" introduced in [**26**], and which was listed earlier as Question 3.2.1. The question is: is it true that the probability that $(n/2)e_1$ is in a geodesic from 0 to ne_1 goes to zero at $n \to \infty$? By a translation, if the answer is no, then

$$\mathbb{P}\left(0 \in \overline{\mathrm{GEO}}(-ne_1, ne_1) \text{ for infinitely many } n\right) > 0.$$

On this event, 0 is in a bigeodesic, and under the differentiability assumptions of Theorem 5.21, one of its ends is directed in the sector corresponding to $\theta = 0$. This is impossible by Theorem 5.21. Therefore under the differentiability assumption of Theorem 5.21 for $\theta = 0$, the answer to the midpoint problem is yes. For more recent work on this problem, the reader is directed to Ahlberg and Hoffman [**5**].

SKETCH OF PROOF OF THEOREM 5.21. The proof we give is more similar to the proof in last-passage percolation [**97**] because, for ease of exposition, we will

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

omit the many applications of the Jordan curve theorem needed to deal with the fact that paths in FPP are not necessarily oriented.

For simplicity, we will take $\theta = \pi/2$ and

$$\pi > \theta_1 \geq \theta \geq \theta_2 > 0.$$

We will first reduce to problem to a the corresponding one on the upper half-plane, where it is easy to put an ordering on infinite geodesics. We will skip this step, although it is the most work, because it does not involve Busemann functions, but just topological arguments. However we mention that this reduction is done using the fact that every path that is directed in S_θ must have at most finitely many intersections with

$$L_0 = \{x : x \cdot e_2 = 0\}.$$

So consider the upper half-plane with vertices V_H and nearest-neighbor edges E_H, and define passage times $T_H(x, y)$ for vertices $x, y \in V_H$ by considering only paths in the upper half-plane. We will content ourselves with proving the theorem for geodesics constructed in the upper half-plane using T_H.

Let $\mathcal{G}_H(x)$ be the collection of infinite geodesics starting from x in the upper half-plane which are directed in S_θ. Because all these geodesics are in the upper half-plane, for each $x \in L_0$, it is possible to define a leftmost (counterclockwisemost) infinite geodesic in $\mathcal{G}_H(x)$, written as $\Gamma^L_{x,H}$. Similarly there is a rightmost infinite geodesic written as $\Gamma^R_{x,H}$. Using our differentiability assumptions, one can show (see [**69**, Proposition 4.7]) that $\Gamma^L_{x,H}$ and $\Gamma^R_{x,H}$ themselves are directed in S_θ and the first lies counterclockwise to the second. The main step is to argue that that

$$\Gamma^L_{0,H} = \Gamma^R_{0,H} \quad \text{a.s..} \tag{5.10}$$

Once we show (5.10), then part 1 of Theorem 5.21 follows from the following lemma, whose proof uses the assumption that $\partial\mathcal{B}$ is differentiable at the endpoints θ_1, θ_2.

LEMMA 5.27. *A.s., the following holds. If (x_n) is any (random) sequence of points directed in S_θ, and Γ is a subsequential limit of upper half-plane geodesics from 0 to x_n, then Γ is directed in S_θ.*

The idea of the proof of this lemma is to use trapping. Since $\partial\mathcal{B}$ is differentiable at θ_1, one can find infinitely many extreme points of $\mathcal{B}$ which converge to v_{θ_1} from outside of S_θ. One then derives an upper half-plane version of the existence of infinite geodesics directed in sectors from [**68**] (see Theorem 5.9), and uses this to construct infinitely many geodesics which are directed in distinct sectors whose endpoints converge to v_{θ_1} from outside S_θ. The subsequential limit Γ cannot intersect any of these geodesics more than once, due to uniqueness of passage times, and so it must be directed "on or to the right" of θ_1. A similar argument works for θ_2 and traps Γ in S_θ.

If we assume (5.10) and use Lemma 5.27, we find that any subsequential limit Γ of half-plane geodesics from 0 to a (random) sequence (x_n) directed in S_θ must itself be directed in S_θ, and must then lie between $\Gamma^L_{0,H}$ and $\Gamma^R_{0,H}$. This would mean all three geodesics $\Gamma, \Gamma^L_{0,H}$, and $\Gamma^R_{0,H}$ are equal and would prove item 1 of Theorem 5.21.

Now we argue for (5.10). Similarly to the definition of leftmost and rightmost geodesics in the upper half-plane, one can with considerably more effort (see [**69**,

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Section 4]) define leftmost and rightmost infinite geodesics from $x \in \mathbb{Z}^2$ in the full-plane; we write these as Γ_x^L and Γ_x^R. One can show (see [**69**, Proposition 4.10]) that for $x \in L_0$, with positive probability,

$$\Gamma_{x,H}^L = \Gamma_x^L,$$

and similarly that with positive probability,

$$\Gamma_{x,H}^R = \Gamma_x^R.$$

This is argued using the fact that any path that is directed in S and starts on L_0 has a last intersection with L_0. One can then use a Licea-Newman style argument (as in Theorem 4.28) to show that each Γ_x^L and Γ_y^L coalesce for $x, y \in \mathbb{Z}^2$ (and similarly for rightmost geodesics) and then deduce this same statement for $\Gamma_{x,H}^L$ and $\Gamma_{y,H}^L$ for $x, y \in V_H$ (and similarly for rightmost upper half-plane geodesics). This allows one to define Busemann functions for $x, y \in V_H$ and $* = L$ or R:

$$B_H^*(x,y) = \lim_{n \to \infty} \left[T_H(x, x_n) - T_H(y, x_n)\right],$$

where (x_n) is the sequence of vertices on $\Gamma_{0,H}^*$. By coalescence, the definition would be the same if we replace $\Gamma_{0,H}^*$ by $\Gamma_{z,H}^*$ for any $z \in V_H$, and this fact implies a horizontal translation covariance similar to that in (5.5) for these Busemann functions. Indeed, if θ is a horizontal integer translation, then

$$B_H^*(\theta(x), \theta(y))(\theta(\omega)) = B_H^*(x,y)(\omega).$$

A similar definition is given for full-plane geodesics and in this case, one may define the Busemann functions on all of $\mathbb{Z}^2$, not just V_H, and they are translation covariant relative to all integer translations.

By the ergodic theorem (like in (5.1)), if we set

$$\Delta_H(x,y) = B_H^L(x,y) - B_H^R(x,y)$$

(and similarly for $\Delta(x,y)$ as the difference of full-plane Busemann functions), then there is a c such that

$$\frac{1}{n}\Delta_H(0, ne_1) \to c \text{ and } \frac{1}{n}\Delta_H(0, -ne_1) \to -c \quad \text{a.s..}$$

Furthermore, one can use equality of full-plane and half-plane geodesics (with positive probability) to show that these limits exist for Δ (the full-plane difference) and are equal to c and $-c$, respectively:

$$\frac{1}{n}\Delta(0, ne_1) \to c \text{ and } \frac{1}{n}\Delta(0, -ne_1) \to -c \quad \text{a.s..}$$

Now (5.10) follows from the next two lemmas.

LEMMA 5.28. *Under our differentiability assumption on $\partial\mathcal{B}$, $c = 0$.*

PROOF. For $* = L, R$ the full-plane function $x \mapsto \mathbb{E}B^*(0,x)$ is linear, and so as usual there is a vector ρ^* such that

$$\mathbb{E}B^*(x,y) = \rho^* \cdot (y - x) \quad \text{for } x, y \in \mathbb{Z}^2.$$

Once again (as in Lemma 5.6), one may upgrade this to a sort of shape theorem: for each $\epsilon > 0$,

$$\mathbb{P}\left(|B^*(0,x) - \rho^* \cdot x| > \epsilon \|x\|_1 \text{ for infinitely many } x \in \mathbb{Z}^2\right) = 0.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

To prove the lemma, we will show that for for $* = L, R$, the line

$$L^* = \{r \in \mathbb{R}^2 : r \cdot \rho^* = 1\}$$

is a supporting line for the limit shape $\mathcal{B}$ at some direction (and therefore any direction) of the sector S_θ. By our differentiability assumptions, there is only one such line, and so we can conclude

$$\rho^L = \rho^R.$$

In other words, this will show that

$$\mathbb{E}B^*(0, e_1) = \rho^L \cdot e_1$$

for both $* = L, R$, and the ergodic theorem will complete the proof.

The proof of the supporting property of L^* is similar to that of Theorem 5.7. Label the last intersections of Γ_0^* with

$$L_0, L_1, L_2, \ldots$$

as

$$x_0^*, x_1^*, x_2^*, \ldots.$$

(Here, $L_k = L_0 + ke_2$.) Then choose a subsequence $(x_{n_k}^*)$ such that $x_{n_k}^*/\mu(x_{n_k})$ converges to some vector $z_0 \in \partial\mathcal{B}$ (whose angle necessarily lies in S_θ). By the standard shape theorem,

$$B^*(0, x_n^*)/\mu(x_n) = T(0, x_n^*)/\mu(x_n) \to 1 \quad \text{as } n \to \infty.$$

On the other hand, this converges on the subsequence (n_k) to $\rho^* \cdot z_0$, so

$$\rho^* \cdot z_0 = 1.$$

In other words, z_0 is both a point of L^* and of $\partial\mathcal{B}$. Furthermore, as usual,

$$\rho^* \cdot z \leq 1$$

for all $z \in \mathcal{B}$, using the fact that Busemann functions are bounded by the passage time:

$$\rho^* \cdot z = \lim_{n\to\infty} \frac{B^*(0, nz)}{n} \leq \lim_{n\to\infty} \frac{T(0, nz)}{n} = \mu(z) \leq 1.$$

This means that L^* is a supporting line for $\mathcal{B}$ at z_0, which is a point with angle in S_θ. This completes the proof. □

LEMMA 5.29. *Almost surely,* $\Delta_H(0, e_1) \leq 0$. *If* $\Gamma_{0,H}^L \neq \Gamma_{0,H}^R$ *with positive probability, then* $c < 0$.

PROOF. If $\Gamma_{0,H}^L \neq \Gamma_{0,H}^R$, then by coalescence, the same must be true for geodesics starting from any point on L_0. Since the edge-weights have a continuous distribution, one then deduces that

$$\Delta(0, e_1) \neq 0.$$

So if we can show that

$$\Delta(0, e_1) \leq 0$$

a.s., one then has

$$\mathbb{E}\Delta(0, e_1) < 0,$$

and the ergodic theorem will prove the lemma.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

To show the a.s. inequality, we use the "paths-crossing" trick. We claim that $\Gamma^R_{0,H}$ and $\Gamma^L_{e_1,H}$ must share a vertex z. Indeed, if

$$\Gamma^L_{0,H} = \Gamma^R_{0,H},$$

then this holds by coalescence. Otherwise, if follows from planarity and the fact that $\Gamma^R_{0,H}$ coalesces with $\Gamma^R_{e_1,H}$, which is "to the right" of $\Gamma^L_{e_1,H}$. Let y be a vertex on $\Gamma^L_{e_1,H}$ beyond z and let y' be a vertex on $\Gamma^R_{0,H}$ beyond z. Then

$$\begin{aligned} T_H(0,y') + T_H(e_1,y) &= T_H(0,z) + T_H(z,y') + T_H(e_1,z) + T_H(z,y) \\ &\geq T_H(0,y) + T_H(e_1,y'). \end{aligned}$$

Rearranging and taking $y, y' \to \infty$ along their respective geodesics, we obtain $B^L_H(0,e_1) \leq B^R_H(0,e_1)$. □

Given (5.10) from the previous two lemmas, we can return to the proof of Theorem 5.21. We already argued for item 1 below (5.10). Coalescence of leftmost and rightmost geodesics implies item 2 of the theorem. For the last item, we use the argument of Theorem 5.15. This argument applies to general cases in which from each point x, there is an infinite geodesic Γ_x such that (a) for $x, y \in \mathbb{Z}^2$, Γ_x and Γ_y coalesce, and (b) there is a line L such that a.s., each Γ_x intersects this line finitely often. In such cases, the conclusion is that the tree formed by the union of the geodesics $(\Gamma_x : x \in \mathbb{Z}^2)$ cannot have infinite backward paths. In other words, a.s., the set $\{y : 0 \in \Gamma_y\}$ is finite. If there were a bigeodesic with one end directed in S_θ, then it would have to be contained in this tree of geodesics. However this tree cannot have an infinite backward path, and this would be a contradiction, so we conclude that a.s. there are no bigeodesics with one end directed in S_θ.

□

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/06

CHAPTER 6

Growth and competition models

6.1. Eden Model and the limit shape in high dimensions

The Eden model was introduced by Eden [**82**] in 1961 to address the following question in cell reproduction:

Starting from a single cell which may divide, its daughter cells divide again and again, what are the structural properties of the resulting colony of cells and how do various possible constraints effect the architecture?

We assume that each cell is identical to every other cell, that each cell is connected to at least one other cell and that each of their locations is specified by a node in a lattice. Eden considered the lattice to be two-dimensional, and reasoned that the close biological counterpart to this process was the growth of bacterial cells or tissue cultures that are constrained from moving. One example given is *Ulva latuca* [**45**], the common sea lettuce, that grows as a sheet of two cells thick and apparently grows only at its periphery. Although sometimes (wrongly) suggested as a possible model for cancer, it is important to point out that there is strong *in vitro* and *in vivo* evidence that Eden's model does not explain any possible tumor growth (see for instance [**41**] and the references therein).

The model can be described as follows. We start with a cell at the origin. Next, we adjoin one of the four neighbor vertices with equal probability. This two-celled configuration has six adjacent nodes. We attach one of these six neighbor vertices with equal probability 1/6. We continue this procedure indefinitely, each time adjoining a single cell uniformly from the collection of neighbor vertices. (See Figure 14.) The construction is similar in any dimension.

The Eden model can be seen as a time-change of a (site) FPP model in $\mathbb{Z}^d$ with exponential mean one passage times on the vertices. Indeed, let $\mathcal{A}(0) = \{0\}$ and, for $n \geq 1$, define $\mathcal{A}(n)$ as the subset of $\mathbb{Z}^d$ after we attach n cells (vertices) to $\mathcal{A}(0)$ according to the rules for the Eden model. In the site FPP model, define the times $\sigma_0 = 0$ and

$$\sigma_k = \inf\left\{t \geq 0 : B(t) \text{ contains } k+1 \text{ points of } \mathbb{Z}^d\right\},$$

where $B(t) := \{x \in \mathbb{Z}^d : T(0,x) \leq t\}$ and the vertex weights have distribution $F(t) = \mathbb{P}(\tau_e \leq t) = 1 - e^{-t}$. Then

$$\left\{\mathcal{A}(k) : k \geq 1\right\} \text{ and } \left\{B(\sigma_k) : k \geq 1\right\} \text{ have the same distribution.} \tag{6.1}$$

One of the most important consequences of (6.1) is that, contrary to the situation with other site-weight distributions, the evolution of the ball $B(t)$ when the weights are exponentially distributed is Markovian. Precisely, in the Eden model, the process

$$B(\sigma_1), B(\sigma_2), B(\sigma_3), \ldots$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

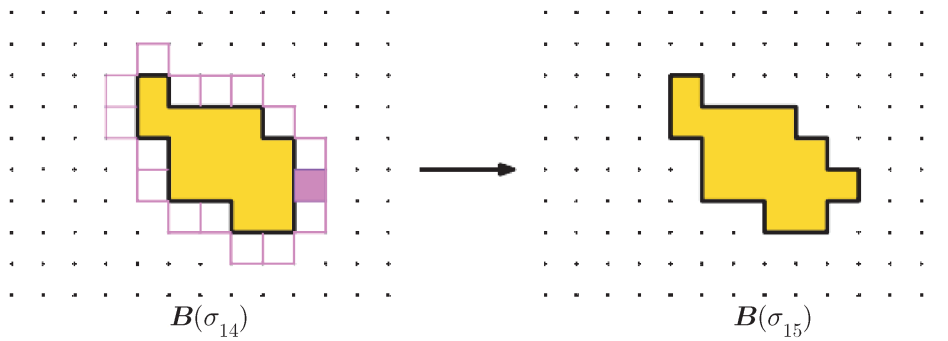


FIGURE 14. The evolution of the Eden model. On the left, the current cluster at time σ_{14} is surrounded by the bold path and the possible 16 neighbor squares that can be added are outlined adjacent to it. We choose one uniformly with probability 1/16. On the right, the cluster is shown after the attachment. Note that the order in which we attached previous boxes does not matter for the decision; only the knowledge of the boundary is needed to establish the next transition. The vertices in the picture are vertices of the dual lattice $\mathbb{Z}^2_*$ and a box represents a vertex of the lattice $\mathbb{Z}^2$.

is Markov on the space of possible shape configurations. Indeed, one need know only the (boundary) of the ball at time σ_i to know its possible transitions at time σ_{i+1}. The standard FPP model with exponential weights on the edges also satisfies the Markov property.

The Markov property of the growing ball allows use of tools that are not available for general weights. This has been explored in the variance bound of Pemantle-Peres [**150**] and the geodesic work of Häggström-Pemantle [**106**]. The reader may think that this powerful property would open doors to a better knowledge of the limit shape. As of today, however, there are few that successfully exploit the Markovian nature of $B(t)$ to obtain limit shape and geodesic results. The state of the art of the Eden model is essentially the same as for any FPP model. In Section 7.2.3, we discuss the last-passage analogue of Eden-type models, where through combinatorics and explicit formulae (achieved through the RSK correspondence), one indeed can say much more. We stress, however, that the methods of that section do not apply to FPP. Connections between FPP and growth processes are further explored in Section 6.2.

We end this section with the few theorems that explicitly use the Markovian nature of the Eden model. The first one is the determination of the growth of the time constant as the dimension diverges.

THEOREM 6.1 (Dhar [**77**]). *For exponential passage times, the quantity $\mu(e_1)$ as a function of d satisfies*

$$\lim_{d\to\infty} \mu(e_1)\frac{d}{\log d} = \frac{1}{2}.$$

The next result rules out certain shapes as possible limit shapes. However it was only proved in very large dimensions.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

THEOREM 6.2 (Kesten [**123**], Corollary 8.4). *If $d \geq 1000000$ then the limit shape of the Eden model is not an Euclidean ball.*

The bound $d \geq 1000000$ in Theorem 6.2 above was later improved used computer-aided proofs by Couronné-Enriquez-Gerin [**60**] and more recently by Bertrand-Pertinand [**33**].

Eden and Richardson observed by simulation that the limit shape appears to be circular in dimension two. However, one of the most long-standing questions remains:

QUESTION 6.1.1. Show that for $d \geq 2$, the limit shape of the Eden model is not a Euclidean ball.

We finish this section by reporting what is known for the limit shape in high dimensions when the edge-weights are not exponential. Assume $\mathbb{E}\tau_e < \infty$, and the existence of some constant $a \in [0, \infty]$ such that,

$$\left| \frac{\mathbb{P}(\tau_e \leq x)}{x} - a \right| \leq C \cdot |\log x|^{-1}, \tag{6.2}$$

for some $C > 0$ in some interval $[0, \epsilon_0]$, $\epsilon_0 > 0$. Here, we understand the case $a = \infty$ to mean

$$\lim_{x \to 0} \frac{\mathbb{P}(\tau_e \leq x)}{x} = \infty. \tag{6.3}$$

The asymptotic behavior of $\mu(e_1)$ as a function of d was determined in [**22**].

THEOREM 6.3 (Auffinger-Tang [**22**], Theorem 1.2). *Assume $\mathbb{E}\tau_e < \infty$ and that* (6.2) *holds for some $a \in [0, \infty]$. The quantity $\mu(e_1)$ as a function of d satisfies*

$$\lim_{d \to \infty} \frac{\mu(e_1) d}{\log d} = \frac{1}{2a}. \tag{6.4}$$

REMARK 6.4. Hypothesis (6.2) is a natural condition on the behavior of the distribution of the edge-weights at 0. It is satisfied by a large collection of examples; for instance, it includes all distributions that have a continuous density near the origin. The bound with $|\log x|^{-1}$ is a weaker condition than any polynomial bound around 0.

REMARK 6.5. If $F(0) > 0$, then (6.3) holds. In this case, Theorem 2.5 and the fact that $p_c = p_c(d)$ decreases to 0 as $d \to \infty$ together imply that for large enough d, one has $\mu(e_1) = 0$. Therefore, in this case (6.4) holds.

The theorem above also excludes the d-dimensional Euclidean ball $\mathsf{B}_2 := \{x \in \mathbb{R}^d : \|x\|_2 \leq \mu(e_1)^{-1}\}$, the d-dimensional cube $\mathsf{B}_\infty := \{x \in \mathbb{R}^d : \|x\|_\infty \leq \mu(e_1)^{-1}\}$, and the d-dimensional diamond $\mathsf{B}_1 := \{x \in \mathbb{R}^d : \|x\|_1 \leq \mu(e_1)^{-1}\}$ as possible limit shapes under the same assumptions. Note that due to convexity we have $\mathsf{B}_1 \subseteq \mathcal{B} \subseteq \mathsf{B}_\infty$.

THEOREM 6.6 (Auffinger-Tang [**22**], Theorem 1.5). *For any distribution satisfying* (2.10) *and $F(0) = 0$ for all $d \geq 2$, and* (6.2) *with $a \in (0, \infty)$, there exists $d_0 \geq 1$ such that for any $d \geq d_0$,*

$$\mathsf{B}_1 \subsetneq \mathcal{B} \subsetneq \mathsf{B}_\infty \text{ and } \mathcal{B} \neq \mathsf{B}_2.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

6.2. First-passage competition models

First-passage percolation is closely related to certain growth and competition models. In fact, the original version of the shape theorem was proved by Richardson [**155**] for a wide class of growth models including an example [**155**, Example 1] (the "Gp model") now known as the 1-type Richardson model. In this model on $\mathbb{Z}^d$, we suppose that the origin houses an infection at time $t = 0$ and all other sites are healthy. At each subsequent time $t = 1, 2, \ldots$, any healthy site with at least one infected neighbor becomes infected independently of all other sites with probability $p \in (0, 1)$. Richardson proved a shape theorem for the infected region at time t as $t \to \infty$ and believed, on the basis of computer simulations, that as p varies from 1 to 0, the limit shape varies from a "diamond to a disk." This model was shown to be equivalent to an FPP model with i.i.d. site-weights with a shifted geometric distribution [**81**]. It is also common to describe the infection as a species that is trying to colonize all sites of the space $\mathbb{Z}^d$.

We can build an infection/species model based on edge FPP in an analogous manner. Let (τ_e) be a realization of passage times. We infect the origin at time 0 and the infection spreads at unit speed across edges, taking time τ_e to cross the edge e. If the edge-weights are exponential, the memoryless property implies that the growth of the infected/colonized region is equal in distribution to a time change of a variant of the Eden process (discussed in the previous section). In this variant, a vertex is added to the current cluster with probability proportional to the number of edges connecting it to the cluster.

Building on the above definition, we may describe first-passage competition models, first introduced in $\mathbb{Z}^2$ by Häggström-Pemantle [**106**] with two competing infections (species). We will consider the same model in $\mathbb{Z}^d$ with $k \in \mathbb{N} \cup \{\infty\}$ different species. Fix k vertices $x_1, \ldots, x_k \in \mathbb{Z}^d$ and at time $t = 0$, infect x_i by an infection of type i. Each species spreads at unit speed as before, taking time τ_e to cross an edge e. An uninhabited site is exclusively and permanently colonized by the first species that reaches it; that is, $y \in \mathbb{Z}^d$ is occupied at time t by the i-th species if $T(y, x_i) \le t$ and $T(y, x_i) < \tau(y, x_j)$ for all $j \ne i$. If multiple x_i's attempt to colonize a site at the same time (an event that happens with probability zero if the passage time distribution is continuous), we use some deterministic rule to break ties.

Now, consider the set colonized by the i-th species starting at vertex x_i:

$$C(x_i) = \{y \in \mathbb{Z}^d \,:\, y \text{ is eventually occupied by species } i\}.$$

We are interested in describing the geometry of these random sets. In particular, much of the research so far has been driven by following questions:

(1) Can multiple species simultaneously succeed in invading an infinite subset of the lattice? In other words, does coexistence (also called mutual unbounded growth) occur?
(2) Assuming coexistence of at least two species, what can we say about the boundary interface? Does it have an asymptotic direction? A scaling limit?
(3) What happens if one species moves with faster speed than the others? Can we show that coexistence is impossible?

These questions will be discussed in the following sections.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

6.3. Competition with the same speed

We start with the first question on coexistence. One says that the distribution ν (with distribution function F) of the edge-weights admits coexistence of k species if for some choice of initial sites $x_1, \ldots, x_k$,

$$\mathbb{P}\bigg(|C(x_i)| = \infty \text{ for all } i = 1, \ldots, k\bigg) > 0. \tag{6.5}$$

The first point is that the choice of the initial sites is not important. One just needs to make sure that with positive probability, a new set of initial sites $y_1, y_2, \ldots, y_k$ are respectively infected by the initial sites $x_1, x_2, \ldots, x_k$. When the edge-weights are exponentially distributed, this can be easily done using the Markov property (see [**106**, Propositon 1.1], when $d = k = 2$). When the passage times are not exponentially distributed, this modification argument follows the same lines as those described in [**91**, Remark 2, p. 312]. One therefore has the following proposition:

PROPOSITION 6.7. *Assume that F is continuous and the support of ν is $[0, \infty)$. Suppose that $x_1, \ldots, x_k$ and $y_1, \ldots, y_k$ are points in $\mathbb{Z}^d$ such that there exist disjoint paths from x_i to y_i for $i = 1, \ldots, k$. Then*

$$\begin{aligned}\mathbb{P}\bigg(|C(x_i)| =&\infty \text{ for all } i = 1, \ldots, k\bigg) > 0 \\ &\Leftrightarrow \mathbb{P}\bigg(|C(y_i)| = \infty \text{ for all } i = 1, \ldots, k\bigg) > 0.\end{aligned}$$

The proposition above deals with choosing the initial sites. One has a similar result if starting with k different finite infected sets instead of single points [**73**]. However, the key and difficult step to prove coexistence is to establish that there indeed exist sources such that (6.5) occurs. Heuristically, the shape theorem gives the intuition that the larger the distance between the sources, the harder it is for one infection to surround the other one. If one hopes to prove that (6.5) holds, then one would like to choose the sources far from each other and hope that their infections will grow at least in different directions. This gives an indication that coexistence is strongly related to the existence of distinct geodesic rays with different asymptotic directions.

When ν is the exponential distribution, Häggström-Pemantle [**106**] proved coexistence of two species (see [**34**] for a review of results on Richardson models, focused on exponential passage times). Shortly thereafter, Garet-Marchand [**91**] and Hoffman [**111**] independently extended these results to prove coexistence of two species for a broad class of translation-invariant measures, including some non-i.i.d. ones. Later, Hoffman [**112**] demonstrated coexistence of four species in two dimensions for a similarly broad class of measures by establishing a relation with the number of sides of the limit shape in the associated FPP model. Recall that the number of sides of the limit shape $\mathcal{B}$ is equal to k if $\partial\mathcal{B}$ is a polygon with k sides. If $\partial\mathcal{B}$ is not a polygon, the number of sides is defined to be infinity.

THEOREM 6.8 (Hoffman [**112**], Theorem 1.6). *Suppose that $\mathcal{B}$ is bounded, $\mathbb{E}\tau_e^{2+\eta} < \infty$ for some $\eta > 0$, and that the edge-weight distribution is ergodic under lattice translations. If the number of sides of $\mathcal{B}$ is at least k then for any $\epsilon > 0$*

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

there exist distinct $x_1, \ldots, x_k$ *such that*

$$\mathbb{P}\Big(|C(x_i)| = \infty \text{ for all } i = 1, \ldots, k \Big) > 1 - \epsilon. \tag{6.6}$$

REMARK 6.9. Although the Theorem 6.8, as stated in [**112**], requires an additional assumption of unique geodesics, the proofs go through without it. J.-B. Gouéré has extended these methods to give a geometric description of infected regions; see [**100**].

We now sketch the proof of the above theorem.

STEP 1 - *Moving to Busemann functions.* We start the proof with geometric considerations. Let the set V consist of all point of differentiability of $\partial\mathcal{B}$ — those $v \in \partial\mathcal{B}$ such that there is a unique line L_v which is tangent to $\mathcal{B}$ through v. For such a v, let $w(v)$ be a unit vector parallel to L_v. Let $L_{n,v}$ be the line through nv in the direction of $w(v)$. One can show (see [**112**, Lemma 2.1]) that if the number of sides of $\partial\mathcal{B}$ is at least k then one can find points $v_1, \ldots, v_k \in V$ such that the lines L_{v_i} are distinct for all i. Fix such v_i's from now on. The first step is to establish the following lemma. Define the Busemann-type function $B_S(x, y)$ for $x, y \in \mathbb{Z}^2$ and $S \subset \mathbb{R}^2$ as

$$B_S(x, y) = T(x, S) - T(y, S).$$

Just as in Chapter 5, from the subadditivity of T and the definition of B_S, we have $B_S(x, y) \leq T(x, y)$ and

$$B_S(x, y) + B_S(y, z) = B_S(x, z). \tag{6.7}$$

Then one has the following.

LEMMA 6.10. *Assume that the number of sides of* $\mathcal{B}$ *is at least* k *and let* $v_i, 1 \leq i \leq k$ *be defined as above. If there exist* $\epsilon > 0$, *and* $x_1, \ldots, x_k$ *such that*

$$\mathbb{P}\Big(B_{L_{n,v_i}}(x_j, x_i) > 0 \ \forall i \neq j \Big) > 1 - \epsilon \tag{6.8}$$

for infinitely many n, *then* (6.6) *holds.*

PROOF. If for all $j \neq 1$

$$B_{L_{n,v_1}}(x_j, x_1) > 0,$$

then one can find $z_n \in L_{n,v_1}$ such that $T(z_n, x_1) < T(z_n, x_j)$ for all $j \neq 1$. Thus the species 1 will infect the site z_n on L_{n,v_1} before all other species will. Under the condition (6.8), one has $\mathbb{P}\left(B_{L_{n,v_1}}(x_j, x_1) \ \forall j \neq 1 \text{ occurs for infinitely many } n\right) > 1 - \epsilon$. So on this event, there is an infinite sequence of z_n's that are infected first by v_1, giving that $|C(x_1)| = \infty$. Repeating the same argument for the other species, we obtain (6.6). $\square$

STEP 2 - *The density argument.* Now, to end the proof of Theorem 6.8, one needs to construct points $x_i, 1 \leq i \leq k$ such that (6.8) holds. The idea is to choose $x_i = Mv_i$ with M large enough. Recall that the lower density of a subset S of $\mathbb{N}$ is

$$\liminf_{n\to\infty} \frac{1}{n} |S \cap \{1, \ldots, n\}|.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

LEMMA 6.11. *There exists $c > 0$ such that for all $\epsilon > 0$ there exists M such that the lower density of n with*

$$\mathbb{P}\Big(B_{L_{n,v_i}}(Mv_j, Mv_i) > cM \quad \forall i \neq j\Big) > 1 - \epsilon$$

is at least $1 - \epsilon$.

PROOF OUTLINE. The idea is that, as in Lemmas 5.5 and 5.11, $B_{L_{n,v_i}}(Mv_j, Mv_i)$ for large M behaves as $f(M(v_j - v_i))$ where f is a linear functional with null space containing $w(v_i)$ such that $f(v_i) = \mu(v_i)$. Therefore, for each pair (v_i, v_j) with $i \neq j$ we will find $w \in \mathbb{Z}^2$ and $c > 0$ (depending on the angle between the points v_i and v_j) such that for all large enough M the events:

$$B_{L_{n,v_i}}(Mv_j, Mw) > Mc(1 - \epsilon) \tag{6.9}$$

and

$$B_{L_{n,v_i}}(Mw, Mv_i) > -M\epsilon \tag{6.10}$$

occur with probability at least $1 - \epsilon$ in a set of values of n with lower density at least $1 - \epsilon$. Then (6.7) implies

$$B_{L_{n,v_i}}(Mv_j, Mv_i) > Mc(1 - 2\epsilon).$$

Establishing (6.9) and (6.10) was the central part of Hoffman's paper (see [**112**, Lemmas 4.4-4.5]) and Hoffman's averaging trick for (6.9), explained in the proof of Lemma 5.5, directly inspired the proof of the formula for the mean of the reconstructed Busemann function (see Lemma 5.11). The proof of (6.10) relies on the assumption that the limit shape is differentiable at v_i. □

Coexistence of infinitely many species is defined similarly and implies coexistence of k species for any k. In [**67**], Damron-Hochman extended the above theorem to include the case $k = \infty$ and used it to construct the first example of a passage time distribution that admits coexistence of infinitely many species. The current state of the art is given by a combination of Theorem 6.8 and Theorem 2.27. It establishes that infinite coexistence is not only possible but is necessary for all measures in $\mathcal{M}_p$.

THEOREM 6.12 (Auffinger-Damron [**19**], Theorem 2.2). *If $\nu \in \mathcal{M}_p$ for $p \in [\vec{p}_c, 1)$ then ν admits coexistence of infinitely many species.*

6.4. Competition with different speeds

If species have different growth rates, one may expect that non-coexistence occurs a.s.. We first consider the original case treated in [**107**], where the passage times are exponentially distributed, and we restrict ourselves to the case of two infections.

The model now has two parameters, denoted by λ_1 and λ_2, indicating the rates of the infections. We assume that $\lambda_1 < \lambda_2$, so species 1 spreads faster than species 2. The infection rule is the same as in the last section. We start with one infected site of each species. An uninfected site becomes type i infected at a rate proportional to the number of infected neighbors of type i and stays type i for all subsequent times.

The best result to date in this case is still from the original paper [**107**]. With one of the intensities fixed, coexistence has probability zero for all but at most

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

countably many values of the other intensity. It is however expected that this should be true for all values of $\lambda_1 \neq \lambda_2$.

QUESTION 6.4.1. Show that for all values of $\lambda_1 \neq \lambda_2$ and all $x_1 \neq x_2$,

$$\mathbb{P}\bigg(|C(x_1)| = |C(x_2)| = \infty\bigg) = 0.$$

It is worth mentioning that at this point, it is known that for almost every value of λ_2, coexistence does not occur, but we are unable to exhibit any single pair (λ_1, λ_2) for which coexistence fails. A possible way to answer this conjecture is to show:

QUESTION 6.4.2. Fix λ_1. Show that the probability of $\{|C(x_1)| = |C(x_2)| = \infty\}$ is monotone in λ_2.

Monotonicity is believed to hold on $\mathbb{Z}^d$, although there are examples of graphs where such a statement does not hold [**74**]. As in Proposition 6.7, the initial placement of the infections does not matter and one could start with any bounded initial sets. Certain special cases where one of the initial sets is unbounded (a half-space against a point, for instance) are treated in [**75**] and extended in [**16**].

As in Section 6.3, one can define competition with different speeds using more general passage times. This was studied by Garet-Marchand in [**93**], where they extended the results of [**106**] when the edge-weight distributions for the two species are stochastically comparable. Under certain conditions on the distributions, in this setting, they established in any dimension d that if the slow species survives, the fast species cannot occupy a very high density of space. For $d = 2$, they show that a.s., one species must finally occupy a set of full density in the plane while the other species occupies only a set of zero density.

6.5. The competition interface

Before ending the chapter, it is worth mentioning that in the past twenty years, there has also been a growing interest in questions related to the asymptotic shape of infected regions.

In the physics literature, Derrida-Dickman [**76**] describe simulations for the Eden model in which two clusters grow into a vacant sector of the plane with angle θ. (See Figure 15.) They obtain values for the roughening exponents of the competition interface depending on θ. Two species competition with mutation and selection was also studied numerically in the Eden model in a paper by Kuhr-Leisner-Frey [**135**] (see the references therein for other variations of the model).

However, precise results are scarce, except for solvable models of last-passage percolation, where there is a coupling between the competition interface and the second class particle in TASEP [**85, 86, 151**]. One of the main difficulties (again!) in the case of FPP is the lack of knowledge about the limit shape $\mathcal{B}$.

One way to define a competition interface is as follows. For a given $\theta \in (0, 2\pi)$, consider the sets

$$\text{Blue} = \left\{z \in \mathbb{Z}^2 : \frac{\theta}{2} < \arg z < \pi\right\}$$

and

$$\text{Red} = \left\{z \in \mathbb{Z}^2 : \pi < \arg z < 2\pi - \frac{\theta}{2}\right\}.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

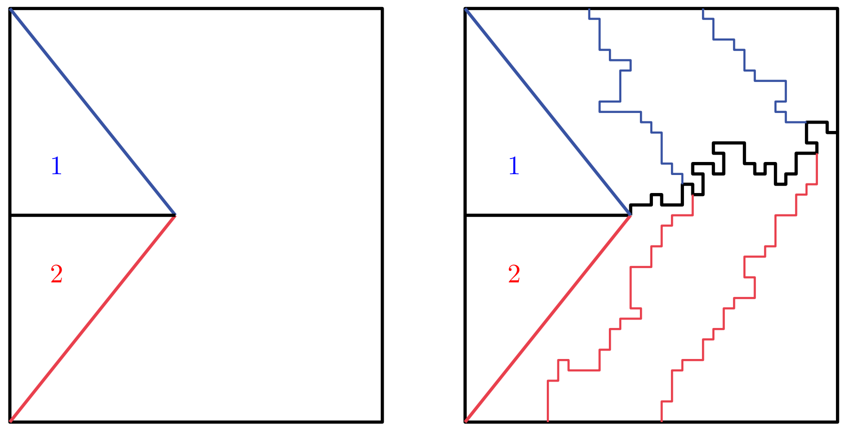


FIGURE 15. Competition interface between two species. On the left, the starting configuration, with initial angle $\theta > \pi$. Above, the region infected by species 1 while, below, the region infected by species 2. Emanating from the apex, the random curve defining the competition interface. The curves connecting the bottom of the square to the top delimit the boundary of the infected region at growing times.

Blue and Red are the initial infected sites. We now color a vertex v in $\mathbb{Z}^2 \setminus (\text{Blue} \cup \text{Red})$ blue if

$$T(v, \text{Blue}) < T(v, \text{Red}),$$

and red otherwise. If the passage time distribution is continuous, a.s. there will be no ties and a color will be assigned to each vertex of $\mathbb{Z}^2$. The competition interface is the unique path on the dual lattice that sits between the blue and red regions. Now for $v \in \mathbb{Z}^2 \setminus (\text{Blue} \cup \text{Red})$ define

$$p(v) = \mathbb{P}(v \text{ is blue})$$

and let the width of the interface at abscissa x be defined as

$$\omega(x) = \sum_{v \,:\, v \cdot e_1 = x} p(v)(1 - p(v)).$$

For x large, [**76**] predicted that the function $\omega(x)$ has power law behavior $\omega(x) \sim x^\zeta$ with exponent ζ satisfying

$$\zeta = \begin{cases} \frac{1}{3} & \text{if } \theta < \pi \\ \frac{2}{3} & \text{if } \theta = \pi \\ 1 & \text{if } \theta > \pi. \end{cases}$$

As far as we know, none of these questions have been investigated except in the case of last-passage percolation with exponential weights. Some results in this exactly solvable case were also obtained in [**86**] (see [**97**] for a study of more general distributions).

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/07

CHAPTER 7

Variants of FPP and related models

In this section, we discuss a sequence of questions on variants of FPP that were not addressed in the previous sections. Instead of giving proofs, we try to summarize the some of the most well-studied connections and variants; however, we refer to the papers for complete details.

7.1. The maximum flow

The problem of the maximum flow in a random network can be thought of as a higher-dimensional analogue of FPP, where optimization over paths is replaced by optimization over higher-dimensional hypersurfaces. We will mostly consider the model in the setting of $\mathbb{Z}^d$ as in the rest of this survey, though we will mention some results on other graphs. Consider the setting of FPP, where edges e are assigned i.i.d. nonnegative weights τ_e; now the interpretation of the weights is different, with τ_e representing "capacity" for flow. That is, τ_e sets the maximal number of units of fluid that can pass through the edge e per unit time.

Consider a connected, finite subset of vertices $\Upsilon \subseteq \mathbb{Z}^d$ and the corresponding induced directed subgraph of $\mathbb{Z}^d$ (denoted also by Υ, with some abuse of notation). That is, the set of edges of Υ is the set of ordered pairs (x, y) such that $x, y \in \Upsilon$ and x and y are endpoints of a standard undirected edge $\{x, y\}$ of the lattice $\mathbb{Z}^d$. We assign to each directed edge (x, y) the edge-weight $\tau_{\{x,y\}}$ of the corresponding undirected edge $\{x, y\}$. Let Λ_1 and Λ_2 be two distinguished and disjoint subsets of Υ, thought of respectively as a source and a sink of fluid. A flow g from Λ_1 to Λ_2 is an assignment of non-negative flow rates $g(e)$ and to all directed edges $e \in \Upsilon$, having the following properties:

(1) $0 \leq |g(e)| \leq \tau_e$ for all e;
(2) if $e = (x, y)$ and $\bar{e} = (y, x)$, then $g(e) = -g(\bar{e})$ for all e;
(3) for each vertex $v \in \Upsilon \setminus (\Lambda_1 \cup \Lambda_2)$, the total inflow equals the total outflow. That is,
$$\sum_w g((w, v)) = \sum_z g((v, z));$$
(4) The net flow out of Λ_1 is equal to the net flow into Λ_2; that is,

$$\sum_{u \in \Lambda_1} \sum_v g((u, v)) = \sum_{w \in \Lambda_2} \sum_z g((z, w)) \ . \tag{7.1}$$

We call the quantity on both sides of (7.1) the "strength" of the flow g, and we abbreviate it by $S(g)$. Because the edges incident to vertices of each Λ_i have finite capacities, there is a maximal flow, with strength

$$\Phi = \Phi(\Upsilon, \Lambda_1, \Lambda_2) = \sup_{\substack{g:\, g \text{ a flow from} \\ \Lambda_1 \text{ to } \Lambda_2 \text{ in } \Upsilon}} S(g) \ .$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

One is typically interested in the properties of Φ and the corresponding maximizing flow g in a limiting sense. For instance, one may consider the case that Υ is the cube $[0,n]^d$ and Λ_1, Λ_2 two opposite faces, and wish to study the maximal g in the limit of large n.

As remarked earlier, this class of problems is similar to an FPP problem; we now make this connection explicit. We say a set E of edges in Υ separates Λ_1 from Λ_2 if Λ_1 and Λ_2 are in different connected components of $\Upsilon \setminus E$. We call E a cut of (Λ_1, Λ_2) if E separates Λ_1 from Λ_2 and if no proper subset of E separates Λ_1 from Λ_2. Now the max-flow min-cut theorem (see [**38**]) states that

THEOREM 7.1. *The following equality holds:*

$$\Phi = \min \left\{ \sum_{e \in E} \tau_e : E \textit{ is a cut of } (\Lambda_1, \Lambda_2) \right\}.$$

The minimal cut formulation of Theorem 7.1 expresses the strength of the maximum flow in a form analogous to a first-passage time, as a minimum of sums of edge-weights over cuts. In the case that $d = 2$ and Υ is a square with Λ_1 and Λ_2 the top and bottom sides, cuts are self-avoiding paths which cross the square from left to right, and the analogy is exact. When $d = 3$, cuts are typically surfaces, so the maximum flow problem is an extension of FPP, replacing the optimization over paths to an optimization over surfaces.

Now, let $d = 3$ and write $\Upsilon = [0,k] \times [0,l] \times [0,m]$. Furthermore, we put $\Lambda_1 = [0,k] \times [0,l] \times \{0\}$ and $\Lambda_2 = [0,k] \times [0,l] \times \{m\}$. The following result is due to Kesten.

THEOREM 7.2 (Kesten [**124**], Theorem 2.12). *Assume that $m(k,l) \to \infty$ as $k \geq l \to \infty$ and that for some $\delta > 0$,*

$$k^{-1+\delta} \log m(k,l) \to 0. \tag{7.2}$$

Then there exists p^*, with $1/27 \leq p^* \leq p_c$ such that for any distribution F with $F(0) < p^*$ and finite exponential moments

$$\lim_{k,l,m \to \infty} \frac{1}{kl} \Phi(\Upsilon, \Lambda_1, \Lambda_2) = \heartsuit \qquad \textit{a.s. and in } L_1 \tag{7.3}$$

for some $\heartsuit > 0$.

In the same paper [**124**, Theorem 2.18], Kesten showed that if $F(0) > 1 - p_c$ (and some other technical conditions on the distribution, like finite sixth moment) $\heartsuit$ exists and is equal to zero. The critical case when $F(0) = 1 - p_c$ was also considered by Zhang [**183**].

THEOREM 7.3 (Zhang [**183**]). *If $F(0) = 1 - p_c$ and if in addition $\mathbb{E}\tau_e < \infty$ then for any $l > 0$,*

$$\lim_{k,m \to \infty} \frac{1}{kl} \Phi(\Upsilon, \Lambda_1, \Lambda_2) = 0$$

and

$$\lim_{k,l,m \to \infty} \frac{1}{kl} \Phi(\Upsilon, \Lambda_1, \Lambda_2) = 0,$$

where in the second limit, k, l, m go to infinity without any restriction (in particular without assuming (7.2)*).*

REMARK 7.4. The arguments in [**183**] are also valid for dimensions $d > 3$.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Assuming finite exponential moments, Zhang improved the above results and proved the following theorem, which shows existence of a limit for general dimensions. Here we use $\Upsilon = [0, k_1] \times \cdots \times [0, k_{d-1}] \times [0, m]$ and $\Lambda_1 = [0, k_1] \times \cdots \times [0, k_{d-1}] \times \{0\}$, $\Lambda_2 = [0, k_1] \times \cdots \times [0, k_{d-1}] \times \{m\}$.

THEOREM 7.5 (Theorem 3, Remark 4 [**186**], Zhang). *Let $d \geq 2$. Assume finite exponential moments and that $m = m(k_1, \ldots, k_{d-1}) \to \infty$ as $k_1, k_2, \ldots, k_{d-1} \to \infty$ in such a way that*

$$\exists\, \delta \in (0,1), \quad \log m(k) \leq \min_{i=1,\ldots,d-1} k_i^{1-\delta}.$$

Then for some $\heartsuit = \heartsuit(F)$, we have

$$\lim_{k_1,\ldots,k_{d-1},m\to\infty} \frac{1}{\prod_{i=1}^{d-1} k_i} \Phi(\Upsilon, \Lambda_1, \Lambda_2) = \heartsuit \qquad a.s. \text{ and in } L_1.$$

Moreover, this limit is positive if and only if $F(0) < 1 - p_c$.

After the work of Kesten and Zhang, remarkable progress has been made over a period of years. Flows between general subsets of $\mathbb{Z}^d$ were studied in the papers of Garet [**90**] and Rossignol-Théret [**158**,**159**] in dimension two. The study culminates in a series of papers of Cerf and Théret [**48**–**50**]. Under finite exponential moments for the passage times, they establish the analogue of Theorem 7.2 for very general networks and derive lower and upper large deviations (of surface order n^{d-1}, and volume order n^d, respectively).

7.2. Variants

7.2.1. Inhomogeneous or ergodic edge-weights. Over the years, some attention has been devoted to models of FPP with non-i.i.d edge-weights. Boivin [**36**] considered the case in which the edge-weight distribution is (stationary and) ergodic under lattice translations and proved an analogue of the shape theorem, Theorem 2.17. (See Björklund [**35**] for the most general known version.) In the special case where the τ_e's are also bounded, this was first noted by Derrienic (see [**126**, page 259]). A remarkable contribution came in the work of Häggström and Meester [**105**], who classified all possible limit shapes in the ergodic case:

THEOREM 7.6 (Häggström-Meester [**105**]). *Let $\mathcal{C}$ be the set of all subsets of $\mathbb{R}^d$ that are compact, convex, and symmetric (satisfy $x \in \mathcal{C} \Rightarrow -x \in \mathcal{C}$) with nonempty interior. Let $\mathcal{C}^*$ be the set of all compact subsets of $\mathbb{R}^d$ with nonempty interior that can arise as limiting shapes for stationary first-passage percolation. Then $\mathcal{C} = \mathcal{C}^*$.*

The proof of the above theorem is constructive. That is, given a set $B \in \mathcal{C}^*$, Häggström and Meester construct a stationary measure $\mathbb{P}$ on edge-weight configurations whose limit shape is B. The measure $\mathbb{P}$ has marginals with bounded support. Their construction never produces i.i.d. edge-weights. The following open question is a rephrasing of Question 2.1.1.

QUESTION 7.2.1. Find a convex, compact set B with non-empty interior and symmetry about the axes, not equal to the ℓ^1 unit ball, and an i.i.d. first-passage model that has B as its limit shape.

Other types of dependent passage times were considered in the literature. Fontes and Newman [**88**] and Chayes and Winfield [**57**] considered dependent edge-weights generated from random colorings of vertices of $\mathbb{Z}^d$. Van den Berg and

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

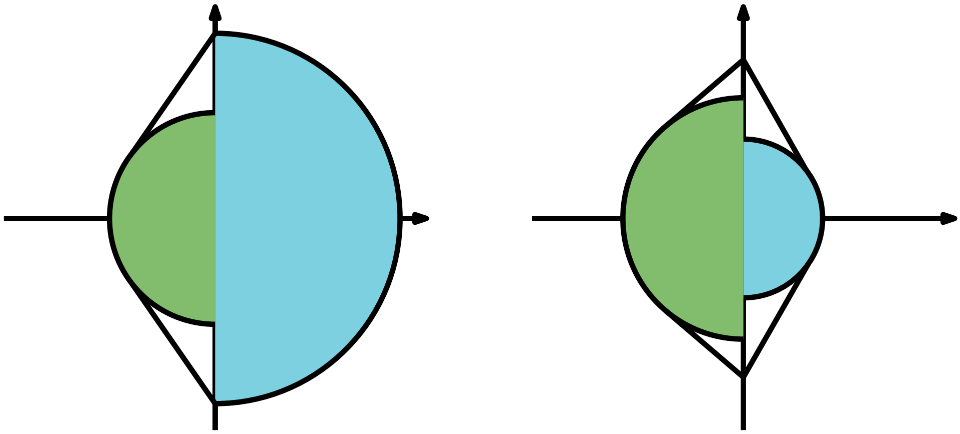


FIGURE 16. Representation of the limit shape in the left-right inhomogeneous model. On the left, the edge distribution F_+ is more variable than F_-. On the right, the case where no such relation is present.

Kiss [**32**] extended the BKS sublinear variance theorem (from Section 3.2.1) to a class of dependent weights that includes passage times that correspond to the minimal number of sign changes in a subcritical Ising landscape, a model investigated by Higuchi and Zhang [**110**].

The following inhomogeneous variation was also studied recently in [**4**]. Consider a version of FPP where edges in the left and right half-planes of the $\mathbb{Z}^2$ lattice are assigned weights independently according to distributions F_- and F_+. Complementing the classical approach of the proof of Theorem 2.17 with large deviation estimates for half-plane passage times as in [**2**], the authors proved a shape theorem. They also relate the limit shape in this version to the shapes arising in the standard homogeneous models corresponding to distributions F_+ and F_-. If either of F_- and F_+ is more variable than the other (see Definition 2.8), then the asymptotic shape is the convex hull of the restriction to the respective half-planes of the asymptotic shapes for F_- and F_+ from the homogeneous models. When no such relation is present, the asymptotic shape equals the convex hull of the two half-shapes and a potentially wider additional line segment along the vertical axis (see Figure 7.2.1). The question introduced in that work is:

QUESTION 7.2.2. Find necessary and sufficient conditions on F_- and F_+ so that the asymptotic speed in the vertical direction is strictly greater than that given by either F_- or F_+ in the standard homogeneous models.

7.2.2. FPP on different graphs. FPP on graphs different from $\mathbb{Z}^d$ has also appeared in recent literature. In addition to work on $\mathbb{Z}^d$ with long-range bonds [**53**], examples include complete graphs, the hypercube $\{0,1\}^n$, and also random graphs such as the Erdős-Rényi random graph and the configuration model. As most of these examples are finite graphs, the questions are different from those considered here.

Given a finite graph $G = (V, E)$ and weights $(\tau_e)_{e \in E}$, the passage time between two vertices u and v is defined analogously to (1.1) as the minimum passage time

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

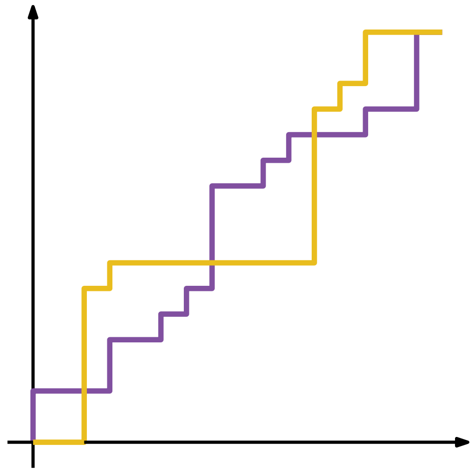


FIGURE 17. Examples of directed paths from the origin to (n, n). These paths are used in oriented first-passage and last-passage percolation models.

among paths between u and v. A few new and old quantities emerge as the main objects of study. First, one would like to understand the typical passage time and the number of edges in a geodesic between two vertices. That is, choosing U and V uniformly at random from the vertices of G, one studies the behavior of the passage time $T(U, V)$ and the length of a geodesic from U to V. Further objects of interest include the flooding time from a vertex u, given by $\max_{v \in V} T(u, v)$, and the the weighted diameter $\max_{u,v \in V} T(u, v)$.

For results in this direction we refer the reader to the notes of R. van der Hofstad [**172**, Chapter 8] and the references therein.

7.2.3. Last-passage percolation. In last-passage percolation (LPP), only oriented paths (those whose vertices have non-decreasing coordinates) are considered. Usually the weights are placed on the vertices instead of the edges and the passage time between two points u and v such that $u \leq v$ (that is, $u \cdot e_j \leq v \cdot e_j$ for all j) is given by

$$L(u, v) = \max_{\Gamma : u \to v} T(\Gamma).$$

Existence of time constants and the shape theorem have similar proofs to those for FPP in Section 2. A major difference is that the process is now superadditive and $\mathcal{B}$ is not convex (its complement in the first orthant is convex). The directness of paths, however, requires that vertices u, v, w be ordered in a non-decreasing fashion so that $L(u, w) \geq L(u, v) + L(v, w)$. The details for the proof of the shape theorem can be found in [**144**], including a necessary extra step to handle the continuity of $\mathcal{B}$ up to the boundary of the first orthant. Here, we will discuss the model in two dimensions and in this subsection, we will use the notation $\mu(x, y) = \mu(xe_1 + ye_2)$ for the time constant.

The importance of this small variation of FPP is that in dimension two there are natural correspondences between LPP models and certain queueing networks, systems of queues in tandem, and interacting particle systems. This connection reaches a deeper level, as precise scaling laws have been obtained for particular

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

distributions. If the passage times are exponentially distributed with mean one, then the time constant can be explicitly computed:

$$\mu(x,y) = (\sqrt{x} + \sqrt{y})^2,$$

as was first shown by Rost [**157**]. If ν is the probability measure of a geometric random variable with parameter p then [**118**]

$$\mu(x,y) = \frac{1}{p}(x + y + 2\sqrt{xy(1-p)}).$$

In both cases, finer asymptotics are available [**118**], as the distribution of

$$\frac{L(0, n(x,y)) - n\mu(x,y)}{n^{1/3}} \tag{7.4}$$

converges to a non-degenerate limit as n goes to infinity. The proof of (7.4) goes through the following special formula that identifies the law of the passage time with the law of the largest eigenvalue of the Laguerre unitary ensemble. Let A be an $n \times n$ matrix with entries that are i.i.d. complex Gaussian random variables with mean zero and variance $1/2$.

THEOREM 7.7 (Johansson [**118**], Proposition 1.4). *If ν is exponentially distributed with mean one then for all $t > 0$,*

$$\mathbb{P}(L(0,(n,n)) \le t) = \mathbb{P}(\lambda_n \le t), \tag{7.5}$$

where λ_n is the largest eigenvalue of the $n \times n$ matrix AA^*.

The identity (7.5) is derived by first considering independent geometric passage times with parameter p. In this case, one can condition on the sum of the weights in the box $[0, n-1] \times [0, n-1]$ (thought of as an $n \times n$ matrix) to show:

(1) The conditional law of the weights is uniform on the space $\mathcal{M}_n(s)$ of all possible non-negative integer-valued $n \times n$ matrices with fixed sum s.

Indeed, letting $A(i,j)$ be the passage time for the vertex (i,j), we have for the matrix $A = (A(i,j))$,

$$\mathbb{P}\Big(A = (a_{ij})\Big) = \prod_{1 \le i,j \le n} (1-p)^{a_{ij}} p = p^{n^2} (1-p)^{\sum_{ij} a_{ij}}.$$

2. The space $\mathcal{M}_n(s)$ is in one-to-one correspondence with the space of generalized permutations of length s on the set $\{1, \ldots, n\}$. The last-passage time from 0 to (n,n) is exactly the length of the longest non-decreasing subsequence in the associated permutation.

These two facts allowed Johansson to use the RSK correspondence [**130**]. It is possible to connect the probability of the event $\{L(0,(n,n)) \le t\}$ to the number of pairs of semi-standard Young tableaux with a fixed size t for the first row. Determining the distribution of the passage times is thus a combinatorial problem and it has an explicit expression for any finite n. By approximating an exponential random variable by a geometric random variable, one derives Theorem 7.7. We refer the reader to the readable Section 2 of [**118**] for the details.

The law of λ_n is well-known to be explicit and amenable to asymptotic analysis through the asymptotics of Laguerre orthogonal polynomials (see [**14**] and the references therein). In particular, (7.4) is a simple combination of the theorem above and the fact that [**118**]

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

$$Z_n := \frac{\lambda_n - 4n}{2^{4/3}n^{1/3}} \Rightarrow W_2 \sim F_2,$$

where

$$F_2(s) = \exp\Big(- \int_s^\infty (x-s)q(x)^2 dx \Big),$$

with q the solution of the Painlevé II differential equation

$$q''(x) = xq(x) + 2q^3(x)$$
$$q(x) \sim Ai(x) \text{ as } x \to +\infty,$$

and $Ai(x)$ denotes the Airy function. The F_2 distribution is called the Tracy-Widom distribution due to the influential paper [**169**]. In the case of geometric passage times, the asymptotic analysis was also carried out in [**118**] using the asymptotics of Meixner polynomials. Once properly rescaled, the passage times fluctuate according to F_2.

As mentioned before, the explicit formula for the limit shape and the connection with TASEP was also explored to understand the asymptotic behavior of infinite geodesics and the competition interface. The reader interested in this direction should look at the work of Ferrari, Martin and Pimentel [**85**] and the references therein.

For passage times that are not geometric or exponential, the results in LPP are at a similar stage as those in FPP. Some recent progress was obtained in the sequence of papers [**97–99**] where, using results from queueing theory, the authors manage to (1) derive variational formulas for the time constant, similar to those in Section 2.2, (2) obtain asymptotic results for infinite geodesics and Busemann functions under mild assumptions on the limit shape.

7.2.4. Rotationally invariant models. The anisotropy of the lattice $\mathbb{Z}^d$ contributes significantly to our lack of description of the limit shape $\mathcal{B}$. In a certain sense, when studying FPP, we are trying to compare the geometry of $\mathbb{R}^d$ — the space where the limit metric μ resides — to the geometry of the lattice. This is rather unnatural as $\mathbb{R}^d$ has a standard isotropic structure, given by the ℓ^2 metric, while $\mathbb{Z}^d$ does not. For instance, in $\mathbb{Z}^d$, someone moving away from the origin in the diagonal direction (towards a point $(n, n, \ldots, n)$) has a different experience than someone going towards a point of the form $(n, 0, \ldots, 0)$.

Other variants of FPP were introduced to exploit the rotational symmetry of $\mathbb{R}^d$. To our knowledge the first example came with the work of Vahidi-Asl and Wierman [**170, 171**], where they considered FPP on the Voronoi tessellation and Delaunay triangulation of the plane. Since this (random) graph is rotationally invariant they determine that the limit shape is a Euclidean ball. Some of the results of Sections 3 and 4 (for instance, existence of directed geodesics and concentration estimates) were adapted by Pimentel [**151**] in this model. We refer the readers to these articles for definitions.

Seven years after the papers of Vahidi-Asl and Wierman, Howard and Newman [**114**] introduced another rotationally invariant model called Euclidean first-passage percolation. It is defined as follows. For $r = (q_1, \ldots, q_k)$ a finite sequence of points

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

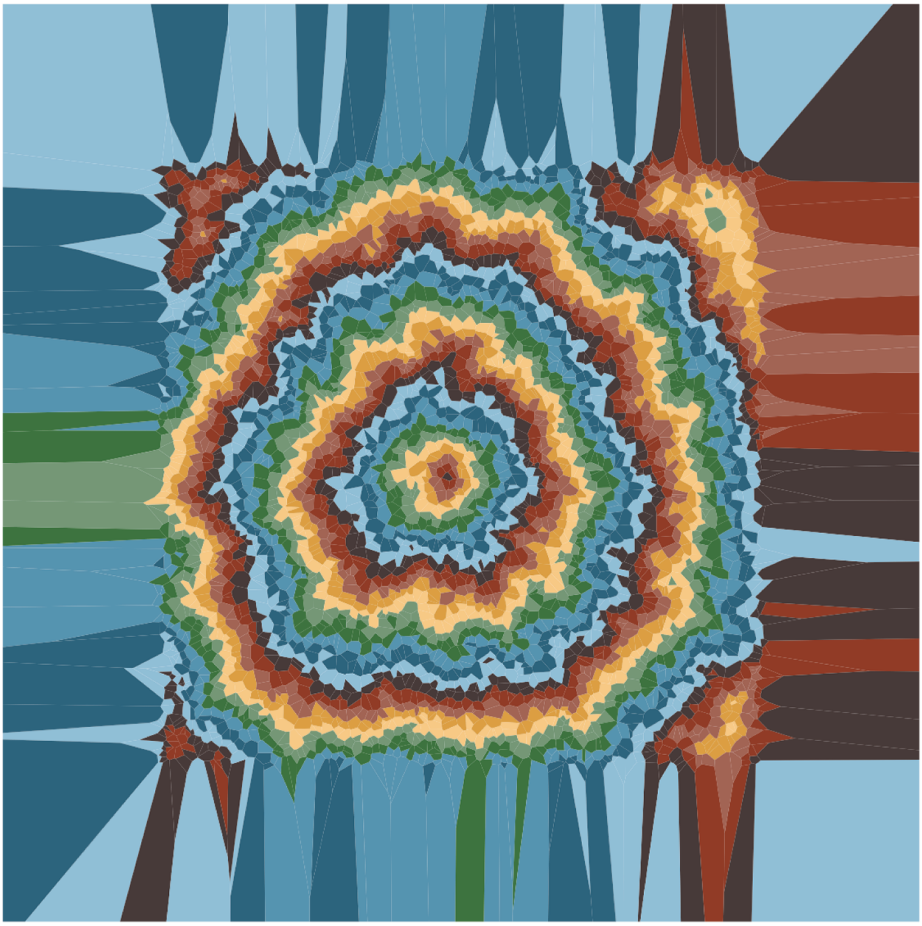


FIGURE 18. Simulation of an FPP model on a Voronoi tesselation with 5000 points chosen uniformly at random in the unit square. Shades represent the distance to the origin.

in $\mathbb{R}^d$ and for $\alpha > 1$, set

$$T(r) = \left(\sum_{j=1}^{k-1} |q_j - q_{j+1}|^\alpha \right)^{1/\alpha}.$$

Let Q be a homogeneous Poisson point process of unit density on all of $\mathbb{R}^d$. For $q, q' \in Q$ we set $T(q, q') = \inf T(r)$ where the infimum runs over all finite sequences $r = (q_1, \ldots, q_k)$ of points of Q such that $q_1 = q$ and $q_k = q'$. T has similar properties to the passage time in lattice FPP and the model also satisfies a shape theorem. Because of the statistical Euclidean invariance of the homogeneous Poisson point process, the asymptotic shape is exactly a Euclidean ball. The condition $\alpha > 1$ is imposed because if $0 < \alpha \leq 1$, then the straight line segment connecting any two Poisson points is a minimizing path for T, and the analysis becomes trivial.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

As a consequence of the differentiability and strict convexity of the limit shape in Euclidean FPP, Howard and Newman established that for $d \geq 2$ and $1 < \alpha < \infty$ a.s. geodesic lines other than other than those with opposite ends in antipodal directions do not exist. In addition, for $d = 2$, $2 \leq \alpha < \infty$, and any deterministically chosen direction θ, they proved that a.s. geodesic lines with opposite ends in directions θ and $-\theta$ do not exist. This result is the state of the art in the direction of verifying the conjecture that a.s. geodesic lines do not exist. However, even for $d = 2$ and $2 \leq \alpha < \infty$, it does not prove the bigeodesic conjecture since it leaves open the possible existence of geodesic lines with (random) angle dependent on the realization of Q. Results about exponents similar to those presented in Section 3 were also obtained in this model without the uniform curvature assumption. The reader is directed to the papers [**113**,**114**] for work on Euclidean FPP and [**23**,**136**] for other more recent rotationally invariant models.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

10.1090/ulect/068/08

CHAPTER 8

Summary of open questions

Here, we list all questions that appear in the book. The list is not exhaustive, but we believe that it provides a good panorama of the open and important areas of the field. For each question, see the referenced page and corresponding section for relevant notation. Problems marked with asterisks represent major open questions.

QUESTION 1 (p. 8). Find a nontrivial distribution for which one can explicitly determine $\mu(e_1)$.

QUESTION 2 (p. 10). Suppose that the support of the distribution of τ_e is unbounded. Let

$$X_n = \max\{\tau_e : e \text{ is in a geodesic from 0 to } ne_1\}.$$

How does X_n scale with n?

QUESTION 3 (p. 10). Extend Theorem 2.12 to the case $d > 2$, $I > 0$ and $\vec{p}_c \leq F(I)$.

QUESTION 4 (p. 17). [**] Show that if F is a continuous distribution then the limit shape is strictly convex.

QUESTION 5 (p. 17). Show that the d-dimensional cube (ℓ^∞ ball) is not a possible limit shape for an FPP model with i.i.d. passage times.

QUESTION 6 (p. 19). Show that for any measure $\nu \in \mathcal{M}_p$ the boundary of the limit shape is not flat outside the percolation cone.

QUESTION 7 (p. 19). Show that if the limit shape of a measure ν has a flat piece then $\nu \in \mathcal{M}_p$, for $p > p_c$ and the flat piece is delimited by the percolation cone.

QUESTION 8 (p. 19). Show that for any measure ν, in direction e_1 the boundary of the limit shape does not contain any segment parallel to the e_2-axis.

QUESTION 9 (p. 21). Find conditions that guarantee the existence of the limit (2.18).

QUESTION 10 (p. 26). [**] Show that for continuously distributed passage times, the limit shape is uniformly curved.

QUESTION 11 (p. 27). Assume $F(0) = 0$. Show that $\mu \geq \mathbb{E}_4\tau_e$.

QUESTION 12 (p. 32). [**] Show that for any $d \geq 2$, under suitable conditions on F, $\chi < 1/2$. If $d = 2$, show that $\chi = 1/3$.

QUESTION 13 (p. 32). Determine whether or not

$$\lim_{d\to\infty} \chi(d) = 0.$$

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

QUESTION 14 (p. 32). [$**$] For suitable d, show that $\chi > 0$.

QUESTION 15 (p. 33). Show that for continuous F and $d \geq 2$,

$$\lim_{n\to\infty} \mathbb{P}((n/2)e_1 \in \overline{\textsc{Geo}}(0, ne_1)) = 0.$$

QUESTION 16 (p. 43). Consider the two-dimensional torus of side length n: the graph $[0, n]^2 \cap \mathbb{Z}^2$ with opposite sides identified, and let T_n be the minimal passage time of all paths that wind around the torus once in the e_1-direction. Show that for some $C > 0$ and all $n \geq 1$,

$$\operatorname{Var} T_n \geq C \log n.$$

QUESTION 17 (p. 52). Show that $\underline{\chi}_2 = \overline{\chi}_{2+\delta}$ for some $\delta > 0$.

QUESTION 18 (p. 67). Assume that the passage times are bounded and not concentrated at a single point. Show that there exists a convex function $I_u(\epsilon)$ such that the following limit holds:

$$\lim_{n\to\infty} \frac{1}{n^d} \log p_n^u(\epsilon) = I_u(\epsilon).$$

QUESTION 19 (p. 71). Let $d = 3$ and $\mathbb{P}(\tau_e = 0) = p_c = 1 - \mathbb{P}(\tau_e = 1)$. Is it true that

$$\mathbb{E}T(0, \partial B(n)) \asymp \log n?$$

QUESTION 20 (p. 72). Is (3.65) optimal? More explicitly, let

$$\gamma_F(d) := \sup \left\{ \alpha : (3.66) \text{ holds for } h_n = n^\alpha \right\}.$$

What is the value of $\gamma_F(d)$?

QUESTION 21 (p. 74). Prove that under no assumptions on F, geodesics a.s. exist in the d-dimensional lattice, $d \geq 3$.

QUESTION 22 (p. 75). Determine necessary and sufficient conditions on F for dimension $d \geq 3$ so that $\rho = \infty$.

QUESTION 23 (p. 81). Do there exist constants C, M such that

$$\mathbb{P}(\overline{\textsc{Geo}}(0, x) \geq M\|x\|_\infty) \leq e^{-C\|x\|_\infty}?$$

Similarly, can one get upper bounds for $\mathbb{E}\overline{\textsc{Geo}}(0, x)^\alpha$ of the order $\|x\|_\infty^\alpha$ for positive $\alpha \neq 1$?

QUESTION 24 (p. 82). [$**$] Show that for continuous F, one has for fixed nonzero $x \in \mathbb{Z}^d$

$$\lim_{n\to\infty} \frac{|\textsc{Geo}(0, nx)|}{n} \quad \text{exists a.s.}$$

QUESTION 25 (p. 84). [$**$] Show that for continuous F, one has

$$\lim_{n\to\infty} \frac{\mathbb{E}D(0, ne_1)}{n} = 0.$$

QUESTION 26 (p. 93). [$**$] Find a distribution of passage times for which (4.29) holds with $d \geq 2$.

QUESTION 27 (p. 95). [$**$] Show that for continuous distributions and $d \geq 2$, there are infinitely many distinct geodesic rays.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

QUESTION 28 (p. 95). Find a distribution that is not in $\mathcal{M}_p$ that has an infinite number of distinct geodesic rays.

QUESTION 29 (p. 96). Is it true that the sequence of geodesics $(\textsc{Geo}(0, ne_1))$ converges to some Γ?

QUESTION 30 (p. 98). Show that $D = [0, 2\pi)$.

QUESTION 31 (p. 98). Show that $0 \in D$.

QUESTION 32 (p. 100). Decide whether the claim of Theorem 4.28 holds in arbitrary dimension.

QUESTION 33 (p. 100). Let $\Gamma(v)$ be a geodesic from 0 to v. Consider the collection of all edges that belong to $\Gamma(v)$ for infinitely many v and let $f(r)$ be the number of such edges which intersect the circle $x^2 + y^2 = r^2$. Does $f(r) \to \infty$ as $r \to \infty$; and, if so, how fast?

QUESTION 34 (p. 101). [$**$] Do geodesic lines exist?

QUESTION 35 (p. 112). Prove that under some conditions on the edge weights the limit (5.4) exists.

QUESTION 36 (p. 113). Show that under general assumptions on the edge-weights (τ_e) and for some $d \geq 2$, there a.s. exists an infinite geodesic with vertices $0 = x_0, x_1, \ldots$ that has an asymptotic direction; that is, the sequence $(x_n/\mu(x_n))$ converges.

QUESTION 37 (p. 133). Show that for $d \geq 2$, the limit shape of the Eden model is not a Euclidean ball.

QUESTION 38 (p. 138). Show that for all values of $\lambda_1 \neq \lambda_2$ and all $x_1 \neq x_2$,

$$\mathbb{P}\Big(|C(x_1)| = |C(x_2)| = \infty\Big) = 0.$$

QUESTION 39 (p. 138). Fix λ_1. Show that the probability of $\{|C(x_1)| = |C(x_2)| = \infty\}$ is monotone in λ_2.

QUESTION 40 (p. 143). Find a convex, compact set B with non-empty interior and symmetry about the axes, not equal to the ℓ^1 unit ball, and an i.i.d. first-passage model that has B as its limit shape.

QUESTION 41 (p. 144). Find necessary and sufficient conditions on F_- and F_+ so that the asymptotic speed in the vertical direction is strictly greater than that given by either F_- or F_+ in the standard homogeneous models.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Bibliography

[1] Ahlberg, D. (2015). Asymptotics of first-passage percolation on one-dimensional graphs. *Adv. in Appl. Probab.*, **47**, 182–209.

[2] Ahlberg, D. (2015). Convergence towards an asymptotic shape in first-passage percolation on cone-like subgraphs of the integer lattice. *J. Theoret. Probab.*, **28**, 198–222.

[3] Ahlberg, D. (2015). A Hsu-Robbins-Erdős strong law in first-passage percolation. *Ann. Probab.*, **43**, 1992–2025.

[4] Ahlberg, D., Damron, M. and Sidoravicius, V. (2016). Inhomogeneous first-passage percolation. *Electron. J. Probab.*, **21**, Paper no. 4, 19 pp.

[5] Ahlberg, D. and Hoffman, C. (2016). Random coalescing geodesics in first-passage percolation. *arXiv:1609.02447.*

[6] Aizenman, M. (1997). On the number of incipient spanning clusters. *Nuclear Phys. B*, **485**, 551–582.

[7] Alexander, K. (1997). Approximation of subadditive functions and convergence rates in limiting-shape results. *Ann. Probab.*, **25**, 30–55.

[8] Alexander, K. (1993). A note on some rates of convergence in first-passage percolation. *Ann. Appl. Probab.*, **3**, no. 1, 81–90.

[9] Alexander, K. (2011). Subgaussian rates of convergence of means in directed first passage percolation. *arXiv: 1101.1549.*

[10] Alexander, K. and Zygouras, N. (2013). Subgaussian concentration and rates of convergence in directed polymers. *Electron. J. Probab.*, **18**, 1–28.

[11] Alm, S. E. (1998). A note on a problem by Welsh in first-passage percolation. *Combin. Probab. Comput.*, **7** (1), 11–15.

[12] Alm, S. E. and Deijfen, M. (2015). First passage percolation on $\mathbb{Z}^2$ - a simulation study. *J. Stat. Phys.*, **161**, 657–678.

[13] Alm, S. E. and Wierman, J. C. (1999). Inequalities for means of restricted first-passage times in percolation theory. *Combin. Probab. Comput.*, **8**, 307–315.

[14] Anderson, G., Guionnet, A. and Zeitouni, O. (2010) *An introduction to random matrices.* Cambridge studies in advanced mathematics, **118**.

[15] Antal, P. and Pisztora, A. (1996). On the chemical distance for supercritical Bernoulli percolation. *Ann. Probab.*, **24**, 1036–1048.

[16] Antunović, T. and Procaccia, E. (2017). Stationary Eden model on Cayley graphs. *Ann. Appl. Probab.*, **27**, 517–549.

[17] Auffinger, A. and Damron, M. (2013). The universal scaling relation for polymers in a random environment and related models *ALEA, Lat. Am. J. Probab. Math. Stat.*, **10** (2), 857–880.

[18] Auffinger, A. and Damron, M. (2014). A simplified proof of the relation between scaling exponents in first-passage percolation. *Ann. Probab.*, **42**, No. 3, 1197–1211.

[19] Auffinger, A. and Damron, M. (2013). Differentiability at the edge of the limit shape and related results in first passage percoaltion. *Probab. Theory Related Fields*, **156**, 193–227.

[20] Auffinger, A., Damron, M. and Hanson, J. (2015). Limiting geodesics for first-passage percolation on subsets of $\mathbb{Z}^2$, *Ann. Appl. Probab.*, **25**, 373–405.

[21] Auffinger, A., Damron, M. and Hanson, J. (2014). Rate of convergence of the mean for sub-additive ergodic sequences. *Adv. Math.*, **285**, 138–181.

[22] Auffinger, A. and Tang, S. (2016). On the time constant of high dimensional first passage percolation. *Electron. J. Probab.*, **21**, 23 pp.

[23] Bakhtin, Y. and Wu, W. (2016). Transversal fluctuations for a first passage percolation model. *arXiv: 1605.05965.*

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

[24] Basdevant, A. Enriquez, N., Gerin, L., Gouéré, J.-B. (2014). The shape of large balls in highly supercritical percolation. *Electron. J. Probab.*, **19**, 1–14.

[25] Benaïm, M. and Rossignol, R. (2008). Exponential concentration for first passage percolation through modified Poincaré inequalities. *: Ann. Inst. H. Poincaré Probab. Statist.*, **44**, 544–573.

[26] Benjamini, I. Kalai, G. and Schramm, O. (2003). First passage percolation has sublinear distance variance. *Ann. Probab.*, **31**, 1970–1978.

[27] Benjamini, I. and Tessera, R. (2015). First passage percolation on nilpotent Cayley graphs and beyond. *Electron. J. Probab.*, **20**, no. 9, 20 pp.

[28] van den Berg, J. (1983). A counterexample to a conjecture of J. M. Hammersley and D. J. A. Welsh concerning first-passage percolation. *Adv. in Appl. Probab.*, **15**, no. 2, 465–467.

[29] van den Berg, J. and Fiebig, U. (1987). On a combinatorial conjecture concerning disjoint occurrences of events. *Ann. Probab.*, **15**, 354–375.

[30] van den Berg, J. and Kesten, H. (1985). Inequalities with applications to percolation and reliability. *J. Appl. Probab.*, **22**, 556–569.

[31] van den Berg, J. and Kesten, H. (1993). Inequalities for the time constant in first-passage percolation. *Ann. Appl. Probab.*, **3**, 56–80.

[32] van den Berg, J. and Kiss, D. (2012). Sublinearity of the travel-time variance for dependent first-passage percolation. *Annals of Probab.*, **40**, 743–764.

[33] Bertrand, Q. and Pertinand, J. (2017). Dimension improvement in Dhar's refutation of the Eden conjecture. *arXiv:1705.08143*.

[34] Blair-Stahn, N. D. (2010). First passage percolation and competition models. *arXiv:1005.0649*.

[35] Bjorklund, M. (2010). The asymptotic shape theorem for generalized first passage percolation. *Ann. Probab.*, **38**, 632–660.

[36] Boivin, D. (1990). First passage percolation: the stationary case. *Probab. Theory Related Fields*, **86**, 491–499.

[37] Boivin, D. and Derrien, J.-M. (2002). Geodesics and recurrent of random walks in disordered systems. *Ann. Probab.*, **7**, 101–115.

[38] Bollobás, B. (1979). *Graph theory.* Graduate Texts in Mathematics. Springer- Verlag, New York. An introductory course.

[39] Boucheron, S, Lugosi, G., and Massart, P. *Concentration inequalities. A nonasymptotic theory of independence.* With a foreword by Michel Ledoux. Oxford University Press, Oxford, 2013. x+481 pp.

[40] Bridson, M. and Haefliger, A. *Metric Spaces of Non-Positive Curvature*, Grundlder Math. Wiss. 319, Springer Verlag, 1999.

[41] Bru, A., Albertos, S. Subiza, J., Garcia, J., and Bru, I. (2003). The Universal Dynamics of Tumor Growth. *Biophys J.*, **85**, 2948–2961.

[42] Busemann, H. *The geometry of geodesics.* Vol. 6. Dover Publications, 1985.

[43] Burago, D., Burago, Y. and Ivanov, S. *A Course in Metric Geometry.* American Mathematical Society, 2001.

[44] Burkholder, D. L. (1973). Discussion of Prof. Kingman's paper. *Ann. Probab.*, **1**, 900–902.

[45] Burrows, E. M. (1991). Seaweeds of the British Isles. *London: Natural History Museum.* ISBN 0-565-00981-8.

[46] Burton, R. M. and Keane, M. (1989). Density and uniqueness in percolation. *Comm. Math. Phys.*, **121**, Number 3, 501–505.

[47] Cerf, R. and Théret, M. (2014). Weak shape theorem in first passage percolation with infinite passage times. *Ann. Inst. Henri Poincaré Probab. Stat.*, **52**, no. 3, 1351–1381.

[48] Cerf, R. and Théret, M. (2011). Law of the large numbers for the maximal flow through a domain of $\mathbb{R}^d$ in first passage percolation. *Trans. Amer. Math. Soc.*, **363**, 3665–3702.

[49] Cerf, R. and Théret, M. (2011). Lower large deviations for the maximal flow through a domain of $\mathbb{R}^d$ in first passage percolation. *Probab. Theory Related Fields*, **150** no. 3, 635–661.

[50] Cerf, R. and Théret, M. (2011). Upper large deviations for the maximal flow through a domain of $\mathbb{R}^d$ in first passage percolation. *Ann. Appl. Probab.*, **21**, No. 6, 2075–2108.

[51] Chatterjee, S. (2013). The universal relation between scaling exponents in first-passage percolation. *Ann. Math. (2)*, **177**, no. 2, 663–697.

[52] Chatterjee, S. and Dey, P. (2009). Central limit theorem for first-passage percolation time across thin cylinders. *Probab. Theory Related Fields*, **156**, no. 3–4, 613–663.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

[53] Chattjerjee, S. and Dey, P. (2015). Multiple phase transitions in long-range first-passage percolation on square lattices. *Comm. Pure Appl. Math.*, **69**, No. 2, 203–256.

[54] Chayes, L. (1991). On the critical behavior of the first passage time in $d \geq 3$. *Helvetica Phys. Acta*, **67**, 1055–1071.

[55] Chayes, J.T., Chayes, L. and Durrett, R. (1986). Critical behavior of the two-dimensional first passage time. *J. Stat. Phys.*, **45**, 933–951.

[56] Chayes, J. T., Chayes, L. and Durrett, R. (1987). Inhomogeneous percolation problems and incipient infinite clusters. *J. Phys. A*, **20**, 1521–1530.

[57] Chayes, L. and Winfield, C. (1993). The density of interfaces: A new first passage problem. *J. Appl. Prob.*, **30**, 851–862.

[58] Coupier, D. (2011). Multiple geodesics with the same direction. *Electron. Commun. Probab.*, **16**, 517–527.

[59] Chow, Y. and Zhang, Y. (2003). Large deviations in first-passage percolation. *Ann. Appl. Probab.*, **13**, 1601–1614.

[60] Couronné, O., Enriquez, N. and Gerin, L. (2011). Construction of a short path in high-dimensional first passage percolation. *Electron. Commun. Probab.*, **16**, 22–28.

[61] Cox, J. T. (1980). The time constant of first-passage percolation on the square lattice. *Adv. Appl. Prob.*, **12**, 864–879.

[62] Cox, J. T. and Durrett, R. (1981). Some limit theorems for percolation with necessary and sufficient conditions. *Ann. Probab.*, **9**, 583–603.

[63] Cox, J. T. and Kesten, H. (1981). On the continuity of the time constant of first-passage percolation. *J. Applied Prob.*, **18**, 809–819.

[64] Cranston, M. Gauthier, D. and Mountford, T. S. (2009). On large deviation regimes for random media models. *Ann. Appl. Probab.*, **19**, 826–862.

[65] Damron, M., Hanson, J. and Sosoe, P. (2016). Subdiffusive concentration in first-passage percolation. *Electron. J. Probab.*, **19**, 27 pp.

[66] Damron, M., Hanson, J. and Sosoe, P. (2015). Sublinear variance in first-passage percolation for general distributions. *Probab. Theory Related Fields*, **163**, 223–258.

[67] Damron, M. and Hochman, M. (2013). Examples of non-polygonal limit shapes in i.i.d. first-passage percolation and infinite coexistence in spatial growth models. *Ann. Appl. Probab.*, **23**, No. 3, 1074–1085.

[68] Damron, M. and Hanson, J. (2014). Busemann functions and infinite geodesics in two-dimensional first-passage percolation. *Comm. Math. Phys.*, **325**, 917–963.

[69] Damron, M. and Hanson, J. (2017). Bigeodesics in first-passage percolation. *Comm. Math. Phys.*, **349**, 753–776.

[70] Damron, M. and Kubota, N. (2016). Rate of convergence in first-passage percolation under low moments. *Stochastic Process. Appl.*, **126**, no. 10, 3065–3076.

[71] Damron, M., Lam, W.-K. and Wang, X. (2015). Asymptotics for $2D$ Critical First Passage Percolation. *arXiv:1505.07544*.

[72] Damron, M. and Tang, P. (2016). Superlinearity of geodesic length in $2D$ critical first-passage percolation. *arXiv:1610.02593*.

[73] Deijfen, M. and Häggström, O. (2006). The initial configuration is irrelevant for the possibility of mutual unbounded growth in the two-type Richardson model. *Combin. Probab. Comput.*, **15**, 345–353.

[74] Deijfen, M. and Häggström, O. (2006). Nonmonotonic coexistence regions for the two- type Richardson model on graphs. *Electron. J. Probab.*, **11**, 331–344.

[75] Deijfen, M. and Häggström, O. (2007). The two-type Richardson model with unbounded initial configurations. *Ann. Appl. Probab.*, **17**, 1639–1656.

[76] Derrida, B. and Dickman, R. (1991). On the interface between two growing Eden clusters. *Journal of Physics A: Mathematical and General*, **24**, 1385–1422.

[77] Dhar, D. (1986). Asymptotic shape of Eden clusters, *On growth and form. Ed. H. E. Stanley and N. Ostrowsky, Martinus Nijhoff.*, 288–292.

[78] Duminil-Copin, H. and Tassion, V. (2015). A new proof of the sharpness of the phase transition for Bernoulli percolation on $\mathbb{Z}^d$. *arXiv:1502.03051*

[79] Durrett, R. (2005). *Probability: theory and examples.* Duxbury Advanced Series. Third Edition.

[80] Durrett, R. (1984). Oriented percolation in two dimensions. *Ann. Probab.*, **12**, 999–1040.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

[81] Durrett, R. and Liggett, T. (1981). The shape of the limit set in Richardson's growth model. *Ann. Probab.*, **9**, 186–193.

[82] Eden, M. A two-dimensional growth process. 1961 *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. IV*, 223–239, *Univ. California Press, Berkeley, Calif.*

[83] Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *Ann. Statist.*, **9**, 586–596.

[84] Falik, D. and Samorodnitsky, A. (2007). Edge-isoperimetric inequalities and influences. *Combin. Probab. Comput.*, **16**, 693–712.

[85] Ferrari, P. A., Martin, J. B, and Pimentel, L. (2009). A phase transition for competition interfaces. *Ann. Appl. Probab.*, **19**, 281–317.

[86] Ferrari, P. A. and Pimentel, L. (2005). Competition interfaces and second class particles. *Ann. Probab.*, **33**, 1235–1254.

[87] Fishburn, P. C. and Shepp, L. A. (1991). On the FKB conjecture for disjoint intersections. *Discrete Math.*, **98**, 105–122.

[88] Fontes, L. and Newman, C.M. (1993). First passage percolation for random colorings of $\mathbb{Z}^d$. *Ann. Appl. Probab.*, **3**, 746–762.

[89] Forgacs, G., Lipowsky R. and Nieuwenhuizen, T. (1991). The behavior of interfaces in ordered and disordered systems. In *Phase transitions and critical phenomena* (C. Domb and J. Lebowitz, eds.) **14** 135–363. Academic, London.

[90] Garet, O. (2009). Capacitive Flows on a 2D random net. *Ann. Appl. Probab.*, **19**, 641–660.

[91] Garet, O. and Marchand, R. (2005). Coexistence in two-type first-passage percolation models. *Ann. Appl. Probab.*, **15**(1A), 298–330.

[92] Garet, O. and Marchand, R. (2004). Asymptotic shape for the chemical distance and first-passage percolation on the infinite Bernoulli cluster. *ESAIM Probab. Stat.,* **8**, 169–199.

[93] Garet, O. and Marchand, R. (2008). First-passage competition with different speeds: positive density for both species is impossible. *Electron. J. Probab.*, **13**, 2118–2159.

[94] Garet, O. and Marchand, R. (2014). Large deviations for the contact process in random environment. *Ann. Probab.*, **42**, 1438–1479.

[95] Garet, O., Marchand, R., Procaccia, E. and Théret, M. Continuity of the time and isoperimetric constants in supercritical percolation. *Electron. Commun. Probab.*, **20**, no. 92, 11 pp.

[96] Garet, O., Gouere, J-B. and Marchand, R. (2013). The number of open paths in oriented percolation. *arXiv:1312:2571.*

[97] Georgiou, N., Rassoul-Agha, F. and Seppäläinen, T. (2015). Geodesics and the competition interface for the corner growth model. *arXiv:1510.00860.*

[98] Georgiou, N., Rassoul-Agha, F. and Seppäläinen, T. (2015). Stationary cocycles and Busemann functions for the corner growth model. *arXiv:1510.00859.*

[99] Georgiou, N., Rassoul-Agha, F. and Seppäläinen, T. (2016). Variational formulas and cocycle solutions for directed polymer and percolation models. *Comm. Math. Phys.*, **346**, no. 2, 741–779.

[100] Gouéré, J.-B. (2007). Shape of territories in some competing growth models. *Ann. Appl. Probab.*, **17**(4), 1273–1305.

[101] Gouéré, J.-B. (2014). Monotonicity in first-passage percolation. *ALEA Lat. Am. J. Probab. Math. Stat.*, **11**, 565–569.

[102] Grimmett, G. (1999). *Percolation.* 2nd edition. *Springer, Berlin.*

[103] Grimmett, G. and Kesten, H. (1984). First-passage percolation, network flow and electrical resistances *Z..Wahrsch. Verw. Gebiete*, **66**, 335–366.

[104] Gross, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.*, **97**, 1061–1083.

[105] Häggström, O. and Meester, R. (1995). Asymptotic shapes for stationary first passage percolation. *Ann. Probab.*, **23**, 1511–1522.

[106] Häggström, O. and Pemantle, R. (1998). First passage percolation and a model for competing spatial growth. *J. Appl. Probab.*, **35**, 683–692.

[107] Häggström, O. and Pemantle, R. (2000). Absence of mutual unbounded growth for almost all parameter values in the two-type Richardson model. *Stochastic Process. Appl.*, **90**, 207–222.

[108] Hammersley, J. and Welsh, D. First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. 1965 *Proc. Internat. Res. Semin., Statist. Lab., Univ. California, Berkeley, Calif.*, 61–110, *Springer-Verlag, New York.*

[109] Hanson, J. (2013). Ph.D. thesis.

[110] Higuchi, Y. and Zhang, Y. (2000). On the speed of convergence for two-dimensional first passage Ising percolation. *Ann. Probab.*, **28**, 353–378.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

[111] Hoffman, C. (2005). Coexistence for Richardson type competing spatial growth models. *Ann. Appl. Probab.*, **15**(1B), 739–747.

[112] Hoffman, C. (2008). Geodesics in first passage percolation. *Ann. Appl. Probab.*, **18**, 1944–1969.

[113] Howard, C. D. Models of first-passage percolation. *Probability on discrete structures,* 125–173, Encyclopaedia Math. Sci., 110, *Springer, Berlin,* 2004.

[114] Howard, C. D. and Newman, C. M., From greedy lattice animals to Euclidean first-passage percolation. *Perplexing Problems in Probability, (R. T. Durrett, ed.), Progr. Probab.*, **44**, Birkhäuser, Boston, 107–119.

[115] Howard, C. D. and Newman, C. M. (2001). Geodesics and spanning trees for Euclidean first-passage percolation. *Ann. Probab.*, **29**, 577–623.

[116] Huse, D. and Henley, C. (1985). Pinning and roughening of domain walls in Ising systems due to random impurities. *Phys. Rev. Lett.*, **54**, 2708–2711.

[117] Huse, D. A., Henley, C. L., Fisher, D. S.: (1985). Huse, Henley, and Fisher respond. *Phys. Rev. Lett.*, **55**, 2924–2924.

[118] Johansson, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.*, **209**, 437–476.

[119] Johansson, K. (2000). Transversal fluctuations for increasing subsequences on the plane. *Probab. Theory Related Fields*, **116**, 445–456.

[120] Kardar, M. (1985). Roughening by impurities at finite temperatures. *Phys. Rev. Lett.*, **55**, 2923–2923.

[121] Kardar, M., Parisi, G. and Zhang, Y. (1986). Dynamic scaling of growing interfaces. *Phys. Rev. Lett.*, **56**, 889–892.

[122] Kardar, M. and Zhang, Y.-C. (1987). Scaling of directed polymers in random media. *Phys. Rev. Lett.* , **56**, 2087–2090.

[123] Kesten, H. (1986). Aspects of first passage percolation. *École d'Été de Probabilités de Saint Flour XIV,* Lecture Notes in Mathematics, **1180**, 125–264.

[124] Kesten, H. (1987). Surfaces with minimal random weight and maximal flows: A higher dimensional version of first passage percolation, *Illinois J. Math.*, **31**, 99–166.

[125] Kesten, H. (1993). On the speed of convergence in first-passage percolation. *Ann. Appl. Probab.*, **3**, 296–338.

[126] Kesten, H. First-passage percolation. *From classical to modern probability,* 93–143, Progr. Probab., 54, *Birkhäuser, Basel,* 2003.

[127] Kesten, H. and Zhang, Y. (1996). A central limit theorem for critical first-passage percolation in two dimensions. *Probab. Theory Related Fields*, **107**, 137–160.

[128] Kingman, J. F. C. (1968). The ergodic theory of sub additive stochastic processes. *J. Roy. Statist. Soc. Ser.*, **30**, 499–510.

[129] Kingman, J. F. C. (1973). Subadditive ergodic theory. *Ann. Probab.*, **1**, 883–899.

[130] Knuth, D. (1970). Permutations, matrices, and generalized Young tableaux. *Pacific J. Math.*, **34**, 709–727.

[131] Krishnan, A. (2016). Variational formula for the time-constant of first-passage percolation I: Homogenization. *Comm. Pure Appl. Math.*, **69**, no. 10, 1984–2012.

[132] Krug, J. (1987). Scaling relation for a growing surface. *Phys. Rev. A*, **36**, 5465–5466.

[133] Krug, J. and Spohn, H. *Kinetic roughening of growing interfaces.* Edited by C. Godrèche. Collection Aléa-Saclay: Monographs and Texts in Statistical Physics, 1. Cambridge University Press, Cambridge, 1992. xvi+588 pp. ISBN: 0-521-41170-X

[134] Kubota, N. (2013). Rates of convergence in first passage percolation with low moment conditions. *arXiv:1306.5917.*

[135] Kuhr, J.-T., Leisner, M. and Frey, E. (2011). Range expansion with mutation and selection: dynamical phase transition in a two-species Eden model. *New Journal of Physics*, **13**.

[136] Lagatta, T. and Wehr, J. (2010). A shape theorem for Riemannian first-passage percolation. *J. Math. Phys.*, **51**, 14 pp.

[137] Lalley, S. (2003). Strict convexity of the limit shape in first-passage percolation. *Electron. Commun. Probab.*, **8**, 135–141.

[138] Lee, S. (1993). An inequality for Greedy Lattice Animals. *Ann. Appl. Probab.*, **3**, 1170–1188.

[139] Licea, C. and Newman, C. M. (1996). Geodesics in two-dimensional first-passage percolation. *Ann. Probab.*, **24**, 399–410.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

[140] Licea, C. and Newman, C. and Piza, M. (1996). Superdiffusivity in first-passage percolation. *Probab. Theory Related Fields*, **106**, 559–591.

[141] Lieb, E. H. and Loss, M. Analysis. Second edition. Graduate Studies in Mathematics, 14. *American Mathematical Society, Providence, RI,* 2001. xxii+346 pp. ISBN: 0-8218-2783-9.

[142] Liggett, T. (1985). An Improved Subadditive Ergodic Theorem. *Ann. Probab.*, **13**, 4, 1279–1285.

[143] Marchand, R. (2002). Strict inequalities for the time constant in first passage percolation. *Ann. Appl. Probab.*, **12**, 1001–1038.

[144] Martin, J. (2004). Limiting shape for directed percolation models. *Ann. Probab.*, **32**, 2908–2937.

[145] Mourrat, J.-C. (2012). Lyapunov exponents, shape theorems and large deviations for the random walk in random potential *ALEA, Lat. Am. J. Probab. Math. Stat.*, **9**, 165–209.

[146] Nakajima, S. (2016). Maximal edge-traversal time in First Passage Percolation. *arXiv:1605.04787*

[147] Newman, C. M. (1995). A surface view of first-passage percolation. *Proc. International Congress of Math.,* Vol. 1, 2 (Zürich, 1994), 1017–1023, Birkhäuser, Basel.

[148] Newman, C. M. *Topics in disordered systems.* Springer, 1997.

[149] Newman, C. M. and Piza, M. (1995). Divergence of shape fluctuations in two dimensions. *Ann. Probab.*, **23**, 977–1005.

[150] Pemantle, R. and Peres, Y. (1994). Planar first-passage percolation times are not tight. *Probability and phase transition (Cambridge, 1993)*, 261-264, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 420, *Kluwer Acad. Publ., Dordrecht*.

[151] Pimentel, L. P. R. (2007). Multitype shape theorems for first passage percolation models. *Adv. Appl. Probab.*, **39**(1), 53–76.

[152] Reimer, D. (2000). Proof of the van den Berg-Kesten conjecture. *Combin. Probab. Comput.*, **9**, 27–32.

[153] Rhee, W. (1995). On Rates of Convergence for Common Subsequences and First Passage Time. *Ann. Appl. Probab.*, **5**, 44–48.

[154] Rockafellar, R. *Convex Analysis.* Princeton University Press, 1970.

[155] Richardson, D. (1973). Random growth in a tessellation. *Proc. Cambridge Philos. Soc.*, **74**, 515–528.

[156] Rossignol, R. (2006). Threshold for monotone symmetric properties through a logarithmic Sobolev inequality. *Ann. Probab.*, **35**, 1707–1725.

[157] Rost, H. (1981). Nonequilibrium behaviour of a many particle process: density profile and local equilibria. *Z. Wahrsch. Verw. Gebiete*, **58**, 41–53.

[158] Rossignol, R. and Théret. (2010). Law of large numbers for the maximal flow through tilted cylinders in two-dimensional first passage percolation. *Stochastic Process. Appl.,* **120**, 873–900.

[159] Rossignol, R. and Théret. (2013). Lower large deviations for the maximal flow through tilted cylinders in two-dimensional first passage percolation. *ESAIM Probab. Stat.,* **17**, 70–104.

[160] Smythe, R. and Wierman, J. C. (1977). First-passage percolation on the square lattice. I. *Advances in Appl. Probability*, **9**, no. 1, 38–54.

[161] Smythe, R. and Wierman, J. C. (1978). First-passage percolation on the square lattice. III. *Advances in Appl. Probability*, **10**, no. 1, 155–171.

[162] Sodin, S. (2014). Positive temperature versions of two theorems on first-passage percolation. *Geometric Aspects of Functional Analysis*, **2116**, Lecture Notes in Mathematics, 441–453.

[163] Steele, J. (1986). An Efron-Stein inequality for nonsymmetric statistics. *Ann. Statist.*, **14**, 753–758.

[164] Steele, J. and Zhang, Y. (2003) Nondifferentiability of the time constants of first-passage percolation. *Ann. Probab.*, **31**, 1028–1051.

[165] Talagrand, M. (1994). On Russo's approximate zero-one law. *Ann. Probab.*, **22**, 1576–1587.

[166] Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. I.H.E.S.*, **81**, 73–205.

[167] Tessera, R. (2014). Speed of convergence in first passage percolation and geodesicity of the average distance. *arXiv:1410.1701.*

[168] Timár, A. (2013). Boundary-connectivity via graph theory. *Proc. Amer. Math. Soc*, **141**, 475–480.

[169] Tracy, C. A.; Widom, H. (1994). Level-spacing distributions and the Airy kernel, *Comm. Math. Phys.*, **159**, 151–174

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

[170] Vahidi-Asl, M. Q. and Wierman, J. C. (1990). First-passage percolation on the Voronoi tessellation and Delaunay triangulation. *Random Graphs '87* , 341–359, Wiley, New York.

[171] Vahidi-Asl, M. Q. and Wierman, J. C. (1992). A shape result for first-passage percolation on the Voronoi tessellation and Delaunay triangulation. *Random Graphs '87* , 247–262, Wiley, New York.

[172] Van der Hofstad, R. (2014). Random Graphs and Complex Networks. Vol. II. *Online at http://www.win.tue.nl/˜rhofstad/NotesRGCNII.pdf.*

[173] Wehr, J. (1997). On the number of infinite geodesics and ground states in disordered systems. *J. Statist. Phys.*, **87**, 439–447.

[174] Wehr, J. and Aizenman, M. (1987). Fluctuations of extensive functions of quenched random couplings. *J. Stat. Phys.*, **60**, 287–306.

[175] Wehr, J. and Woo, J. (1998). Absence of geodesics in first-passage percolation on a half-plane. *Ann. Probab.*, **26**, 358–367.

[176] Wierman, J. C., (1980). Weak moment conditions for time coordinates in first-passage percolation models. *J. Appl. Probab.*, **17**, 968–978.

[177] Wierman, J. C. and Reh, W. (1978). On conjectures in first passage percolation theory. *Ann. Probab.*, **6**, 388–397.

[178] Wolf, D. E. and Kertesz, J. (1987). Noise reduction in Eden models: I. *J. Phys. A*, **20**, L257-L261.

[179] Wüthrich, M. V. (1998). Scaling identity for crossing Brownian motion in a Possonian potential. *Probab. Theory Related Fields,* **112**, 299–319.

[180] Yao, C.-L. (2015). Limit theorems for critical first-passage percolation on the triangular lattice. *arXiv: 1602.00065.*

[181] Zabolitsky, J. G. and Stauffer, D. (1986). Simulation of large Eden clusters. *Phys. Rev. A* **34** 1523–1530.

[182] Zhang, Y. (1993). A shape theorem for epidemics and forest fires with finite range interactions. *Ann. Probab.*, **21**, 1755–1781.

[183] Zhang, Y. (2000). Critical behavior for maximal flows on the cubic lattice. *J. Stat. Phys.* **98**, 799–811.

[184] Zhang, Y. (2006). The divergence of fluctuations for shape in first passage percolation. *Probab. Theory Related Fields*, **136**, 298–320.

[185] Zhang, Y. (1999). Double behavior of critical first-passage percolation. *Perplexing Problems in Probability*, 143–158.

[186] Zhang, Y. (2016). Limit theorems for maximum flows on a lattice. To appear in *Probab. Theory Related Fields.*

[187] Zhang, Y. (2010). On the concentration and the convergence rate with a moment condition in first-passage percolation. *Stochastic Process. Appl.*, **120**, 1317–1341.

[188] Zhang, Y. (2008). Shape fluctuations are different in different directions. *Ann. Probab.*, **36**, 331–362.

[189] Zhang, Y. (1995). Supercritical behaviors in first-passage percolation. *Stochastic Process. Appl.*, **59**, 251–266.

[190] Zhang, Y. and Zhang, Y. (1984). A limit theorem for $N_{0,n}/n$ in first-passage percolation. *Ann. Probab.*, **12**, 1068–1076.

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms

Photo by Nkir Sasune, Photographer

Photo courtesy of Joseph Rabinoff

Photo courtesy of Jillian Roberts

First-passage percolation (FPP) is a fundamental model in probability theory that has a wide range of applications to other scientific areas (growth and infection in biology, optimization in computer science, disordered media in physics), as well as other areas of mathematics, including analysis and geometry. FPP was introduced in the 1960s as a random metric space. Although it is simple to define, and despite years of work by leading researchers, many of its central problems remain unsolved.

In this book, the authors describe the main results of FPP, with two purposes in mind. First, they give self-contained proofs of seminal results obtained until the 1990s on limit shapes and geodesics. Second, they discuss recent perspectives and directions including (1) tools from metric geometry, (2) applications of concentration of measure, and (3) related growth and competition models. The authors also provide a collection of old and new open questions. This book is intended as a textbook for a graduate course or as a learning tool for researchers.

ISBN 978-1-4704-4183-8

9 781470 441838

ULECT/68

For additional information and updates on this book, visit

www.ams.org/bookpages/ulect-68

www.ams.org

Licensed to Princeton Univ. Prepared on Thu Jun 9 08:22:22 EDT 2022for download from IP 128.112.200.107.
License or copyright restrictions may apply to redistribution; see https://www.ams.org/publications/ebooks/terms