Galois Martingales and the Hyperbolic Subset of the p-adic Mandelbrot Set

by

Rafe Jones

A. B., Amherst College, 1998

Sc. M., Brown University, 2001

Thesis

Submitted in partial fulfillment of the requirements for

the Degree of Doctor of Philosophy

in the Department of Mathematics at Brown University

Providence, Rhode Island

May 2005

This dissertation by Rafe Jones is accepted in its present form by
the Department of Mathematics as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date _____ _____

Joseph Silverman, Director

Recommended to the Graduate Council

Date _____ _____

Michael Rosen, Reader

Date _____ _____

Jill Pipher, Reader

Approved by the Graduate Council

Date _____ _____

Dean of the Graduate School

The vita of Rafe Jones

Rafe Jones came into the world on October 10, 1976, in Knoxville, Tennessee. After growing up in Knoxville and in Berea, Kentucky, he set out for New England to attend Amherst College in Amherst, Massachusetts. During the summers he worked on a research project with professor Jan Pearce of Berea College, initially responding "anything but number theory" when asked what he wanted to work on. After receiving a Bachelor's Degree in Mathematics and French from Amherst in 1998, he headed to Paris, France, to spend a very enjoyable year as a visiting student at the Ecole Normale Superieure in Paris. With his French well-honed, he returned to the U.S. to begin graduate studies in math at Brown University. There, he received his Master's Degree in 2001. By this time his views on number theory had evidently softened, as he embarked on a thesis in that area under the supervision of Joseph Silverman. In the course of his studies at Brown, he has been supported by a University Fellowship, several Teaching Assistantships and Fellowships, and a Dissertation Fellowship.

Acknowledgements

The completion of this thesis owes much to the many people who have supported and encouraged me during its preparation.

First and foremost, I would like to thank my advisor, Joseph Silverman. His wealth of knowledge and keen insight were instrumental both in helping me find a good problem and then solve it. Just as important were his good humor, vast patience, and clear explanations, which never failed to leave me inspired after our weekly meetings. Thanks, Joe, for being so generous with your time, your enthusiasm, and your good humor.

I'm also grateful to my readers, Mike Rosen and Jill Pipher. Special thanks goes to Mike Rosen, who had many helpful discussions with me on various aspects of my thesis. Thanks also go to Jonathan Wise, Graeme Wilkin, Michelle Manes, Ben Brubaker, Brian Munson, Tim Record, Dev Sinha, all the Mikes, and the many other grad students and postdocs at Brown who've shared their knowledge and many good times. Particular thanks go to Aaron Hoffman, who first tipped me off about the existence of martingales during a post-basketball-game conversation. And hearty thanks too to Natalie Johnson, Doreen Pappas, and Audrey Aguiar, who have bailed me out of all manner of logistical jams over the years. Thank you also to Heinrich Hock and Matthias Vom Hau, my housemates and good friends for the past five years. I will dearly miss living at 440 Wayland Ave.

I would be remiss not to mention the two most important mentors in my pre-grad-school mathematical life: Jan Pearce, whose energy infected me during high school and several summers in college, and David Armacost, whose guidance of my undergraduate thesis led to an experience that convinced me I wanted to be a mathematician.

Finally, thanks to my parents Libby and Roger and my brother Gil for their longstanding encouragement and love. Most of all, thanks to Kate Stange, who has had to deal first hand with so many of the ups and downs I went through during my thesis research. You are a wonderful companion.

Contents

Introduction

This thesis has four main elements: dynamics, algebraic number theory (in function fields), group theory, and probability. The dynamics provides inspiration for the problem, the algebraic number theory allows an important reformulation, the group theory does the heavy lifting, and the probability theory synthesizes the group theory into a form that solves the problem. Chapter 1 gives the dynamical background, Chapter 2 the number-theoretic translation, and Chapter 3 the group-theoretic results. Chapter 4 deals almost solely with probability, while Chapter 5 applies the probability to give a solution to the problem.

In this introduction, we sketch of the origins of the problem, state the problem, and give a detailed outline of our solution. The origins of the problem and the first stages of the solution lie in the field of algebraic dynamics, which can be broadly defined as the study of function iteration over algebraic/arithmetical sets, such as algebraic number fields and rings, polynomial rings, finite fields, p-adic fields, algebraic curves, etc. This relatively new field evolved naturally from the field of complex dynamics, whose roots stretch back to the work of Julia and Fatou in the early 1900s. Complex dynamics enjoyed an explosion of deep results in the 1980s; one could note in particular Benoit Mandelbrot's popularization of the remarkable set that bears his name, Dennis Sullivan's No Wandering Domains theorem [36], and Curt McMullen's proof of the nonexistence of generally convergent algorithms in degree larger than three [23].

In the 1990s, some attention began to turn to p-adic analogues of the blossoming complex theory. Michael Herman and Fields medalist Jean-Christophe Yoccoz [17] had already gone in this direction with their 1983 paper on a non-Archimedean version of Siegel's linearization theorem [34]. Rob Benedetto proved a partial analogue of the No Wandering Domains theorem in 2000 [6] and Juan Rivera-Letelier, a student of Yoccoz, fleshed out much of the theory of p-adic Fatou and Julia sets in the early part of the 2000s [29, 30]. An unexpected impetus for exploring p-adic dynamics has come from physicists, who began to explore what the world looked like through p-adic eyes. The first paper whose title contained the phrase "p-adic dynamics" appeared in the Physics literature in 1989 [37].

The work presented in this thesis stems from a question regarding the degree to which p-adic dynamics and complex dynamics are analogous. This question deals with the p-adic analogue of the complex Mandelbrot set. The complex Mandelbrot set, defined to be

$$M = \{c \in \mathbb{C} : 0 \text{ has a bounded orbit under iteration of } z^2 + c\}$$

not only created a stir among researchers, but its striking images reached an audience far beyond the mathematical community. It has assumed a place as one of the most widely recognized mathematical objects. Thus the p-adic analogue of M, at first blush, seems to promise a similar treasure trove of complexity. This promise unfortunately proves false, as the analogous set is simply the closed unit disk (for $p \neq 2$).

Peering into the set more closely, however, reveals a particular subset that is more interesting. It is the subset of parameter values c such that $z^2 + c$ is hyperbolic; see below for a definition of hyperbolicity and see (1) for a definition of the set. Its complex analogue has been much studied [1, 14, 19, 22]. This complex analogue is a large subset of M, accounting for at least 96% of the area [11], and is conjectured to be the interior of M [24]. The problem this thesis sets out to resolve is to determine the "size" of the hyperbolic subset of the p-adic Mandelbrot set. We show that it is in a certain sense a measure zero subset, contrasting sharply with the complex case.

The first hurdle is to say what is meant by size: there is no suitable notion of measure on the p-adic analogue of $\mathbb{C}$ because it is not locally compact. The way around this, and the first step in the solution of the above problem, is to use the reduction homomorphism to translate the problem into one of dynamics over $\overline{\mathbb{F}}_p$. In $\overline{\mathbb{F}}_p$ we define density measures closely related to the well-known Dirichlet density and natural density (see e.g. [20]). The proof then follows a path through algebraic number theory, then the Galois theory of function fields, and eventually into the realm of stochastic processes, where it reaches its conclusion. This method of proof appears to be highly unusual, and may be fruitful in answering other density questions regarding dynamically defined sets. We now give a detailed outline of the argument.

In Chapter 1 we begin with some definitions and background. We call a rational function *hyperbolic* if all its critical points are attracted to attracting cycles (see page 8 for definitions of these terms). Hyperbolic maps have many nice properties, and are the subject of the biggest unsolved conjecture in complex dynamics (see page 9 for a statement and [24] for more detail). The map $z^2 + c$ has critical points at infinity and 0, and infinity is an attracting fixed point. Therefore $z^2 + c$ is hyperbolic if and only if 0 is attracted to an attracting cycle. Thus the set

$$\mathcal{H}(\mathbb{C}) = \{c \in M : 0 \text{ is attracted to an attracting cycle of } z^2 + c\} \tag{1}$$

is the *hyperbolic subset* of M. We examine the analogous set defined over the field $\mathbb{C}_p$, which is the

smallest complete, algebraically closed extension of $\mathbb{Q}_p$; it is therefore the p-adic analogue of $\mathbb{C}$. Let M_p be the natural analogue of M over $\mathbb{C}_p$. It's easily seen that $M_p = \{c \in \mathbb{C}_p : |c| \leq 1\}$ provided that $p \neq 2$. (Proposition 1.2). *Throughout this discussion, we take p to be a prime different from 2.*

However, the subset

$$\mathcal{H}(\mathbb{C}_p) = \{c \in M_p : 0 \text{ is attracted to an attracting cycle of } z^2 + c\}$$

is not so easily characterized. Letting $\phi : \{|c| \leq 1\} \to \overline{\mathbb{F}}_p$ be the reduction homomorphism (see (1.2)), we establish in Corollary 1.6 that $\mathcal{H}(\mathbb{C}_p) = \phi^{-1}(\mathcal{H}(\overline{\mathbb{F}}_p))$, where

$$\mathcal{H}(\overline{\mathbb{F}}_p) = \{\alpha \in \overline{\mathbb{F}}_p : 0 \text{ is periodic under iteration of } x^2 + \alpha\}.$$

(Note that by periodic we mean that the orbit of 0 is a cycle; some authors refer to this as purely periodic.) We define two notions of density for subsets of $\overline{\mathbb{F}}_p$, called Dirichlet density and natural density, that are closely related to the densities of the same names defined for subsets of primes in $\mathbb{F}_p[x]$. We denote Dirichlet density by δ and natural density by D (see (1.6) and (1.7) for the definitions). Our main result (Theorem 1.7) is that $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$, and its proof is the principal goal of all of our subsequent work. We also establish, using a similar method, a companion result: we show $D(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$ for $p \equiv 3 \bmod 4$, and we conjecture that $D(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$ holds for all $p \neq 2$.

In Chapter 2, we take the first steps toward a proof by giving two translations of the problem. Although the definition of $\mathcal{H}(\overline{\mathbb{F}}_p)$ says that the forward orbit of 0 under iteration of $x^2 + \alpha$ is a cycle, we focus in Section 2.1 on the *inverse* orbit of 0 under $x^2 + \alpha$. Clearly for any $\alpha \in \overline{\mathbb{F}}_p$ the forward orbit of 0 under iteration of $x^2 + \alpha$ is contained in $\mathbb{F}_p(\alpha)$. If 0 is periodic, however, this forward orbit coincides with one branch of the inverse orbit of 0 in $\overline{\mathbb{F}}_p$, and thus 0 has nth preimages in $\mathbb{F}_p(\alpha)$ for each $n \geq 1$. We show (Proposition 2.1) that the converse is also true. Therefore, setting $f_\alpha = x^2 + \alpha$, and denoting by $f_\alpha^{-n}(0)$ the set of nth preimages of 0 (in $\overline{\mathbb{F}}_p$) under iteration of f_α, we have

$$\mathcal{H}(\overline{\mathbb{F}}_p) = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset \text{ for all } n \geq 1\}.$$

We can therefore define a sequence of sets

$$\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset\} \tag{2}$$

that serve as progressively better "approximations" of $\mathcal{H}(\overline{\mathbb{F}}_p)$ in the sense that $\mathcal{I}_n \supseteq \mathcal{I}_{n+1}$ and $\mathcal{H}(\overline{\mathbb{F}}_p) = \bigcap_n \mathcal{I}_n$. We show that if $\delta(\mathcal{I}_n)$ exists for all n and $\lim_{n\to\infty} \delta(\mathcal{I}_n) = 0$, then $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$.

In Section 2.2, we prove that $\delta(\mathcal{I}_n)$ exists and give a method of computing it using the Galois groups of certain algebraic extensions of $\mathbb{F}_p(x)$. Our main tool in this endeavor is the Tchebotarev Density theorem for function fields. In order to use it, we find a set of primes in $\mathbb{F}_p[x]$ that is

expressible in terms of the Artin symbol (see page 25 for a definition), and whose Dirichlet density (in the sense of (2.3)) is equal to $\delta(\mathcal{I}_n)$. Note that the set $f_\alpha^{-n}(0)$ consists of the roots of f_α^n, and thus $f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset$ if and only if the factorization of f_α^n over $\mathbb{F}_p(\alpha)$ contains at least one linear term. There is a thus a close relationship between $\mathcal{I}_n$ and the following set of primes in $\mathbb{F}_p[x]$:

$$I_n = \{\mathfrak{p} \subseteq \mathbb{F}_p[x] : f_x^n \text{ has a linear factor mod } \mathfrak{p}\},$$

where $f_x = y^2 + x \in \mathbb{F}_p(x)[y]$. Indeed, we show that $\delta(\mathcal{I}_n) = \delta(I_n)$, where this second Dirichlet density is the usual one for sets of primes in a function field (2.3). We then use some standard arguments in algebraic number theory to show that I_n differs by only finitely many primes from a set of primes defined in terms of the Artin symbol (2.14). This allows us to apply the Tchebotarev Density theorem. Let K_n be the splitting field of f_x^n over $K = \mathbb{F}_p(x)$, and $G_n = \mathrm{Gal}\,(K_n/K)$. In Theorem 2.18 we show that $\delta(\mathcal{I}_n)$ exists for all n and equals

$$\frac{1}{\#G_n} \#\left\{g \in G_n : g \text{ fixes at least one root of } f_x^n\right\}. \tag{3}$$

The same statement hold for $D(\mathcal{I}_n)$, provided that the extensions K_n/K are geometric for all n (see Definition 2.14), a statement we conjecture to hold for $p \neq 2$ but which we can only show when $p \equiv 3 \bmod 4$ (Corollary 3.40). To illustrate Theorem 2.18, we give here a few values of $\delta(\mathcal{I}_n)$. We can describe the $n = 1$ case completely: the roots of f_x are $\{\sqrt{-x}, -\sqrt{-x}\}$, which we label $\{a_1, a_2\}$. Clearly we have $G_1 = \{e, (a_1 \;\; a_2)\}$, whence $\delta(\mathcal{I}_1) = 1/2$. In Example 2.19, page 29, we work out the case $n = 2$, showing that $\delta(\mathcal{I}_2) = 3/8$. Only with significantly more work (Theorem 3.2, (5.28), and Corollary 5.11) can we show $\delta(\mathcal{I}_3) = 39/128$. Moreover we note in Chapter 3 (see the discussion on page 56) that for $n > 7$, $\delta(\mathcal{I}_n)$ may depend on the prime p, and cannot in general be easily computed.

In Chapter 3 we undertake an analysis of the groups G_n. We do this through the study of $H_n = \mathrm{Gal}\,(K_n/K_{n-1})$. In (3.2), we show that K_n is obtained from K_{n-1} by adjoining the square roots of 2^{n-1} elements. Thus $|H_n| \leq 2^{2^{n-1}}$, and we call H_n *maximal* if this inequality is an equality. One principal result of the chapter is that for all $p \neq 2$, H_n is maximal for n squarefree, and if $p \equiv 3 \bmod 4$ then H_n is maximal for all n (Theorem 3.2). The other result that is of central importance in later chapters is Corollary 3.23 at the end of Section 3.2 (see below for explanation).

In Section 3.1, we prove some basic properties about f_x^n, and we introduce the polynomials $p_n \in \mathbb{F}_p[x]$, defined by $p_1 = x$ and $p_n = p_{n-1}^2 + x$ for $n \geq 2$, which play an important role. We show that the discriminant of f_x^n is a product of powers of p_i for $i \leq n$ (Proposition 3.9) and use this to prove that G_n is not alternating, i.e. composed of even permutations (Corollary 3.10). In Section 3.2 we examine the center of G_n. We note that G_n is a 2-group, and we establish a series of propositions about 2-groups acting on certain sets. This culminates in Theorem 3.22, which states

that if G_n is not alternating then a certain permutation of the roots of f_x^n must lie in G_n. Indeed, as shown in Corollary 3.23, this permutation must lie in H_n, proving that H_n is nontrivial for all n.

In Section 3.3 we use abelian Kummer Theory to show that H_n is maximal if and only if p_n is a square in K_{n-1} (Theorem 3.27). This relies heavily on Corollary 3.23. In Section 3.4 we use the work of the previous sections to show that H_n is maximal if and only if a certain element $\Phi_n \in \mathbb{F}_p[x]$ is a not a square in $K = \mathbb{F}_p(x)$ (Theorem 3.38). Specifically, Φ_n is the primitive part of p_n:

$$\Phi_n = \prod_{d|n} (p_d)^{\mu(n/d)}.$$

One can easily show that the degree of Φ_n is odd when n is squarefree (Corollary 3.30), establishing the maximality of H_n for squarefree n. When n is not squarefree one cannot rule out the possibility that Φ_n is a square in K. However, the facts that Φ_n is separable over $\mathbb{Q}$ for all n (see the proof of Theorem 3.2, page 54) and the degree of Φ_n grows like 2^n suggest this is unlikely. We thus conjecture that H_n is maximal for all n. In section 3.5 we adapt an argument of Stoll [35] to show that if $p \equiv 3 \bmod 4$ then H_n is maximal for all n.

The results of Chapter 3 provide some insight into the structure of H_n and therefore G_n, but there is no obvious way to use them to compute the limit as $n \to \infty$ of the expression in (3). In Chapters 4 and 5 we build a stochastic process where the main results of Chapter 3 have a natural interpretation. The tools of the theory of stochastic processes then allow us to prove this process is eventually 0 with probability 1, and this is enough to establish Theorem 1.7.

Chapter 4 focuses on proving the process we seek exists and fleshing out definitions and standard results from the theory of stochastic processes. In Section 4.1 we recall that a discrete-time stochastic process (or simply *process* for short) is a sequence $\{X_n\}_{n\geq 0}$ of random variables defined on a common probability space. We consider only processes whose random variables take positive-integer values. Such a process can be thought of as a game of chance, with X_n denoting a gambler's score at turn n. We prove that there exists a process where the probability of the gambler having score t at turn n is determined by the structure of G_n. Specifically,

$$\mathbf{P}(X_n = t) = \frac{1}{\#G_n} \#\left\{g \in G_n : g \text{ fixes } t \text{ roots of } f_x^n\right\} \tag{4}$$

We actually prove an even stronger property (see (4.3)). From (4) and the remark immediately before (3) it follows that

$$\delta(\mathcal{I}_n) = \mathbf{P}(X_n > 0). \tag{5}$$

For more on why probability theory is a relatively natural tool in this context, see the introduction to Chapter 4 on page 59.

The main work of section 4.1 is to show that a process satisfying (4) exists. We call this process the Galois process of the iterates of f, or GP(f) for short. (Because there is no possibility of ambiguity, we drop the x from previous notation and simply write $f = y^2 + x$.) In Section 4.2 we give some probabilistic background, including definitions of martingales and Markov chains and one version of the basic martingale convergence theorem (Theorem 4.9). In Section 4.3 we present the basic theory of branching processes, which not only is useful in Chapter 5 but also illustrates some of the definitions of Section 4.2.

In Chapter 5 all of the threads come together. In Section 5.1 we use Corollary 3.23, which guarantees the existence of a certain type of element in each H_n, to establish that GP(f) is a martingale. Thus GP(f) is the first known example of a class of processes we call *Galois martingales* (see the remark on page 80), giving some justification for the first two words of this thesis' title. The martingale convergence theorem then shows that with probability 1 the sequence $\{X_n(\omega)\}_{n\geq 0}$ is eventually constant (where ω is any element in the underlying probability space).

In Section 5.2 we compute, under the assumption that H_n is maximal, the conditional distribution of X_n given $X_{n-1} = t$ for any value of t. Thus, using the metaphor of the gambler, when n is such that H_n is maximal we have explicit information about the probability that the gambler's score goes from t to t' at turn n. Using Theorem 3.2, which says that H_n is maximal when n is squarefree, we show that for any $t \geq 1$ and $m \geq 1$,

$$\mathbf{P}\left(X_n = t \text{ for all } n \geq m\right) = 0.$$

This is Theorem 5.8. It quickly follows that GP(f) converges to 0 with probability 1. We deduce from this in Section 5.3 that $\lim_{n\to\infty} \mathbf{P}(X_n > 0) = 0$, and then from (5) and the remark following (2) we conclude that $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$, which proves Theorem 1.7. Finally, we show that under the assumption that H_n is maximal for all n, GP(f) is a particularly simple branching process. We also give explicit values for $\delta(\mathcal{I}_n)$ under this assumption (see (5.28) and Corollary 5.11).

Chapter 1

A Subset of the p-adic Mandelbrot Set

The complex Mandelbrot set $M(\mathbb{C})$ is obtained by considering the polynomials $f_c(z) = z^2 + c$, where c is a complex number. Writing f_c^n for the nth iterate of f_c, we define $M(\mathbb{C})$ as the set of parameter values c such that the orbit of 0 remains bounded:

$$\{c \in \mathbb{C} : f_c^n(0) \not\to \infty \text{ as } n \to \infty\}.$$

There are several equivalent characterizations of the Mandelbrot set; see [8] for details. We're interested in the related set

$$\mathcal{H}(\mathbb{C}) = \{c \in \mathbb{C} : f_c \text{ has an attracting cycle in } \mathbb{C}\}.$$

We recall that a cycle of a rational function R is a collection of points $\zeta_1, \ldots, \zeta_n$ such that $R(\zeta_i) = \zeta_{i+1}$ for $1 \leq i \leq n-1$ and $R(\zeta_n) = \zeta_1$. This is equivalent to each ζ_i being a root of $R^n(z) - z$, where R^n is the nth iterate of R. We refer to any point in a cycle as *periodic*. Some authors use the phrase *purely periodic* to describe such points, so as to underline the distinction with points that are not in a cycle but map into one under iteration, often called preperiodic.

Throughout this thesis, we apply the term periodic only to points that lie in a cycle

We say that the cycle $\zeta_1, \ldots, \zeta_n$ has period n. If, moreover, each ζ_i is a root of $R^n(z) - z$ but not of $R^m(z) - z$ for any $m < n$, we say the cycle has *primitive period* n (some authors use the term *exact period* n). This is equivalent to each ζ_i being a root of $R^n(z) - z$ but not of $R^m(z) - z$ for any m dividing n. Note also that the cycle $\zeta_1, \ldots, \zeta_n$ has primitive period n if and only if the ζ_i are distinct.

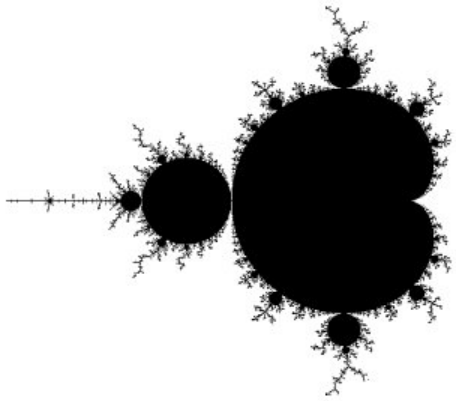


Figure 1.1: M, the complex Mandelbrot set. Also this thesis' coolest picture.

The above cycle is attracting if we have $|(R^n)'(\zeta_i)| < 1$ for any i (this is equivalent to the inequality holding for all i). It is *super-attracting* if $|(R^n)'(\zeta_i)| = 0$ for any i. Each point ζ_i of an attracting cycle lies in an open set with the following property: for any $u \in U_i$, $R^{nm}(u) \to \zeta_i$ as $n \to \infty$ (see [5]). We refer to the largest such U_i as the *immediate basin of attraction* of ζ_i. The union of the sets $U_1, \ldots, U_n$ is called the immediate basin of attraction for the attracting cycle $\zeta_1, \ldots, \zeta_n$.

The relation between $\mathcal{H}(\mathbb{C})$ and $M(\mathbb{C})$ arises from the following Theorem (see e.g. [5, Sec. 9.3]):

Theorem 1.1. *Let R be a rational function of degree at least two. Then the immediate basin of attraction for each attracting cycle of R contains a critical point of R.*

Thus every attracting cycle must attract a critical point. Let's return to the case of f_c. Its only critical points are 0 and ∞, with the latter being fixed for all c. So if f_c has an attracting cycle, it must attract the only nonfixed critical point, namely 0. Thus the orbit of 0 approaches the cycle, and in particular remains bounded. This shows that $\mathcal{H}(\mathbb{C}) \subset M(\mathbb{C})$. Indeed, we have

$$\mathcal{H}(\mathbb{C}) = \{c \in M(\mathbb{C}) : 0 \text{ is attracted to an attracting cycle of } z^2 + c\}.$$

However, $\mathcal{H}(\mathbb{C})$ is a proper subset of $M(\mathbb{C})$, as can be seen by taking $c = -2$: under $z^2 - 2$, after two iterations 0 lands on the fixed point 2, which is not attracting. Thus there can be no attracting cycles, but on the other hand $f_c^n(0) \not\to \infty$.

Notice that if $c \in \mathcal{H}(\mathbb{C})$, then every critical point of f_c is attracted to an attracting cycle. Rational functions with this property are known as *hyperbolic* [24]. If $c \in M(\mathbb{C}) - \mathcal{H}(\mathbb{C})$, then the only attracting cycle of f_c is infinity (which is a fixed point), yet the critical point 0 does not go to infinity. Hence $\mathcal{H}(\mathbb{C})$ consists of all c in $M(\mathbb{C})$ for which f_c is hyperbolic. We thus refer to $\mathcal{H}(\mathbb{C})$ as the *hyperbolic subset* of $M(\mathbb{C})$.

We can visualize pieces of $\mathcal{H}(\mathbb{C})$ by setting

$$\mathcal{H}(\mathbb{C})^{(i)} = \{c \in \mathbb{C} : f_c \text{ has an attracting cycle of primitive period } i \text{ in } \mathbb{C}\}.$$

Then $\mathcal{H}(\mathbb{C})^{(1)}$ is the cardioid at the center of $M(\mathbb{C})$, and $\mathcal{H}(\mathbb{C})^{(2)}$ is the circle immediately to its left; see Figure 1.1. The set $\mathcal{H}(\mathbb{C})^{(3)}$ consists of three disjoint open disks tangent to the cardioid. Subsequent $\mathcal{H}(\mathbb{C})^{(i)}$ are also disjoint unions of open disks, and the disks are increasingly smaller and more numerous (see [8, p.316] or, more colorfully, [9, Chapter 17]).

We are interested in the relative size of $\mathcal{H}(\mathbb{C})$ to $M(\mathbb{C})$. Both $\mathcal{H}(\mathbb{C})$ and $M(\mathbb{C})$ are Lebesgue measurable, and the measure of $\mathcal{H}(\mathbb{C})$ exceeds 1.503 while the measure of $M(\mathbb{C})$ is less than 1.562 [11]. Thus the hyperbolic subset $\mathcal{H}(\mathbb{C})$ accounts for nearly all of the measure of $M(\mathbb{C})$. In [24] it is conjectured that $\mathcal{H}(\mathbb{C})$ is the interior of $M(\mathbb{C})$. Indeed, this is a special case of the famous conjecture that hyperbolic maps are dense in the set of rational functions of degree n – the biggest unsolved conjecture in complex dynamics – and even this simplest possible special case remains open. Since $\mathcal{H}(\mathbb{C})$ is suspected of being the interior of $M(\mathbb{C})$, it is reasonable to suppose that $\mathcal{H}(\mathbb{C})$ accounts for all of $M(\mathbb{C})$ except a set of measure 0. But the boundary $\partial M(\mathbb{C})$ is a complicated creature: Shishikura ([33]) has shown that it has Hausdorff dimension 2, and some believe that it has positive 2-dimensional Lebesgue measure. At any rate, it is true that $\mathcal{H}(\mathbb{C})$ is at least a very large subset of $M(\mathbb{C})$.

In this thesis we consider the relative size of the hyperbolic subset of the natural p-adic analogue of the Mandelbrot set. The p-adic equivalent of the complex numbers is $\mathbb{C}_p$, the smallest complete, algebraically closed extension of $\mathbb{Q}_p$. We use only the following information about $\mathbb{C}_p$ (see [31] for details):

1. There is a unique absolute value $|\cdot|$ on $\mathbb{C}_p$ that extends the p-adic absolute value on $\mathbb{Q}_p$. This absolute value satisfies the ultrametric inequality $|x+y| \leq \max\{|x|, |y|\}$. Moreover, if $|x| \neq |y|$, then
$$|x+y| = \max\{|x|, |y|\}. \tag{1.1}$$

2. The ultrametric inequality implies that the closed unit disk $D = \{x \in \mathbb{C}_p : |x| \leq 1\}$ is a subring of $\mathbb{C}_p$. The open unit disk $U = \{x \in \mathbb{C}_p : |x| < 1\}$ is the unique maximal ideal in D. The quotient D/U is isomorphic to $\overline{\mathbb{F}}_p$, the algebraic closure of the finite field with p elements. There is thus a natural homomorphism
$$\phi : D \to \overline{\mathbb{F}}_p \tag{1.2}$$
called the *reduction homomorphism*. Following convention, we sometimes denote $\phi(x)$ by $\overline{x}$.

We can extend ϕ to a mapping from $\mathbb{P}^1(\mathbb{C}_p)$ to $\mathbb{P}^1(\overline{\mathbb{F}}_p)$ by setting $\phi(x) = \infty$ for any x with $|x| > 1$.

For $c \in \mathbb{C}_p$, we set $f_c(x) = x^2 + c$. We define the p-adic Mandelbrot set $M(\mathbb{C}_p)$ in the obvious way:

$$M(\mathbb{C}_p) = \{c \in \mathbb{C}_p : f_c^n(0) \not\to \infty \text{ as } n \to \infty\}.$$

We also define

$$\mathcal{H}(\mathbb{C}_p) = \{c \in \mathbb{C}_p : f_c \text{ has an attracting cycle in } \mathbb{C}_p\}.$$

We wish to show that in fact

$$\mathcal{H}(\mathbb{C}_p) = \{c \in M(\mathbb{C}_p) : 0 \text{ is attracted to an attracting cycle of } z^2 + c\}, \tag{1.3}$$

whence $\mathcal{H}(\mathbb{C}_p)$ is legitimately the hyperbolic subset of $M(\mathbb{C}_p)$, i.e., the subset consisting of all of the hyperbolic maps.

However, Theorem 1.1 does not hold in general in the p-adic setting. For instance, the map $f(x) = x^p$ has an infinite number of attracting cycles, yet only two critical points. Thus we have to do a bit of work to show that $\mathcal{H}(\mathbb{C}_p) \subset M(\mathbb{C}_p)$ and that f_c is hyperbolic for all $c \in \mathcal{H}(\mathbb{C}_p)$. First we determine $M(\mathbb{C}_p)$, a task whose easiness contrasts sharply with its complex equivalent.

Proposition 1.2. *For all p, we have $M(\mathbb{C}_p) = D$.*

Proof: First take $|c| > 1$, and consider f_c. We show by induction that $|f_c^n(0)| = |c|^{2^{n-1}}$. This is clear for $n = 1$. For the induction step, assume $n > 1$ and note that

$$|f_c^n(0)| = \left|f_c(f_c^{n-1}(0))\right| = \left|(f_c^{n-1}(0))^2 + c\right|.$$

By the inductive hypothesis, we have $\left|(f_c^{n-1}(0))^2\right| = |c|^{2^{n-1}}$, and this is greater than $|c|$ since $n > 1$ and $|c| > 1$. Thus by (1.1) we have $\left|(f_c^{n-1}(0))^2 + c\right| = |c|^{2^{n-1}}$. This finishes the induction, and we now have $|f_c^n(0)| \to \infty$ as $n \to \infty$, so $c \notin M(\mathbb{C}_p)$. Thus $M(\mathbb{C}_p) \subseteq D$.

Now suppose that $c \in D$, i.e. $|c| \leq 1$. Then $|f_c(0)| \leq 1$, and if we suppose that $\left|f_c^{n-1}(0)\right| \leq 1$, then

$$|f_c^n(0)| = \left|(f_c^{n-1}(0))^2 + c\right| \leq \max\left\{\left|(f_c^{n-1}(0))\right|^2, |c|\right\} \leq 1,$$

by the ultrametric inequality. Thus $|f_c^n(0)| \leq 1$ for all n, so we have $c \in M(\mathbb{C}_p)$. Hence $M(\mathbb{C}_p) = D$. ■

We next show that if $p \neq 2$, then $c \notin M(\mathbb{C}_p)$ implies $c \notin \mathcal{H}(\mathbb{C}_p)$. Before doing this, we state a useful equation that follows from the chain rule. Let K be a field, $x \in K$, and $f_c = x^2 + c \in K[x]$.

In the following we take f_c^0 to be the identity:

$$(f_c^n)'(x) = \prod_{i=0}^{n-1} f_c'(f_c^i(x)) = 2^n \prod_{i=0}^{n-1} f_c^i(x) \tag{1.4}$$

Moreover, if $x_1, \ldots, x_n$ is a cycle in K under f_c, then for any k with $1 \leq k \leq n$ we have

$$(f_c^n)'(x_k) = 2^n \prod_{i=0}^{n-1} x_i \tag{1.5}$$

Proposition 1.3. *Suppose that $p \neq 2$ and $|c| > 1$. Then f_c has no attracting cycles in $\mathbb{C}_p$.*

Proof: We show first that all points x_0 whose absolute value is not $\sqrt{|c|}$ are attracted to infinity. If $|x_0| > \sqrt{|c|}$, then by (1.1) we have $\left|x_0^2 + c\right| = \left|x_0^2\right| > |c|$, so $|f_c(x_0)| > |c|$. Using the same argument as in the proof of Proposition 1.2 one can show that $\lim_{n\to\infty} f_c^n(x_0) = \infty$. Now take $|x_0| < \sqrt{|c|}$. Then $|f_c(x_0)| = |c|$, so $\left|f_c^2(x_0)\right| = |c|^2 > |c|$. The argument of the preceding paragraph then applies to show that $\lim_{n\to\infty} f_c^n(x_0) = \infty$.

The orbit of any cycle is bounded, and so must remain within the set $\{x_0 \in \mathbb{C}_p : |x_0| = \sqrt{|c|}\}$. Let x_1 belong to a cycle of f_c. Then using (1.4) and $p \neq 2$ we have

$$|(f_c^n)'(x_1)| = \prod_{i=0}^{n-1} \left|f_c^i(x_1)\right| = |c|^{\frac{n}{2}} > 1.$$

Hence the cycle to which x_1 belongs is not attracting. ■

This Proposition shows that $\mathcal{H}(\mathbb{C}_p) \subseteq M(\mathbb{C}_p)$ for $p \neq 2$. When $p = 2$ this is not true. Indeed, if we take $1 < |c| < 4$ then one can easily show that the fixed points x_1, x_2 of f_c both have absolute value $\sqrt{|c|}$, which is less than 2. Hence $|f_c'(x_i)| = |2x_i| < 1$. This is only one of many respects in which the case $p = 2$ differs from other primes, and we prefer to focus on behavior of typical primes. *Hence from now on we assume $p \neq 2$.*

It remains to show for $p \neq 2$ that $\mathcal{H}(\mathbb{C}_p) \neq M(\mathbb{C}_p)$, and we also wish to show that f_c is hyperbolic for $c \in \mathcal{H}(\mathbb{C}_p)$. Both of these statements follow from a theorem of Rivera-Letelier [28], which we state below

First we develop some terminology. A rational function $R \in \mathbb{C}_p(x)$ may be considered as a map $\mathbb{P}^1(\mathbb{C}_p) \to \mathbb{P}^1(\mathbb{C}_p)$. We can represent it in homogeneous coordinates as

$$R([x, y]) = [f(x, y), g(x, y)],$$

where f and g are relatively prime homogeneous polynomials with coefficients in D, and $\deg(f) = \deg(g) = d = \deg(R)$. We can also assume that at least one coefficient of f or g has absolute value 1. We set $\overline{R}([x, y]) = \left[\overline{f}(x, y), \overline{g}(x, y)\right]$, where $\overline{f}$ and $\overline{g}$ are the functions obtained by applying the

reduction homomorphism to each coefficient of f and g, respectively. We call $\overline{R}$ the *reduction* of R; note that $\overline{R}$ is a map $\mathbb{P}^1(\overline{\mathbb{F}}_p) \to \mathbb{P}^1(\overline{\mathbb{F}}_p)$. We say R has *good reduction* if $\deg(R) = \deg(\overline{R})$, i.e. if the only common zero of $\overline{f}$ and $\overline{g}$ in $\overline{\mathbb{F}}_p \times \overline{\mathbb{F}}_p$ is $(x, y) = (0, 0)$.

Note that any monic polynomial with coefficients in D must have good reduction. Hence f_c has good reduction if $c \in M(\mathbb{C}_p)$.

We make one more observation before stating the theorem. Since $D/U \cong \overline{\mathbb{F}}_p$, we can write D as a disjoint union of open balls of radius 1. Indeed, for $\alpha \in \overline{\mathbb{F}}_p$, define $B_\alpha = \phi^{-1}(\alpha) = \{x \in \mathbb{C}_p : \overline{x} = \alpha\}$, and note that for any $\alpha \in \overline{\mathbb{F}}_p$ and any x_0 such that $\overline{x_0} = \alpha$ we have $B_\alpha = \{x \in \mathbb{C}_p : |x_0 - x| < 1\}$ (this follows from the ultrametric inequality). We then have

$$D = \bigcup_{\alpha \in \overline{\mathbb{F}}_p} B_\alpha.$$

Theorem 1.4 (Rivera-Letelier [28]). *Suppose $R \in \mathbb{C}_p(x)$ has good reduction, and for each $\alpha \in \mathbb{P}^1(\overline{\mathbb{F}}_p)$, set $B_\alpha = \phi^{-1}(\alpha) = \{c \in \mathbb{P}^1(\mathbb{C}_p) : \overline{c} = \alpha\}$. Then*

1. *$R(B_\alpha) = B_{\overline{R}(\alpha)}$ for all $\alpha \in \mathbb{P}^1(\overline{\mathbb{F}}_p)$.*
2. *B_α contains a point of an attracting cycle if and only if α is periodic under $\overline{R}$ and $(\overline{R}^k)'(\alpha) = 0$, where k is the primitive period of α. In this case B_α contains a unique periodic point β and B_α is the immediate basin of attraction of β.*
3. *If p does not divide $\deg(R)$ then the immediate basin of attraction of any attracting cycle of R contains a critical point.*

We can quickly derive several corollaries from Part 3 of Theorem 1.4. First, it follows immediately that f_c is hyperbolic for $c \in \mathcal{H}(\mathbb{C}_p)$. Second, we may show $\mathcal{H}(\mathbb{C}_p) \neq M(\mathbb{C}_p)$ just as in the complex case by taking $c = -2$. Under $f_{-2} = x^2 - 2$, 0 lands on the fixed point 2 after one interation, and $|f'_{-2}(2)| = |4| = 1$ since $p \neq 2$. Thus the critical point 0 is not attracted to an attracting cycle, so f_{-2} has no attracting cycles in $\mathbb{C}_p$. Finally, using the same argument as in the complex case, we have that f_c is not hyperbolic if $c \in M(\mathbb{C}_p) - \mathcal{H}(\mathbb{C}_p)$.

Theorem 1.4 also shows that we may obtain a great deal of information about f_c from its reduction $\overline{f_c}$.

Corollary 1.5. *Let $c \in D$. Then f_c has a unique attracting cycle if 0 is periodic under iteration of $\overline{f_c}$ and no attracting cycles otherwise. Moreover, if 0 is periodic of primitive period n under iteration of $\overline{f_c}$, then the unique attracting cycle of f_c has primitive period n.*

Proof: The fact that f_c has at most one attracting cycle follows immediately from part 3 of Theorem 1.4. We now show that f_c has an attracting cycle of order n if and only if 0 is periodic of primitive period n under iteration of $\overline{f_c}$. This is sufficient to prove the Corollary.

Suppose first that f_c has an attracting cycle of order n, and let $x_1, \ldots, x_n$ be the points in this cycle. Since $|c| \leq 1$, all points outside D are attracted to infinity, so the cycle must lie in D. We then have by (1.5),

$$|(f_c^n)'(x_1)| = \prod_{i=0}^{n-1} |x_i| .$$

Since the cycle is attracting, this expression is less than 1. However, because the cycle is in D, $|x_i| \leq 1$ for each i. Hence $|x_j| < 1$ for some j, so $\overline{x_j} = 0$. By part 2 of Theorem 1.4, no two x_i can be in the same B_α, so the reductions $\overline{x_1}, \ldots, \overline{x_n}$ are distinct. Thus $\overline{x_1}, \ldots, \overline{x_n}$ is a cycle under $\overline{f_c}$ containing 0.

Now suppose that $\alpha_1, \ldots, \alpha_n$ is a cycle of order n under $\overline{f_c}$ and that $\alpha_j = 0$. Using (1.5) gives us

$$(f_c^n)'(\alpha_k) = 2^n \prod_{i=0}^{n-1} \alpha_i$$

for any k. Since $\alpha_j = 0$, this product is 0 for all k, so by part 2 of Theorem 1.4 each B_{α_i} contains a point x_i of an attracting cycle. By part 3 of Theorem 1.4 there can be at most one attracting cycle of f_c, so $x_1, \ldots, x_n$ is an attracting cycle. Each x_i is in a different ball B_{α_i}, and thus the points of the cycle are distinct. ■

Corollary 1.5 gives us a tidy characterization of $\mathcal{H}(\mathbb{C}_p)$. It says that one may determine if some $c \in M(\mathbb{C}_p)$ belongs to $\mathcal{H}(\mathbb{C}_p)$ by examining only the behavior of $\overline{f_c}$. Specifically,

$$c \in \mathcal{H}(\mathbb{C}_p) \Longleftrightarrow B_{\overline{c}} \subset \mathcal{H}(\mathbb{C}_p) \Longleftrightarrow 0 \text{ is periodic under iteration of } x^2 + \overline{c}.$$

Let us set

$$\mathcal{H}(\overline{\mathbb{F}}_p) = \overline{\mathcal{H}(\mathbb{C}_p)} = \{\alpha \in \overline{\mathbb{F}}_p : 0 \text{ is periodic under } x^2 + \alpha\};$$

we call $\mathcal{H}(\overline{\mathbb{F}}_p)$ the hyperbolic locus of $\overline{\mathbb{F}}_p$. We have the following relation between $\mathcal{H}(\overline{\mathbb{F}}_p)$ and $\mathcal{H}(\mathbb{C}_p)$:

Corollary 1.6. *Let $\phi : D \to \overline{\mathbb{F}}_p$ be the reduction homomorphism. Then*

$$\mathcal{H}(\mathbb{C}_p) = \phi^{-1}(\mathcal{H}(\overline{\mathbb{F}}_p)) = \bigcup_{\alpha \in \mathcal{H}(\overline{\mathbb{F}}_p)} B_\alpha.$$

We note that $\mathcal{H}(\overline{\mathbb{F}}_p)$ is not empty, as it clearly contains 0. It is also not all of $\overline{\mathbb{F}}_p$, since we may choose $\alpha = -2 \in \mathbb{F}_p$. Under $x^2 + \alpha$, 0 maps to the fixed point 2 after two iterations, and $2 \neq 0$ since $p \neq 2$. Hence $-2 \notin \mathcal{H}(\overline{\mathbb{F}}_p)$ for all $p \geq 2$.

We now focus our attention on the relative size of the hyperbolic subset $\mathcal{H}(\mathbb{C}_p) \subseteq M(\mathbb{C}_p)$. The field $\mathbb{C}_p$ is not locally compact (see e.g. [31]), so one cannot define a Haar measure on it, and thus we must look elsewhere for a notion of size. Corollary 1.6 provides us with a solution: use the reduction homomorphism to move the problem to $\overline{\mathbb{F}}_p$.

In $\overline{\mathbb{F}}_p$, there are two notions of density that interest us. We refer to the first as the Dirichlet density, as it corresponds to the density measure of that name in number fields and function fields. For its definition we need a bit of terminology. Given $\alpha \in \overline{\mathbb{F}}_p$, let $\deg \alpha = [\mathbb{F}_p(\alpha) : \mathbb{F}_p]$, and define $N(\alpha)$ to be $p^{\deg \alpha}$. If $\mathcal{S} \subseteq \overline{\mathbb{F}}_p$, we define the Dirichlet density $\delta(\mathcal{S})$ to be

$$\delta(\mathcal{S}) = \lim_{s \to 1^+} \frac{\sum_{\alpha \in \mathcal{S}} (\deg \alpha)^{-1} N(\alpha)^{-s}}{\sum_{\alpha \in \overline{\mathbb{F}}_p} (\deg \alpha)^{-1} N(\alpha)^{-s}}. \tag{1.6}$$

The second notion of density is called natural density and is more immediately satisfying. It, too, corresponds to similarly named densities in number fields and function fields. Intuitively, as k grows, $\mathbb{F}_{p^k}$ offers progressively better "approximations" of $\overline{\mathbb{F}}_p$. We therefore define the natural density $D(\mathcal{S})$ of $\mathcal{S} \subseteq \overline{\mathbb{F}}_p$ to be

$$D(\mathcal{S}) = \lim_{k \to \infty} \frac{\#\{\mathcal{S} \cap \mathbb{F}_{p^k}\}}{p^k}. \tag{1.7}$$

Clearly for a given $\mathcal{S}$ neither $\delta(\mathcal{S})$ nor $D(\mathcal{S})$ need necessarily exist. If $D(\mathcal{S})$ exists then a quick computation shows that $\delta(\mathcal{S})$ exists and that the two are equal. On the other hand, there are easily described sets that have Dirichlet density but no natural density. Dirichlet density is the less intuitive of the two and much harder to calculate, but it is more flexible and applies to many more sets.

Because of the connection between $\mathcal{H}(\overline{\mathbb{F}}_p)$ and $\mathcal{H}(\mathbb{C})$, we call $\delta(\mathcal{H}(\overline{\mathbb{F}}_p))$ the *hyperbolic density* of $M(\mathbb{C}_p)$. The main business of this thesis is to show:

Theorem 1.7. *For all $p \neq 2$, $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$.*

We also show that $D(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$ when $p \equiv 3 \bmod 4$, and conjecture that this holds for all p (see the remarks following the proof of Theorem 1.7 on page 85). In the next chapter we begin the perhaps unexpectedly far-reaching journey of proving these statements. For now, though, we take a look at a few concrete pieces of $\mathcal{H}(\overline{\mathbb{F}}_p)$ and from these determine the analogues in $\mathcal{H}(\mathbb{C}_p)$ of some of the previously mentioned pieces of the complex Mandelbrot set.

Pursuing an analogy with the complex case, we make the following definitions:

$$\begin{aligned} \mathcal{H}(\mathbb{C}_p)^{(i)} &= \{c \in \mathbb{C}_p : x^2 + c \text{ has an attracting cycle of primitive period } i\} \\ \mathcal{H}(\overline{\mathbb{F}}_p)^{(i)} &= \{\alpha \in \overline{\mathbb{F}}_p : x^2 + \alpha \text{ has a super-attracting cycle of primitive period } i\} \end{aligned}$$

Clearly $\mathcal{H}(\mathbb{C}_p)$ is the disjoint union of the $\mathcal{H}(\mathbb{C}_p)^{(i)}$ and $\mathcal{H}(\overline{\mathbb{F}}_p)$ is the disjoint union of the $\mathcal{H}(\overline{\mathbb{F}}_p)^{(i)}$. By Corollary 1.5 we have for every i,

$$\mathcal{H}(\mathbb{C}_p)^{(i)} = \phi^{-1}(\mathcal{H}(\overline{\mathbb{F}}_p)^{(i)}) = \bigcup_{\alpha \in \mathcal{H}(\overline{\mathbb{F}}_p)^{(i)}} B_\alpha. \tag{1.8}$$

For $\alpha \in \overline{\mathbb{F}}_p$, we set $f_\alpha(x) = x^2 + \alpha$. We then have for all p,

$$\begin{aligned}\mathcal{H}(\overline{\mathbb{F}}_p)^{(1)} &= \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha(0) = 0\} = \{0\} \\ \mathcal{H}(\overline{\mathbb{F}}_p)^{(2)} &= \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^2(0) = 0 \text{ and } f_\alpha(0) \neq 0\} = \{\alpha \in \overline{\mathbb{F}}_p : \alpha^2 + \alpha = 0 \text{ and } \alpha \neq 0\} = \{-1\}\end{aligned}$$

By (1.8) we then have $\mathcal{H}(\mathbb{C}_p)^{(1)} = B_0$ and $\mathcal{H}(\mathbb{C}_p)^{(2)} = B_{-1}$. Note that $\mathcal{H}(\mathbb{C}_p)^{(1)}$ is the p-adic equivalent of the main cardioid of the complex Mandelbrot set, while $\mathcal{H}(\mathbb{C}_p)^{(2)}$ is the p-adic equivalent of the circle immediately to the left of the cardioid.

For any i, $\mathcal{H}(\overline{\mathbb{F}}_p)^{(i)}$ is the set of $\alpha \in \overline{\mathbb{F}}_p$ such that $f_\alpha^i(0) = 0$ and $f_\alpha^k(0) \neq 0$ for all $k \mid i$ with $k < i$. We can restate this by setting $f_t = x^2 + t$ and considering the polynomials in t resulting from the evaluation of $f_t^i(x)$ at $x = 0$. This family of polynomials plays an important role in Chapter 3, and we thus define

$$p_i(t) = f_t^i(0).$$

We then have that $\mathcal{H}(\overline{\mathbb{F}}_p)^{(i)}$ consists of all $\alpha \in \overline{\mathbb{F}}_p$ that are roots of p_i but not also roots of p_k for some $k \mid i$ with $k < i$. We say that such α have *primitive Mandelbrot period i.* This notion plays an important role in Section 3.4; here we content ourselves with a few examples. Note that $p_1 = t$ and $p_2 = t^2 + t$, and since these polynomials split completely over $\mathbb{Q}$ their roots are the same for all p. Hence $\mathcal{H}(\mathbb{C}_p)^{(1)}$ and $\mathcal{H}(\mathbb{C}_p)^{(2)}$ do not depend on p. When $i \geq 3$ the roots of p_i depend on the prime p. Indeed, $p_3 = t^4 + 2t^3 + t^2 + t$, and the roots of $p_3/p_1 = t^3 + 2t^2 + t + 1$ are those with primitive Mandelbrot period 3. When $p = 3$ we must move to a degree 3 extension of $\mathbb{F}_p$ in order to get all the roots of p_3. When $p = 5$ a degree 2 extension will do, while when $p = 101$ we have the three roots $4, 37$, and 58 in $\mathbb{F}_p$.

One could hope to gain information about $\delta(\mathcal{H}(\overline{\mathbb{F}}_p))$ by analyzing the size of the extension of $\mathbb{F}_p$ necessary to contain $\mathcal{H}(\overline{\mathbb{F}}_p)^{(i)}$. However, this question seems complicated, as there may be points of very high primitive Mandelbrot period even in $\mathbb{F}_p$.

Thus we will take a different approach to determining $\delta(\mathcal{H}(\overline{\mathbb{F}}_p))$. Although the sets $\mathcal{H}(\overline{\mathbb{F}}_p)^{(i)}$ are defined by looking at the forward orbit of 0 under $x^2 + \alpha$, our technique is to examine the *inverse* orbit. We lay out the details in the next chapter.

Chapter 2

Two Reformulations of the Problem

In Chapter 1, we defined the hyperbolic density of the p-adic Mandelbrot set $M(\mathbb{C}_p)$ to be $\delta(\mathcal{H}(\overline{\mathbb{F}}_p))$, where δ is the Dirichlet density (see (1.6)) and

$$\mathcal{H}(\overline{\mathbb{F}}_p) = \{\alpha \in \overline{\mathbb{F}}_p : 0 \text{ is periodic under } x^2 + \alpha\}$$

Our main task is to show Theorem 1.7, namely $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$. In this chapter we give two reformulations of the problem of determining $\delta(\mathcal{H}(\overline{\mathbb{F}}_p))$ and $D(\mathcal{H}(\overline{\mathbb{F}}_p))$ (see (1.7)). First, by considering the inverse orbit of 0 under $x^2 + \alpha$, we relate these quantities to the densities of certain subsets of $\overline{\mathbb{F}}_p$ that are invariant under the action of $\mathrm{Gal}\,(\overline{\mathbb{F}}_p/\mathbb{F}_p)$ (see (2.1) and Corollary 2.6). We then show that these sets have densities equal to the densities (in an appropriate sense – see (2.3) and (2.4)) of certain sets of primes in the ring $\mathbb{F}_p[x]$. These sets of primes turn out to have a nice characterization in terms of the Artin symbol (see page 25 for a definition). Thus we may apply the Tchebotarev Density theorem to reduce our original problem to purely Galois-theoretic considerations.

Before diving in, we make a notational remark: *For this entire chapter we take p to be a prime number different from 2.*

2.1 The Inverse Orbit of Zero

Let $\alpha \in \overline{\mathbb{F}}_p$, and set $f_\alpha = x^2 + \alpha$. We begin with the simple observation that the forward orbit of 0 under f_α, defined to be $\{f_\alpha^n(0) : n = 1, 2, 3, \ldots\}$, is contained in $\mathbb{F}_p(\alpha)$. The same is not true of the

inverse orbit of 0, i.e. the set

$$\{\beta \in \overline{\mathbb{F}}_p : f_\alpha^n(\beta) = 0 \text{ for some } n \geq 1\}.$$

Indeed, for many α there is an n such that no root of f_α^n is contained in $\mathbb{F}_p(\alpha)$. Dynamically, this means that 0 has no nth preimages in $\mathbb{F}_p(\alpha)$. We now show that the collection of α for which there is **no** such n is precisely $\mathcal{H}(\overline{\mathbb{F}}_p)$.

Proposition 2.1. *Let $\alpha \in \overline{\mathbb{F}}_p$, and for each $n \geq 1$ let $f_\alpha^{-n}(0)$ be the set of roots of f_α^n in $\overline{\mathbb{F}}_p$. Let $\mathcal{S}$ be any finite subset of $\overline{\mathbb{F}}_p$ containing $\mathbb{F}_p(\alpha)$. Then $\alpha \in \mathcal{H}(\overline{\mathbb{F}}_p)$ if and only if*

$$f_\alpha^{-n}(0) \cap \mathcal{S} \neq \emptyset$$

for all $n \geq 1$.

Proof: Suppose first that $\alpha \in \mathcal{H}(\overline{\mathbb{F}}_p)$, so that 0 has a periodic forward orbit $c_1, c_2, \ldots, c_{m-1}, c_m = 0$. Then $f_\alpha^j(0) = c_j$ for $j = 1, \ldots, m-1$, and $f_\alpha^m(0) = 0$. Clearly this entire orbit lies in $\mathbb{F}_p(\alpha)$, and thus in $\mathcal{S}$. For any $n > 0$, we may choose l with $m \geq lm - n > 0$, and this gives

$$0 = f_\alpha^{lm}(0) = f_\alpha^n(f_\alpha^{lm-n}(0)) = f_\alpha^n(c_{lm-n}).$$

Thus $c_{lm-n} \in f_\alpha^{-n}(0)$, showing that $f_\alpha^{-n}(0) \cap \mathcal{S}$ is nonempty.

Conversely, suppose that $f_\alpha^{-n}(0) \cap \mathcal{S}$ is nonempty for all $n \geq 1$. Then for each $n \geq 1$ we may find a root b_n of f_α^n in $\mathcal{S}$. Since $\mathcal{S}$ is finite, there must exist $n_1 < n_2$ such that $b_{n_1} = b_{n_2}$. Thus $f_\alpha^{n_2}(b_{n_1}) = f_\alpha^{n_2}(b_{n_2}) = 0$. Combining this with $f_\alpha^{n_1}(b_{n_1}) = 0$ yields

$$f_\alpha^{n_2-n_1}(0) = f_\alpha^{n_2-n_1}(f_\alpha^{n_1}(b_{n_1})) = f_\alpha^{n_2}(b_{n_1}) = 0.$$

Hence 0 is periodic under f_α. ■

Corollary 2.2. $\mathcal{H}(\overline{\mathbb{F}}_p) = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset \text{ for all } n \geq 1\}$.

Proof: Put $\mathcal{S} = \mathbb{F}_p(\alpha)$ in Proposition 2.1. ■

Corollary 2.3. *For any $k \geq 1$ we have*

$$\mathcal{H}(\overline{\mathbb{F}}_p) \cap \mathbb{F}_{p^k} = \{\alpha \in \mathbb{F}_{p^k} : f_\alpha^{-n}(0) \cap \mathbb{F}_{p^k} \neq \emptyset \text{ for all } n \geq 1\}.$$

Proof: If $\alpha \in \mathcal{H}(\overline{\mathbb{F}}_p) \cap \mathbb{F}_{p^k}$, then put $\mathcal{S} = \mathbb{F}_{p^k}$ in Proposition 2.1. Since $\mathbb{F}_p(\alpha) \subseteq \mathbb{F}_{p^k}$, we have $f_\alpha^{-n}(0) \cap \mathbb{F}_{p^k} \neq \emptyset$ for all $n \geq 1$. To show the reverse inclusion, if $\alpha \in \mathbb{F}_{p^k}$ and $f_\alpha^{-n}(0) \cap \mathbb{F}_{p^k} \neq \emptyset$ for all $n \geq 1$, then Proposition 2.1 gives that $\alpha \in \mathcal{H}(\overline{\mathbb{F}}_p)$. ■

Using the characterizations given in Corollaries 2.2 and 2.3, we now describe supersets of $\mathcal{H}(\overline{\mathbb{F}}_p)$, which we use later in this chapter to give upper bounds for $\delta(\mathcal{H}(\overline{\mathbb{F}}_p))$ and $D(\mathcal{H}(\overline{\mathbb{F}}_p))$. For each $n \geq 1$, define

$$\mathcal{I}_n \quad = \quad \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset\} \tag{2.1}$$

Note that $\alpha \in \mathcal{I}_n$ if f_α has an nth preimage that is "simple" in the sense that it is contained in $\mathbb{F}_p(\alpha)$ rather than an extension. These sets play a crucial role in much of the rest of this thesis. This is because they may be thought of as successively better "approximations" of $\mathcal{H}(\overline{\mathbb{F}}_p)$, as the next Proposition shows.

Proposition 2.4. *For each $n \geq 1$ we have $\mathcal{I}_n \supseteq \mathcal{I}_{n+1}$, and $\mathcal{H}(\overline{\mathbb{F}}_p) = \bigcap_{n\geq 1} \mathcal{I}_n$.*

Proof: let $\alpha \in \mathcal{I}_{n+1}$, and take $\beta \in \mathbb{F}_p(\alpha)$ such that $f_\alpha^{n+1}(\beta) = 0$. Then $f_\alpha^n(f_\alpha(\beta)) = 0$, so $f_\alpha(\beta)$ is a root of f_α^n, and because $\beta \in \mathbb{F}_p(\alpha)$ it follows that $f_\alpha(\beta) \in \mathbb{F}_p(\alpha)$. Therefore $\alpha \in \mathcal{I}_n$. That $\mathcal{H}(\overline{\mathbb{F}}_p) = \bigcap_{n\geq 1} \mathcal{I}_n$ follows immediately from Corollary 2.2. ■

We wish to use Proposition 2.4 to connect $\delta(\mathcal{I}_n)$ and $\delta(\mathcal{H}(\overline{\mathbb{F}}_p))$ on the one hand, and $D(\mathcal{I}_n)$ and $D(\mathcal{H}(\overline{\mathbb{F}}_p))$ on the other. To do this, we need a basic property of Dirichlet and natural density, which we give in the following proposition.

Proposition 2.5. *Let $\mathcal{S}, \mathcal{T}_1, \mathcal{T}_2, \ldots$ be subsets of $\overline{\mathbb{F}}_p$, and suppose that $\mathcal{S} \subseteq \mathcal{T}_i$ for each i. Suppose also that $\delta(\mathcal{T}_i)$ exists and $\lim_{i\to\infty} \delta(\mathcal{T}_i) = 0$. Then $\delta(\mathcal{S})$ exists and equals zero. A similar statement holds for natural density.*

Proof: Define

$$a_{\mathcal{S}}(s) = \frac{\sum_{\alpha\in\mathcal{S}} (\deg\alpha)^{-1} N(\alpha)^{-s}}{\sum_{\alpha\in\overline{\mathbb{F}}_p} (\deg\alpha)^{-1} N(\alpha)^{-s}}.$$

Define similar functions $a_{\mathcal{T}_i}(s)$ in the obvious way. Since $\mathcal{S} \subseteq \mathcal{T}_i$ and all sums involved are positive wherever they are defined, it follows immediately that $a_{\mathcal{S}}(s) \leq a_{\mathcal{T}_i}(s)$ for $s > 1$. Taking lim sups and using the assumption that $\delta(\mathcal{T}_i)$ exists gives

$$\limsup_{s\to 1^+} a_{\mathcal{S}}(s) \leq \limsup_{s\to 1^+} a_{\mathcal{T}_i}(s) = \lim_{s\to 1^+} a_{\mathcal{T}_i}(s) = \delta(\mathcal{T}_i).$$

Since $\lim_{i\to\infty} \delta(\mathcal{T}_i) = 0$ and $a_{\mathcal{S}}(s) \geq 0$ for $s > 1$, it follows that $\limsup_{s\to 1^+} a_{\mathcal{S}}(s) = 0$. Therefore $\lim_{s\to 1^+} a_{\mathcal{S}}(s) = 0$, proving that $\delta(\mathcal{S}) = 0$.

For the natural density version, $\mathcal{S} \subseteq \mathcal{T}_i$ implies that $\#(\mathcal{S} \cap \mathbb{F}_{p^k}) \leq \#(\mathcal{T}_i \cap \mathbb{F}_{p^k})$. Dividing by p^k and taking lim sups as $k \to \infty$ gives

$$\limsup_{k\to\infty} \frac{\#(\mathcal{S} \cap \mathbb{F}_{p^k})}{p^k} \leq \limsup_{k\to\infty} \frac{\#(\mathcal{T}_i \cap \mathbb{F}_{p^k})}{p^k} = \lim_{k\to\infty} \frac{\#(\mathcal{T}_i \cap \mathbb{F}_{p^k})}{p^k} = D(\mathcal{T}_i).$$

Because $\lim_{i\to\infty} D(\mathcal{T}_i) = 0$, it follows that $\limsup_{k\to\infty} \#(\mathcal{S} \cap \mathbb{F}_{p^k})/p^k = 0$, whence $D(\mathcal{S}) = 0$. ■

Corollary 2.6. *Suppose that for all $n \geq 1$, $\delta(\mathcal{I}_n)$ exists and $\lim_{n\to\infty} \delta(\mathcal{I}_n) = 0$. Then $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$. Similarly, if $D(\mathcal{I}_n)$ exists for all $n \geq 1$ and $\lim_{n\to\infty} D(\mathcal{I}_n) = 0$, then $D(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$.*

Proof: Immediate from Proposition 2.5 and the fact that $\mathcal{H}(\overline{\mathbb{F}}_p) = \bigcap_{n\geq 1} \mathcal{I}_n$ (Proposition 2.4). The natural density case is similar. ■

In order to show that $\delta(\mathcal{I}_n)$ and $D(\mathcal{I}_n)$ exist and to study them, we use the Tchebotarev Density theorem for function fields. However, this theorem deals with the densities of sets of prime ideals in the ring $\mathbb{F}_p[x]$, and so cannot be directly applied to $\mathcal{I}_n$. We thus wish to relate $\mathcal{I}_n$ to a set of primes in $\mathbb{F}_p[x]$, and this is our task for the remainder of this section. In the next section we elaborate on the application of Tchebotarev's theorem.

Recall that $\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset\}$. The condition $f_\alpha^{-n}(0) \cap \mathbb{F}_{p^k} \neq \emptyset$ is equivalent to the polynomial $f_\alpha^n(y) \in \mathbb{F}_p(\alpha)[y]$ having a root in $\mathbb{F}_p(\alpha)$. This, in turn, is equivalent to the factorization of $f_\alpha^n(y)$ over $\mathbb{F}_p(\alpha)$ having a linear term. Now the minimal polynomial of α over $\mathbb{F}_p$ is some irreducible $\pi_\alpha \in \mathbb{F}_p[x]$, and it generates a prime ideal $\mathfrak{p}_\alpha$. There is an obvious isomorphism

$$\phi : \mathbb{F}_p[x,y]/\mathfrak{p}_\alpha \xrightarrow{\sim} \mathbb{F}_p(\alpha)[y],$$

and $\phi^{-1}(f_\alpha^n(y)) = f_x^n(y)$, where $f_x = y^2 + x \in \mathbb{F}_p[x,y]$. Thus $f_\alpha^n(y)$ has a linear factor in $\mathbb{F}_p(\alpha)$ if and only if f_x^n has a linear factor mod $\mathfrak{p}_\alpha$. We sum up this discussion in the following equation:

$$\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_x^n \text{ has a linear factor mod } \mathfrak{p}_\alpha\}. \tag{2.2}$$

Note that (2.2) shows that whether α belongs to $\mathcal{I}_n$ depends only on its minimal polynomial π_α. Thus $\alpha \in \mathcal{I}_n$ if and only if all its conjugates are in $\mathcal{I}_n$; another way of saying this is that $\mathcal{I}_n$ is invariant under the action of $\mathrm{Gal}\,(\overline{\mathbb{F}}_p/\mathbb{F}_p)$.

We wish to relate the density of a Galois-invariant subset of $\overline{\mathbb{F}}_p$ with the density of a corresponding set of primes in $\mathbb{F}_p[x]$. To be able to do this we need a notion of density of a set of primes in $\mathbb{F}_p[x]$. There are two such notions, called Dirichlet and natural density, and they are quite similar to the definitions we have given for subsets of $\overline{\mathbb{F}}_p$. Put $A = \mathbb{F}_p[x]$, let P be the set of primes in A, and recall that if $\mathfrak{p} \in P$, then $N\mathfrak{p} = p^{\deg \mathfrak{p}}$. Given a set $T \subseteq P$, we define

$$\delta(T) = \lim_{s\to 1^+} \frac{\sum_{\mathfrak{p}\in T} N\mathfrak{p}^{-s}}{\sum_{\mathfrak{p}\in P} N\mathfrak{p}^{-s}}, \tag{2.3}$$

$$D(T) = \lim_{k\to\infty} \frac{\#\{\mathfrak{p} \in T : \deg \mathfrak{p} = k\}}{\#\{\mathfrak{p} \in P : \deg \mathfrak{p} = k\}}. \tag{2.4}$$

The relationship of these notions of density is the same as in the case of subsets of $\overline{\mathbb{F}}_p$ (see discussion on page 14). Indeed, we now show that these notions of density are essentially the same as the ones we defined for subsets of $\overline{\mathbb{F}}_p$, only restricted to Galois-invariant subsets. Recall that for $\alpha \in \overline{\mathbb{F}}_p$, we defined $\mathfrak{p}_\alpha$ to be the ideal (π_α), where π_α is the minimal polynomial for α over $\mathbb{F}_p$. Recall also that we defined $\deg \alpha = [\mathbb{F}_p(\alpha) : \mathbb{F}_p]$ (see (1.6)) and $N(\alpha) = p^{\deg \alpha}$.

Lemma 2.7. *Let $\mathcal{S} \subseteq \overline{\mathbb{F}}_p$ be invariant under the action of* $\mathrm{Gal}\,(\overline{\mathbb{F}}_p/\mathbb{F}_p)$*, and let*

$$T = \{\mathfrak{p} \in P : \mathfrak{p} = \mathfrak{p}_\alpha \text{ for some } \alpha \in \mathcal{S}\}.$$

Then

$$\sum_{\alpha \in \mathcal{S}} (\deg \alpha)^{-1} N(\alpha)^{-s} = \sum_{\mathfrak{p} \in T} N\mathfrak{p}^{-s}$$

and

$$\#(S \cap \mathbb{F}_{p^k}) = k \cdot \#\{\mathfrak{p} \in T : \deg \mathfrak{p} = k\} + O(p^{k/2}).$$

Proof: Consider the map $\psi : \overline{\mathbb{F}}_p \to P$ that takes α to $\mathfrak{p}_\alpha$. The Galois invariance of $\mathcal{S}$ is equivalent to $\mathcal{S}$ being the full inverse image of T under ψ. We thus have

$$\sum_{\alpha \in \mathcal{S}} (\deg \alpha)^{-1} N(\alpha)^{-s} = \sum_{\mathfrak{p} \in T} \sum_{\alpha \in \psi^{-1}(\mathfrak{p})} (\deg \alpha)^{-1} N(\alpha)^{-s}. \tag{2.5}$$

Now for any $\alpha \in \psi^{-1}(\mathfrak{p})$ we have $\deg \alpha = \deg \pi_\alpha = \deg \mathfrak{p}_\alpha = \deg \mathfrak{p}$. Thus $N(\alpha) = N\mathfrak{p}$. Hence the inner sum in the right-hand side of (2.5) is repeated addition of the same quantity, and the right-hand side becomes $\sum_{\mathfrak{p} \in T} N\mathfrak{p}^{-s}$. This proves the first statement of the Lemma.

To prove the second statement, we use the fact that $\mathbb{F}_{p^k}$ consists of all $\alpha \in \overline{\mathbb{F}}_p$ with $\deg \alpha$ dividing k. This gives

$$S \cap \mathbb{F}_{p^k} = \psi^{-1}(\{\mathfrak{p} \in T : \deg \mathfrak{p} \mid k\}).$$

Because $\mathcal{S} = \psi^{-1}(T)$ and the preimage of $\mathfrak{p} \in T$ contains $\deg \mathfrak{p}$ elements, the above equation gives

$$\begin{aligned} \#(S \cap \mathbb{F}_{p^k}) &= \sum_{j \mid k} j \cdot \#\{\mathfrak{p} \in T : \deg \mathfrak{p} = j\} \\ &= k \cdot \#\{\mathfrak{p} \in T : \deg \mathfrak{p} = k\} + \sum_{j \mid k, j < k} j \cdot \#\{\mathfrak{p} \in T : \deg \mathfrak{p} = j\} \end{aligned} \tag{2.6}$$

For each j there are at most p^j/j irreducible polynomials over $\mathbb{F}_p$ of degree j (because all such polynomials must split in $\mathbb{F}_{p^j}$). Since $j \mid k$, $j < k$ implies $j \le k/2$, the right-hand sum in (2.6) is $O(p^{k/2})$. ■

Proposition 2.8. *Let $\mathcal{S} \subseteq \overline{\mathbb{F}}_p$ be invariant under the action of* $\mathrm{Gal}\,(\overline{\mathbb{F}}_p/\mathbb{F}_p)$*, let P be the set of all primes in $\mathbb{F}_p[x]$, let*

$$T = \{\mathfrak{p} \in P : \mathfrak{p} = \mathfrak{p}_\alpha \text{ for some } \alpha \in \mathcal{S}\},$$

and suppose that $\delta(T)$ exists. Then $\delta(\mathcal{S}) = \delta(T)$, where the the left-hand side is the Dirichlet density for subsets of $\overline{\mathbb{F}}_p$ defined in (1.6) *and the right-hand side is the Dirichlet density for subsets of P defined in* (2.3)*. If moreover $D(T)$ exists, then $D(\mathcal{S}) = D(T)$, with the two densities being the natural densities defined in* (1.7) *and* (2.4)*, respectively.*

Proof: To prove the Dirichlet density part of the Proposition, apply the first statement of Lemma 2.7 twice, once to the numerator in the expression for $\delta(\mathcal{S})$ and once to the denominator (take $\mathcal{S} = \overline{\mathbb{F}}_p$). This gives

$$\frac{\sum_{\alpha\in\mathcal{S}}(\deg\alpha)^{-1}N(\alpha)^{-s}}{\sum_{\alpha\in\overline{\mathbb{F}}_p}(\deg\alpha)^{-1}N(\alpha)^{-s}} = \frac{\sum_{\mathfrak{p}\in T}N\mathfrak{p}^{-s}}{\sum_{\mathfrak{p}\in P}N\mathfrak{p}^{-s}}.$$

Taking limits as $s \to 1^+$ and using the existence of $\delta(T)$ completes the proof.

To show the natural density component of the Proposition, divide both sides of the second statement of Lemma 2.7 by p^k to get

$$\frac{\#(S \cap \mathbb{F}_{p^k})}{p^k} = \frac{1}{p^k}\left(k \cdot \#\{\mathfrak{p} \in T : \deg\mathfrak{p} = k\} + O(p^{k/2})\right). \tag{2.7}$$

The prime number theorem for polynomials over $\mathbb{F}_p$ [32, Theorem 2.2] states that

$$k \cdot \#\{\mathfrak{p} \in P : \deg\mathfrak{p} = k\} = p^k + O(p^{k/2}).$$

Thus the right-side of (2.7) is equal to

$$\frac{k \cdot \#\{\mathfrak{p} \in T : \deg\mathfrak{p} = k\} + O(p^{k/2})}{k \cdot \#\{\mathfrak{p} \in P : \deg\mathfrak{p} = k\} + O(p^{k/2})}.$$

Taking limits as $k \to \infty$ then proves what we want. ■

We apply Proposition 2.8 to the set $\mathcal{I}_n$, which by (2.2) satisfies

$$\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_x^n \text{ has a linear factor mod } \mathfrak{p}_\alpha\}.$$

Recall that $\mathfrak{p}_\alpha$ is the ideal generated by the minimal polynomial of α, so that each $\mathfrak{p} \in P$ is $\mathfrak{p}_\alpha$ for $\deg\mathfrak{p}$ values of α. Define

$$I_n = \{\mathfrak{p} \in P : f_x^n \text{ has a linear factor mod } \mathfrak{p}\}, \tag{2.8}$$

and note that $I_n = \{\mathfrak{p} \in P : \mathfrak{p} = \mathfrak{p}_\alpha \text{ for some } \alpha \in \mathcal{I}_n\}$. By Proposition 2.8 we have that $\delta(\mathcal{I}_n) = \delta(I_n)$ provided the latter exists, and similarly for natural densities. We then get the following immediate consequence of Corollary 2.6:

Proposition 2.9. *Suppose that $\delta(I_n)$ exists for each $n \geq 1$ and that $\lim_{n\to\infty} \delta(I_n) = 0$. Then $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$. Similarly, if $D(I_n)$ exists for each $n \geq 1$ and $\lim_{n\to\infty} D(I_n) = 0$, then $D(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$.*

With Proposition 2.9 we have succeeded in reducing our original density question to one involving the density of a set of primes in $\mathbb{F}_p[x]$. In the next section we apply the Tchebotarev Density theorem to show that $\delta(I_n)$ exists and to obtain a concrete expression for it in terms of Galois-theoretic information.

2.2 Applying the Tchebotarev Density Theorem

The Tchebotarev Density theorem for function fields gives the density of certain sets of primes in $\mathbb{F}_p[x]$. These sets can be described in terms of the Artin symbol, which we define on page 25. We must therefore show that I_n (see (2.8)) has the same density as a set of primes whose Artin symbol has a certain image. That is the main goal of this section. We close the section with a statement of Tchebotarev's theorem and apply it to give a formula for $\delta(I_n)$ in purely Galois-theoretic terms. By Proposition 2.8, this also gives a similar formula for $D(\mathcal{I}_n)$.

We use the following notation throughout this section: put $A = \mathbb{F}_p[x]$, $K = \mathbb{F}_p(x)$, and let P be the set of primes in A. Let $f_x = y^2 + x \in K[y]$, β a root in $\overline{K}$ of f_x^n, $L = K(\beta)$, and B the integral closure of A in L. We begin by giving a set $I_n' \subseteq P$ that differs from I_n by only finitely many primes and has a characterization in terms of ideal factorizations in B. This is mostly accomplished in the next Proposition, whose proof comes straight out of standard algebraic number theory.

Proposition 2.10. *Let L, K, B, and A be as above. With at most finitely many exceptions, every prime $\mathfrak{p}$ in A satisfies these properties:*

1. *$\mathfrak{p}B = \mathfrak{P}_1\mathfrak{P}_2\cdots\mathfrak{P}_r$, where the $\mathfrak{P}_i$ are distinct primes of B. For each i, denote by $d(\mathfrak{P}_i/\mathfrak{p})$ the residue class degree $[B/\mathfrak{P}_i : A/\mathfrak{p}]$.*
2. *$\overline{f} = f_1 f_2 \cdots f_r$, where $\overline{f}$ is the image of $f_x^n(y)$ in $\mathbb{F}_p[x,y]/\mathfrak{p}$, and the f_i are distinct irreducible polynomials in $\mathbb{F}_p[x,y]/\mathfrak{p}$ with $\deg f_i = d(\mathfrak{P}_i/\mathfrak{p})$.*

Proof: We present this argument in detail, although all its components can be found in any standard number theory text such as [20]. We begin by noting that $D_{L/K}(A[\beta]) = c^2 D_{L/K}(B)$ for some $c \in A$. Since $L = K(\beta)$, $A[\beta]$ is an A-module of full rank. Therefore $D_{L/K}(A[\beta]) \neq 0$, whence $c \neq 0$.

We now set $S = A - \mathfrak{p}$, which is a multiplicative set. As usual, denote $S^{-1}A$ by $A_\mathfrak{p}$ and $S^{-1}B$ by $B_\mathfrak{p}$. Since any basis for an A-module M is also a basis for the $S^{-1}A$-module $S^{-1}M$, we have

$$D_{L/K}(A_\mathfrak{p}[\beta]) = c^2 D_{L/K}(B_\mathfrak{p}).$$

If c^2 is a unit in $A_\mathfrak{p}$, it follows from a basic property of discriminants (see [20] Ch. 3) that

$$A_\mathfrak{p}[\beta] = B_\mathfrak{p}. \tag{2.9}$$

Clearly c^2 is a unit in $A_\mathfrak{p}$ as long as $\mathfrak{p}$ does not divide the principal ideal (c). Since $c \neq 0$, this holds for all but finitely many $\mathfrak{p}$ in A.

We now use the fact that

$$\mathfrak{p}B = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_1^{e_1} \iff \mathfrak{p}B_\mathfrak{p} = \mathfrak{Q}_1^{e_1} \cdots \mathfrak{Q}_r^{e_r} \tag{2.10}$$

where $\mathfrak{P}_i$ and $\mathfrak{Q}_i$ are primes of B and $B_\mathfrak{p}$, respectively, with identical residue class degrees. Since only finitely many $\mathfrak{p}$ ramify in B, we may assume that (2.9) and statement 1 of the Proposition both hold (this omits only finitely many $\mathfrak{p}$ from consideration). It follows from (2.10) and statement 1 that $\mathfrak{p}B_\mathfrak{p} = \mathfrak{Q}_1\mathfrak{Q}_2 \cdots \mathfrak{Q}_r$, where the $\mathfrak{Q}_i$ are distinct primes in $B_\mathfrak{p}$. Using (2.10) once again, we see that it now suffices to show that the degrees of the irreducible factors of $\overline{f}$ in $\mathbb{F}_p[x, y]/\mathfrak{p}$ correspond to the residue class degrees of the $\mathfrak{Q}_i$.

Let f_i be an irreducible factor of $\overline{f}$, and let γ be a root of f_i. There is a natural surjective ring homomorphism

$$A_\mathfrak{p}[\beta] \to (A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p})\,[\gamma].$$

The image of this map is isomorphic to $\mathbb{F}_{p^k}(\gamma)$, so its kernel is a maximal ideal $\mathfrak{Q}$ of $A_\mathfrak{p}[\beta] = B_\mathfrak{p}$. Clearly $\mathfrak{Q} \supseteq \mathfrak{p}A_\mathfrak{p}$, so $\mathfrak{Q} \cap A_\mathfrak{p} = \mathfrak{p}A_\mathfrak{p}$. On the other hand, by (2.9) we have

$$B_\mathfrak{p}/\mathfrak{Q} = (A_\mathfrak{p}[\beta])/\mathfrak{Q} \cong (A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p})\,[\gamma],$$

so we have $[B_\mathfrak{p}/\mathfrak{Q} : A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}] = \deg f_i$.

Hence for each irreducible factor f_i of $\overline{f}$, there is an ideal of $B_\mathfrak{p}$ lying over $\mathfrak{p}A_\mathfrak{p}$ with residue class degree $\deg f_i$. Moreover, these ideals are the only ones lying over $\mathfrak{p}A_\mathfrak{p}$, because the sum of their residue class degrees is $\sum_i \deg f_i = \deg \overline{f} = [L : K]$. Finally, because $\mathfrak{p}B_\mathfrak{p}$ is unramified, the f_i must be distinct. ■

In light of Proposition 2.10, we define

$$I_n' = \{\mathfrak{p} \in P : \mathfrak{p}B \text{ is divisible by a prime of residue class degree } 1\}. \tag{2.11}$$

Proposition 2.10 shows that I_n and I_n' differ by at most finitely many primes, and it is an easy exercise to show that both Dirichlet and natural density do not change if one alters a set by finitely many primes. We now wish to express I_n' in terms of the Artin symbol.

To define the Artin symbol, we move from considerations of the field $L = K(\beta)$ to the splitting field K_n of f_x^n over K. We show in chapter 3 that f_x^n is irreducible and separable over K (see

Proposition 3.3). Therefore K_n/K is a Galois extension, and we denote its Galois group by G_n. We engage in the standard abuse of language whereby "primes of [some extension of K]" is a shorthand for "primes in the integral closure of A in [some extension of K]."

The Artin symbol maps an unramified prime $\mathfrak{p}$ in A to a conjugacy class of G_n. To describe which conjugacy class, we state several basic results and refer the reader to [32] for proofs. For each prime $\mathfrak{p}$ of A and each $g \in G_n$, g permutes the primes of K_n above $\mathfrak{p}$. Moreover, the action of G_n on these primes is transitive [32, Proposition 9.2]. If a prime $\mathfrak{Q}$ of K_n lies above $\mathfrak{p}$, the stabilizer of $\mathfrak{Q}$ under this action, i.e. $\{g \in G_n : g(\mathfrak{Q}) = \mathfrak{Q}\}$, is called the *decomposition group* of $\mathfrak{Q}$ over $\mathfrak{p}$, denoted $Z(\mathfrak{Q}/\mathfrak{p})$. We denote by J the Galois group of $B/\mathfrak{Q}$ over $A/\mathfrak{p} \cong \mathbb{F}_{p^{\deg \mathfrak{p}}}$. If $\mathfrak{Q}^2 \nmid \mathfrak{p}$, then it is shown [32, Theorem 9.6] that

$$Z(\mathfrak{Q}/\mathfrak{p}) \cong J.$$

By the theory of finite fields, J is cyclic of order $[B/\mathfrak{Q} : A/\mathfrak{p}] = d(\mathfrak{Q}/\mathfrak{p})$, the residue class degree of $\mathfrak{Q}$. Moreover, J is generated by the Frobenius automorphism $\phi_{\mathfrak{p}}$ that maps $x \in B/\mathfrak{Q}$ to $x^{\deg \mathfrak{p}}$. We denote by $(\mathfrak{Q}, K_n/K)$ the element of $Z(\mathfrak{Q}/\mathfrak{p})$ whose image in J is $\phi_{\mathfrak{p}}$. We note that

$$|(\mathfrak{Q}, K_n/K)| = |Z(\mathfrak{Q}/\mathfrak{p})| = d(\mathfrak{Q}/\mathfrak{p}). \tag{2.12}$$

We call $(\mathfrak{Q}, K_n/K)$ the Artin symbol of $\mathfrak{Q}$; note that it is a single element of G_n. The price of pushing our definition down to $\mathfrak{p}$ is that we must replace the single element by a conjugacy class:

Proposition 2.11. *For any $g \in G_n$, $(g\mathfrak{Q}, K_n/K) = g(\mathfrak{Q}, K_n/K)g^{-1}$.*

Proof: : See [32], Proposition 9.10

From this proposition and the transitivity of the action of G_n on the primes over $\mathfrak{p}$, it follows that the set $\{(\mathfrak{Q}, K_n/K) : \mathfrak{Q} \text{ above } \mathfrak{p}\}$ fills out a conjugacy class in G_n. We denote this conjugacy class by $(\mathfrak{p}, K_n/K)$, and call it the *Artin symbol of* $\mathfrak{p}$.

Each $g \in G_n$ permutes the roots of f_x^n, and this permutation can be decomposed into disjoint cycles. Moreover, conjugation preserves the lengths of the cycles in this decomposition: if $g(\beta_1) = \beta_2$, then hgh^{-1} maps $h(\beta_1)$ to $h(\beta_2)$. Therefore the conjugacy class $(\mathfrak{p}, K_n/K)$ consists of elements with identical cycle decompositions. The next result relates the lengths of the cycles in this decomposition to the factorization of $\mathfrak{p}B$. Thus the characterization we seek of I_n' (see (2.11)) in terms of the Artin symbol is achieved.

Proposition 2.12. *Let $A = \mathbb{F}_p[x]$, $K = \mathbb{F}_p(x)$, $f_x = y^2 + x$, and $L = K(\beta)$ where β is a root of f_x^n. Let B be the integral closure of A in L, K_n the splitting field of f_x^n over K, and G_n the Galois group of K_n/K. Suppose that $\mathfrak{p}$ is an unramified prime in A with $\mathfrak{p}B = \mathfrak{P}_1 \cdots \mathfrak{P}_r$, where $\mathfrak{P}_i$ is a prime in B. Then any element of $(\mathfrak{p}, K_n/K)$ acts on the roots of $f_x^n(y)$ as a product $\sigma_1 \cdots \sigma_r$ of disjoint cycles, with σ_i having length $d(\mathfrak{P}_i/\mathfrak{p})$.*

Proof: Suppose $L = K(\beta_1)$ for some root β_1 of f_x^n. Let $\beta_2, \ldots, \beta_m$ be the other roots, so that the set of roots of f_x^n is $\mathcal{R}_n = \{\beta_1, \ldots, \beta_m\}$. Suppose $g \in (\mathfrak{p}, K_n/K)$, and let $\sigma_1 \cdots \sigma_r$ be the disjoint cycle decomposition of g as a permutation of $\mathcal{R}_n$. Fix i with $1 \leq i \leq r$, and let $|\sigma_i|$ be the length of the cycle σ_i. We show that there is a prime $\mathfrak{P}_i$ in B lying over $\mathfrak{p}$ such that $d(\mathfrak{P}_i/\mathfrak{p}) = |\sigma_i|$.

Let β_j be an element of the cycle σ_i. Since f_x^n is irreducible (Proposition 3.3), the action of G_n on $\mathcal{R}_n$ is transitive, so we may choose a $t_i \in G_n$ such that $t_i(\beta_j) = \beta_1$. Then $h_i = t_i g t_i^{-1} \in (\mathfrak{p}, K_n/K)$, and h_i maps $t_i(g^{-1}(\beta_j))$ to $t_i(\beta_j) = \beta_1$. Thus the disjoint cycle decomposition of h_i is $\sigma_1' \cdots \sigma_r'$, where $\beta_1 \in \sigma_i'$ and $|\sigma_i'| = |\sigma_i|$.

Let $\mathfrak{Q}_i$ be a prime of K_n with $(\mathfrak{Q}_i, K_n/K) = h_i$. Denote by $\mathfrak{P}_i$ the prime in B lying under $\mathfrak{Q}_i$, and note that both $\mathfrak{Q}_i$ and $\mathfrak{P}_i$ lie over $\mathfrak{p}$. By the multiplicativity of degrees in towers of field extensions we have

$$d(\mathfrak{Q}_i/\mathfrak{p}) = d(\mathfrak{Q}_i/\mathfrak{P}_i) d(\mathfrak{P}_i/\mathfrak{p}). \tag{2.13}$$

Note that by (2.12) we have $d(\mathfrak{Q}_i/\mathfrak{p}) = |h_i|$ and $d(\mathfrak{Q}_i/\mathfrak{P}_i) = |Z(\mathfrak{Q}_i/\mathfrak{P}_i)|$. Defining V to be the Galois group of K_n/L, it is shown in Proposition 9.8 of [32] that $|Z(\mathfrak{Q}_i/\mathfrak{P}_i)| = |Z(\mathfrak{Q}_i/\mathfrak{p}) \cap V| = |\langle h_i \rangle \cap V|$. From this and (2.13) we have

$$d(\mathfrak{P}_i/\mathfrak{p}) = \frac{|h_i|}{|\langle h_i \rangle \cap V|}.$$

Since $L = K(\beta_1)$, V is simply the stabilizer of β_1 in G_n. Since $\beta_1 \in \sigma_i'$, we have $h_i^{|\sigma_i'|} \in V$ and $h_i^l \notin V$ for $l < |\sigma_i'|$. Thus $\langle h_i \rangle \cap V = \left\langle h_i^{|\sigma_i'|} \right\rangle$. Also, because σ_i' is a cycle in the decomposition of h_i, we have that $|\sigma_i'|$ divides $|h_i|$, whence $\left| h_i^{|\sigma_i|} \right| = |h_i| \,/\, |\sigma_i'|$. Putting all this together yields

$$d(\mathfrak{P}_i/\mathfrak{p}) = \frac{|h_i|}{\left| \left\langle h_i^{|\sigma_i'|} \right\rangle \right|} = \frac{|h_i|}{|h_i| \,/\, |\sigma_i'|} = |\sigma_i'| = |\sigma_i|\,,$$

which completes the proof. ■

We now relate I_n' to another set of primes in A. Let $P_U \subseteq P$ be the set of primes of A unramified in K_n. We say that $(\mathfrak{p}, K_n/K)$ fixes a root of f_x^n if every g (equivalently any g) in $(\mathfrak{p}, K_n/K)$ fixes a root of f_x^n. Consider the set

$$I_n'' = \{\mathfrak{p} \in P_U : (\mathfrak{p}, K_n/K) \text{ fixes a root of } f_x^n\}. \tag{2.14}$$

It is clear from (2.11) and Proposition 2.12 that I_n'' differs from I_n' by only finitely many primes (the primes in A that ramify in K_n). Hence $\delta(I_n'') = \delta(I_n') = \delta(I_n)$ and similarly for natural density.

We are at last in a position to use the Tchebotarev Density theorem. There are two versions of the theorem, one dealing with Dirichlet density and the other with natural density. We state the Dirichlet version first, specializing to the case of importance to us.

Theorem 2.13 (Tchebotarev, first form). *Let $K = \mathbb{F}_p(x)$, $f_x = y^2 + x$, K_n/K the splitting field of f_x^n, and P_U the set of primes of K unramified in K_n. Put $G_n = \mathrm{Gal}\,(K_n/K)$, and let C be a conjugacy class in G_n. Then*

$$\delta(\{\mathfrak{p} \in P_U : (\mathfrak{p}, K_n/K) = C\}) = \frac{\#C}{\#G_n}.$$

As one might guess, the second version of Tchebotarev's theorem deals with natural density. In its simplest form it requires that the Galois extension be *geometric*, that is, that it not contain any nontrivial extension of the constant field (which is $\mathbb{F}_p$ in our case). More precisely:

Definition 2.14. *Let F be a field, and let L be an extension of $F(x)$. The algebraic closure E of F in L is called the* constant field *of L; equivalently, $E = \overline{F} \cap L$. If $E = F$ then we call the extension $L/F(x)$* geometric*.*

Theorem 2.15 (Tchebotarev, second form). *Let $K = \mathbb{F}_p(x)$, $f_x = y^2 + x$, K_n/K the splitting field of f_x^n, and P_U the set of primes of K unramified in K_n. Put $G_n = \mathrm{Gal}\,(K_n/K)$, and let C be a conjugacy class in G_n. Assume that K_n/K is geometric. Then*

$$\#\{\mathfrak{p} \in P_U : \deg \mathfrak{p} = k \text{ and } (\mathfrak{p}, K_n/K) = C\} = \frac{\#C}{\#G_n}\frac{p^k}{k} + O\left(\frac{p^{k/2}}{k}\right).$$

For partial proofs of these theorems and references to complete treatments, see [32, Ch. 9]. We make two remarks. First, Theorem 2.15 implies that

$$D(\{\mathfrak{p} \in P_U : (\mathfrak{p}, K_n/K) = C\}) = \frac{\#C}{\#G_n}. \tag{2.15}$$

To prove this, all that is needed is a slight alteration of the argument in the proof of Proposition 2.8 where we showed that $D(\mathcal{S}) = D(T)$. Second, it is a fact of the universe that if the extension K_n/K is not geometric, then the natural density in (2.15) does not exist in general. This follows from the general natural density form of Tchebotarev's theorem (see [12][1]). Unfortunately, it is no easy matter to find the exact constant field in a complicated extension. In the next chapter, we show:

Proposition 2.16. *Let $K = F(x)$, where F is a field of characteristic 0 or $p \equiv 3 \bmod 4$. Put $f_x = y^2 + x \in K[y]$, and let K_n/K the splitting field of f_x^n. Then K_n/K is geometric.*

See Corollary 3.40 for a proof. We make the following conjecture, which has so far eluded proof:

Conjecture 2.17. *Proposition 2.16 holds for all F of characteristic $\neq 2$.*

We now complete the main task of this section, and indeed this chapter:

[1] This article by Fried is rather opaque. It seems to me there is an opening for a nice expository piece on the nuances of the general theorem.

Theorem 2.18. *Let $K = \mathbb{F}_p(x)$, $f_x = y^2 + x$, and K_n/K the splitting field of f_x^n. Set $G_n = \mathrm{Gal}\,(K_n/K)$, and let*

$$\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset\},$$

where $f_\alpha = x^2 + \alpha$. Then

$$\delta(\mathcal{I}_n) = \frac{1}{\#G_n} \#\left\{g \in G_n : g \text{ fixes at least one root of } f_x^n\right\}. \tag{2.16}$$

If in addition K_n/K is geometric, then $D(\mathcal{I}_n)$ exists and also equals the expression in (2.16).

Proof: Recall that we defined $I_n'' = \{\mathfrak{p} \in P_U : (\mathfrak{p}, K_n/K)$ fixes a root of $f_x^n\}$, and we showed (see the discussion after (2.14) and (2.11)) that $\delta(I_n'') = \delta(I_n)$, where I_n is a set satisfying $\delta(I_n) = \delta(\mathcal{I}_n)$ (see Proposition 2.8 and equation 2.8).

Using this and Theorem 2.13, and denoting by $\mathcal{C}$ the collection of conjugacy classes of G_n each of whose elements fixes at least one root of f_x^n, we have

$$\delta(\mathcal{I}_n) = \delta(I_n'') = \sum_{C \in \mathcal{C}} \frac{\#C}{\#G_n} = \frac{1}{\#G_n} \sum_{C \in \mathcal{C}} \#C.$$

Now $g \in G_n$ fixes a root of f_x^n if and only if every element of its conjugacy class does the same. Thus $g \in \bigcup_{C \in \mathcal{C}} C$ if and only if g fixes at least one root of f_x^n. Equation (2.16) follows.

For the natural density case, the same chain of reasoning implies that $D(\mathcal{I}_n) = D(I_n'')$, and by equation (2.15) and the geometricity of K_n/K it follows that $D(I_n'')$ exists. Therefore $D(\mathcal{I}_n)$ exists, whence it must equal $\delta(\mathcal{I}_n)$. ■

Theorem 2.18 provides a key ingredient in the proof of Theorem 1.7, which is our main result. Corollary 2.6 states that to show $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$, it is enough to show that $\lim_{n\to\infty} \delta(\mathcal{I}_n) = 0$. Theorem 2.18 in turn reduces this problem to one involving only the structure of certain Galois groups, which we examine in Chapter 3. However, Theorem 2.18 is of interest in its own right.

Consider the case where K_n/K is geometric, which is provably true in many cases (see Corollary 3.39 and the remarks following). Theorem 2.18 then gives the natural density of $\mathcal{I}_n$, i.e.

$$\lim_{k\to\infty} \frac{\#\mathcal{I}_n \cap \mathbb{F}_{p^k}}{p^k}.$$

Thus for large k, the proportion of $\alpha \in \mathbb{F}_{p^k}$ such that 0 has at least one "simple" nth preimage under f_α stabilizes, and its limit is given by purely Galois-theoretic information. By "simple" we mean that it is not necessary to pass to an extension of $\mathbb{F}_p(\alpha)$. By the prime number theorem for $\mathbb{F}_p[x]$ [32, Theorem 2.2], the above proportion is approximately equal, for large k, to the proportion of $\alpha \in \mathbb{F}_{p^k}$ that have an nth preimage under f_α in $\mathbb{F}_{p^k}$. These proportions are easy to compute, and

doing so shows that k does not have to be very large to get results that are extremely close to the predicted limits (see the table on page 89).

We give an illustration of this discussion in the example below.

Example 2.19. Consider the case $n = 2$. We prove in Chapter 3 that K_2/K is geometric (for all $p \neq 2$). Let us label the roots of $f_x^2 = (y^2 + x)^2 + x$ as follows:

$$\begin{aligned} \sqrt{-x + \sqrt{-x}} &\longleftrightarrow 1 \\ -\sqrt{-x + \sqrt{-x}} &\longleftrightarrow 2 \\ \sqrt{-x - \sqrt{-x}} &\longleftrightarrow 3 \\ -\sqrt{-x - \sqrt{-x}} &\longleftrightarrow 4 \end{aligned}$$

It follows from Theorem 3.2 that under this labeling G_2 is a subgroup of S_4 of order 8 that contains $\{e, (1\ \ 2), (3\ \ 4), (1\ \ 2)(3\ \ 4)\}$ as well as four elements that interchange the sets $\{1, 2\}, \{3, 4\}$ and therefore have no fixed points. Let k be large, and for $\alpha \in \mathbb{F}_{p^k}$, let i_α be the number of 2nd preimages of 0 (under f_α) in $\mathbb{F}_{p^k}$. Since $e, (1\ \ 2)$, and $(3\ \ 4)$ all have fixed points, Theorem 2.18 shows that $i_\alpha > 0$ for about 3/8 of the $\alpha \in \mathbb{F}_{p^k}$. In fact, we obtain more information: $i_\alpha = 4$ for about 1/8 of the $\alpha \in \mathbb{F}_{p^k}$, and $i_\alpha = 2$ for about 1/4 of the $\alpha \in \mathbb{F}_{p^k}$.

Chapter 3

The Structure of G_n

In Chapter 2 we gave a method for proving our main result (Theorem 1.7), which states that $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$ for $p \neq 2$. Corollary 2.6 reduced the problem to showing that $\lim_{n\to\infty} \delta(\mathcal{I}_n) = 0$, where $\mathcal{I}_n$ is the set of $\alpha \in \overline{\mathbb{F}}_p$ such that 0 has an nth preimage in $\mathbb{F}_p(\alpha)$ under iteration of $f_\alpha = x^2 + \alpha$, i.e.,

$$\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset\}.$$

Theorem 2.18 then reduced the determination of $\delta(\mathcal{I}_n)$ to properties of the Galois group G_n of the splitting field over $K = \mathbb{F}_p(x)$ of the nth iterate of $y^2 + x$, specifically

$$\delta(\mathcal{I}_n) = \frac{1}{\#G_n} \#\left\{g \in G_n : g \text{ fixes at least one root of } f_x^n\right\}. \tag{3.1}$$

In this chapter we illuminate enough of the structure of G_n to make significant progress towards finding $\lim_{n\to\infty} \delta(\mathcal{I}_n)$.

In order to prove certain results that we need (particularly Corollary 3.39), it is necessary to work in greater generality than the previous two chapters. We let F be any field of characteristic $\neq 2$, and set $K = F(x)$ and $A = F[x]$. Put $f(y) = y^2 + x \in K[y]$ (we drop the subscript x of the previous chapter), and denote the nth iterate of this polynomial by f^n. Let K_n be the splitting field of f^n over K, and let $G_n = \mathrm{Gal}\,(K_n/K)$. *This notation is in force throughout this chapter* (although we restate it occasionally when stating important theorems).

This chapter contains an examination of the structure of G_n, which we approach through a study of the subgroups $H_n = \mathrm{Gal}\,(K_n/K_{n-1})$. This is a natural object of study, since G_n has a composition series with quotients $H_m, m \leq n$. It turns out that K_n is obtained from K_{n-1} by adjoining the square roots of 2^{n-1} different elements (see (3.2)). Thus H_n is isomorphic to an m_n-fold direct product of $\mathbb{Z}/2\mathbb{Z}$, where $m_n \leq 2^{n-1}$. If $m_n = 2^{n-1}$, we say H_n is *maximal*. We make the following conjecture:

Conjecture 3.1. *For arbitrary F of characteristic $\neq 2$, H_n is maximal for all n.*

If Conjecture 3.1 holds, then one can give a formula for the quantity in (3.1). Indeed, we show in Chapter 5 using the theory of branching processes (see Corollary 5.11 and the comments on the preceding page), that Conjecture 3.1 implies that $1 - \lim_{k\to\infty} \delta(\mathcal{I}_n)$ is equal to the nth iterate of $\frac{1}{2} + \frac{1}{2}z^2$ evaluated at $z = 0$. One can easily show that this approaches 1 as n goes to infinity, which would prove Theorem 1.7.

We remark that in general G_n is isomorphic to a subgroup of each of the following groups (which are themselves isomorphic): the n-fold wreath product of $\mathbb{Z}/2\mathbb{Z}$, a Sylow 2-subgroup of S_{2^n}, and the automorphism group of the complete binary tree of height n. If H_m is maximal for each $m \leq n$, then G_n is isomorphic to the full group in each of these three settings. Thus Conjecture 3.1 would give very explicit information about G_n for all n.

Unfortunately, our attempts to prove Conjecture 3.1 in full generality encounter serious obstacles, which we discuss immediately following Theorem 3.38. However, we do show the following:

Theorem 3.2. *For arbitrary F of characteristic $\neq 2$, H_n is maximal for squarefree n. Moreover, if F has characteristic 0 or characteristic $p \equiv 3 \bmod 4$ then H_n is maximal for all n.*

We note that the characteristic 0 case of Theorem 3.2 has already been proved by Odoni [26], who in fact handles a much larger class of iterative-type towers. The proof of Theorem 3.2 is the main aim of this chapter. In subsequent chapters we parlay Theorem 3.2 into a proof of Theorem 1.7, though using means far more complicated than would be necessary if Conjecture 3.1 could be proven.

The methods that lead to a proof of Theorem 3.2 also allow us to show that if H_m is maximal for all $m \leq n$, then K_n/K is geometric (Corollary 3.39). Using Theorem 2.18 and Theorem 3.2, this allows us to show that $D(\mathcal{I}_n)$ exists and equals the expression in (3.1) in the case $\operatorname{char} F = 0$ and $\operatorname{char} F = p \equiv 3 \bmod 4$.

This chapter unfolds in five parts. First, we give some elementary results on the polynomials f^n and the sequence $\{f^n(0) : n = 1, 2, 3, \ldots\}$. In the second part, we prove that a specific permutation of the roots of f^n must be in G_n, a result that plays a vital role in subsequent chapters. In the third and fourth parts we give a series of results that show that the maximality of H_n depends on whether a certain element of $F[x]$ is a square. This quickly leads to a proof of the first two statements of Theorem 3.2. To handle the case $\operatorname{char} F = p \equiv 3 \bmod 4$ in Theorem 3.2, we give in section five an adaptation of a clever trick of Stoll [35] originally used to show that Galois groups of iterates of $x^2 + 1$ are maximal over $\mathbb{Q}$.

3.1 Preliminaries

We let F be any field of characteristic $\neq 2$, and set $K = F(x)$ and $A = F[x]$. Put $f(y) = y^2 + x \in K[y]$, and denote the nth iterate of this polynomial by f^n. In this section we give some properties of f^n. Our first one is fundamental:

Proposition 3.3. *For each n, $f^n(y)$ is irreducible over K.*

Proof: We show by induction that each f^n is a monic Eisenstein polynomial with respect to the prime $(x) \subset A$. Clearly f is a monic Eisentein polynomial with respect to (x). Now suppose the same is true of f^{n-1}, i.e. $f^{n-1}(y) = y^d + x(a_{d-1}y^{d-1} + \cdots + a_1 y + a_0)$ with $a_i \in A$ and x not dividing a_0. Then

$$\begin{aligned} f^n &= (f^{n-1}(y))^2 - x \\ &= y^{2d} + xy^d(a_{d-1}y^{d-1} + \cdots + a_0) + x^2(a_{d-1}y^{d-1} + \cdots + a_0)^2 - x \\ &= y^{2d} + x\left[y^d(a_{d-1}y^{d-1} + \cdots + a_0) + x(a_{d-1}y^{d-1} + \cdots + a_0)^2 - 1\right]. \end{aligned}$$

The constant term of the polynomial in the brackets is $xa_0^2 - 1$, and since x clearly does not divide this, we have that f^n is a monic Eisenstein polynomial with respect to (x). ■

Note that it is possible to show that many quadratic polynomials over various fields have the property proved for f above: every iterate is irreducible [2, 3].

Since $\deg f^n = 2^n$, which is not divisible by p, Proposition 3.3 implies that f^n is also separable (any inseparable irreducible polynomial must have degree a power of the characteristic). Thus G_n acts as a transitive group of permutations on the set of 2^n roots of f^n, which we denote

$$\mathcal{R}_n = \{\text{roots of } f^n\}.$$

These facts play important roles in the rest of the chapter.

It is an easy matter to write out $\mathcal{R}_n$ explicitly in terms of $\mathcal{R}_{n-1}$. Suppose $\mathcal{R}_{n-1} = \{\beta_1, \ldots, \beta_{2^{n-1}}\}$, and note that $f^n(y) = f^{n-1}(f(y)) = f^{n-1}(y^2 + x)$. Thus for $1 \leq i \leq 2^{n-1}$ the roots of $y^2 + x = \beta_i$ are in $\mathcal{R}_n$, whence

$$\mathcal{R}_n = \left\{\pm\sqrt{-x + \beta_1}, \ldots, \pm\sqrt{-x + \beta_{2^{n-1}}}\right\}. \tag{3.2}$$

This shows that K_n is obtained from K_{n-1} by adjoining the square roots of 2^{n-1} elements, whence H_n is isomorphic to a direct product of $\mathbb{Z}/2\mathbb{Z}$ of rank at most 2^{n-1}. Since $G_i/H_i \sim G_{i-1}$ for all $1 \leq i \leq n$, it follows that $|G_n| = \Pi_{i=1}^n |H_i|$, which proves the following:

Proposition 3.4. *G_n is a 2-group.*

Though we will not need it in the sequel, we can use induction to come up with an explicit description of $\mathcal{R}_n$:

$$\mathcal{R}_n = \left\{ \pm\sqrt{-x \pm \sqrt{-x \pm \cdots \pm \sqrt{-x}}} \right\},$$

where there are n root signs.

Another useful property of f^n is this:

Proposition 3.5. *For each n, f^n is a polynomial in y^2.*

Proof: Clear for $n = 1$. For $n \geq 2$, first note that by the definition of f^n, we have $f^n(y) = f^{n-1}(f(y)) = f^{n-1}(y^2 - x)$. By induction, f^n is a polynomial in y^2. ■

We now establish some relations among elements of $A = \mathbb{F}_p[x]$ of the form $f^n(0)$. These elements turn out to encode important information about the groups H_n. We give them a more succinct name:

Definition 3.6. *For each n, let $p_n = f^n(0) \in A$.*

Clearly $p_1 = x$. Moreover, we have $p_n = f^n(0) = (f^{n-1}(0))^2 + x = (p_{n-1})^2 + x$, and more generally

$$p_n = f^{n-i}(p_i) \tag{3.3}$$

for all $n \geq 2$ and $1 \leq i \leq n-1$.

The polynomial $\sigma^n(x) - x$ ($\sigma \in F[x]$) and p_n have some similarities. The roots of $\sigma^n(x) - x$ are all points periodic under σ with period dividing n; they have been studied by Morton and Patel in [25]. On the other hand, $c \in \overline{F}$ is a root of p_n if and only if $f_c^n(0) = 0$, where $f_c = y^2 + c$. Thus the roots of p_n consist of all $c \in \overline{F}$ such that the period of 0 under iteration of f_c divides n. This implies that if an irreducible $q \in A$ divides p_n, then it must also divide p_{mn} for any $m \geq 1$. The following fundamental property of p_n gives us much more:

Proposition 3.7. *Let $q \in A$ be irreducible, and suppose $\mathrm{ord}_q(p_n) = e \geq 1$. Then for every $m \geq 1$, we have $\mathrm{ord}_q(p_{mn}) = e$.*

Proof: Induction on m. The proposition is trivial if $m = 1$. Now suppose inductively that $\mathrm{ord}_q(p_{(m-1)n}) = e$. By (3.3) we have $p_{mn} = f^{(m-1)n}(p_n)$. Proposition 3.5 tells us $f^{(m-1)n}(y)$ is a polynomial in y^2, so we can write

$$f^{(m-1)n}(y) = y^2 g(y) + f^{(m-1)n}(0) = y^2 g(y) + p_{(m-1)n},$$

for some $g(y) \in K[y]$. Hence putting $y = p_n$ we have

$$p_{mn} = p_n^2 g(p_n) + p_{(m-1)n}.$$

Now $\mathrm{ord}_q\left[(p_n)^2(g(p_n))\right] \geq 2e$, and by our inductive hypothesis $\mathrm{ord}_q(p_{(m-1)n}) = e$. Since $e \geq 1$, the first summand vanishes to higher order at q than the second, so we conclude $\mathrm{ord}_q(p_{mn}) = e$. ■

Since $p_1 = x$, Proposition 3.7 shows that $\mathrm{ord}_x(p_m) = 1$ for all $m \geq 1$. This gives us a useful corollary:

Corollary 3.8. *For each n, p_n is not a square in K.*

Here are the first few elements of the sequence $\{p_n\}$, along with their factorizations in $\mathbb{Z}[x]$:

n	p_n	factorization in $\mathbb{Z}[x]$
1	x	x
2	$x^2 + x$	$x(x+1)$
3	$x^4 + 2x^3 + x^2 + x$	$x(x^3 + 2x^2 + x + 1)$
4	$x^8 + 4x^7 + 6x^6 + 6x^5 + 5x^4 + 2x^3 + x^2 + x$	$x(x+1)(x^6 + 3x^5 + 3x^4 + 3x^3 + 2x^2 + 1)$

We close this section of preliminary results with a computation of $\mathrm{Disc}\, f^n$. The answer turns out to be essentially a product of powers of p_i for $1 \leq i \leq n$.

First we set some notation: let $\beta_1, \ldots, \beta_{2^{n-1}}$ be the roots of f^{n-1}, and let $\pm\alpha_i$ be the roots of $y^2 + x = \beta_i$. Thus the roots of f^n are $\pm\alpha_1, \ldots, \pm\alpha_{2^{n-1}}$. Order these roots as follows:

$$\alpha_1, -\alpha_1, \alpha_2, -\alpha_2, \ldots, \alpha_{2^{n-1}}, -\alpha_{2^{n-1}}.$$

By definition $\mathrm{Disc}\,(f^n)$ is the product of all of the $(r_k - r_l)^2$, where r_k and r_l are respectively the kth and lth roots in the above list and $k < l$. Thus $\mathrm{Disc}\,(f^n)$ has a factor of $(\alpha_i - (-\alpha_i))^2 = (2\alpha_i)^2$ for $i = 1, \ldots, 2^{n-1}$, and a factor of

$$(\alpha_i - \alpha_j)^2(\alpha_i + \alpha_j)^2(-\alpha_i - \alpha_j)^2(-\alpha_i + \alpha_j)^2 = (\alpha_i^2 - \alpha_j^2)^4$$

for all $1 \leq i < j \leq 2^{n-1}$. Thus we have

$$\begin{aligned}
\mathrm{Disc}\,(f^n) &= \left(\prod_{i=1}^{2^{n-1}} 4\alpha_i^2\right)\left(\prod_{1\leq i<j\leq 2^{n-1}} (\alpha_i^2 - \alpha_j^2)\right)^4 \\
&= 2^{2^n}\left(\prod_{i=1}^{2^{n-1}} \alpha_i^2\right)\left(\prod_{i<j}(x + \beta_i - (x + \beta_j))\right)^4 \\
&= 2^{2^n}\left(\prod_{i=1}^{2^{n-1}} \alpha_i^2\right)\mathrm{Disc}\,(f^{n-1})^2. \qquad (3.4)
\end{aligned}$$

Note that $\prod_{i=1}^{2^{n-1}} \alpha_i^2$ is just the product of all of the roots of f^n, which is $\pm f^n(0)$; since f^n has even degree the sign is positive. By definition $p_n = f^n(0)$, so from equation (3.4) we have

$$\operatorname{Disc} f^n = 2^{2^n} p_n \operatorname{Disc} (f^{n-1})^2. \tag{3.5}$$

Proposition 3.9. *For each n we have*

$$\begin{aligned} \operatorname{Disc} (f^n) &= 2^{2m_n} \prod_{i=1}^{n} (p_i)^{2^{n-i}} \\ &= 2^{2m_n} p_n \left(\prod_{i=1}^{n-1} (p_i)^{2^{n-i}} \right)^2 \end{aligned} \tag{3.6}$$

where $m_1 = 1$ and $m_n = 2^{n-1} + 2m_{n-1}$ for $n > 1$.

Proof: We induct on n. The proposition is clear for $n = 1$. Suppose we have

$$\operatorname{Disc} (f^{n-1}) = 2^{2m_{n-1}} \prod_{i=1}^{n-1} (p_i)^{2^{n-1-i}}.$$

Then by (3.5),

$$\begin{aligned} \operatorname{Disc} (f^n) &= 2^{2^n} p_n \operatorname{Disc} (f^{n-1})^2 \\ &= 2^{2^n + 4m_{n-1}} p_n \prod_{i=1}^{n-1} (p_i)^{2 \cdot 2^{n-1-i}} \\ &= 2^{2(2^{n-1} + 2m_{n-1})} \prod_{i=1}^{n} (p_i)^{2^{n-i}}. \end{aligned}$$

■

Finally we have a corollary that pertains to the structure of G_n. We say that G_n is *alternating* if it is comprised entirely of even permutations of $\mathcal{R}_n$.

Corollary 3.10. *For each n, G_n is not alternating.*

Proof: It is a standard fact from algebra (see e.g. [10, Proposition 14.34]) that the Galois group of the splitting field of a separable polynomial is alternating if and only if the discriminant of the polynomial is a square in the base field. From Proposition 3.9 we see that the squarefree part of $\operatorname{Disc} f^n$ is p_n. However by Corollary 3.8, p_n is not a square in K for any n. ■

We return to the discussion of p_n in section 3.4

3.2 On the Center of G_n

As in the previous section, let $\operatorname{char} F \neq 2, K = F(x)$, $f = y^2 + x$, and f^n be the nth iterate of f. Let K_n be the splitting field of f^n over K, and put $G_n = \operatorname{Gal}(K_n/K)$. In this section we consider G_n as a group of permutations on the roots $\mathcal{R}_n$ of f^n. We show that each element of G_n actually induces a permutation on a collection of 2-element subsets of $\mathcal{R}_n$. This information allows us to show that the center of G_n must contain a particular permutation of $\mathcal{R}_n$. From this it follows that H_n is nontrivial, which is the main result of this section.

We begin with a bit of standard theory of permutation groups (see e.g. [40]).

Definition 3.11. *Let G be a group acting on a set S. A* block *of G is a subset Δ of S with the property that for each $\sigma \in G$, either $\sigma(\Delta) = \Delta$ or $\sigma(\Delta) \cap \Delta = \emptyset$. If Δ is a block of G, then we call $\{\sigma(\Delta) \mid \sigma \in G\}$ a* complete block system (CBS) *of G.*

Example 3.12. Suppose that $F = \mathbb{F}_p, p \neq 2$, and consider the group G_2 described in Example 2.19. One can show that

$$G_2 \cong \{e, (1\ \ 2), (3\ \ 4), (1\ \ 2)(3\ \ 4), (1\ \ 3\ \ 2\ \ 4), (1\ \ 4\ \ 2\ \ 3), (1\ \ 3)(2\ \ 4), (1\ \ 4)(2\ \ 3)\}.$$

The set $\{1, 2\}$ is a block of G_2 since either $\sigma(\{1, 2\}) = \{1, 2\}$ or $\sigma(\{1, 2\}) = \{3, 4\}$ for any $\sigma \in G_2$. The collection $\{\{1, 2\}, \{3, 4\}\}$ is a CBS for G. ■

If G has a CBS, then all its elements must permute the blocks of the CBS. Note that if G is transitive, then a CBS must be a partition of S into blocks of equal numbers of elements. In fact, if G is transitive then a partition of S is a CBS for G if and only if each $\sigma \in G$ permutes the subsets belonging to the partition.

We know as a consequence of Proposition 3.3 that G_n is transitive. We now find a CBS for G_n consisting of two-element sets. As noted at the beginning of Section 3.1, if $\mathcal{R}_{n-1} = \{\beta_1, \ldots, \beta_{2^{n-1}}\}$ then

$$\mathcal{R}_n = \left\{\pm\sqrt{-x + \beta_1}, \ldots, \pm\sqrt{-x + \beta_{2^{n-1}}}\right\}.$$

Set $\Gamma_i = \{\pm\sqrt{-x + \beta_i}\}$. Any $\sigma \in G_n$ must permute the roots of f^{n-1}, whence it must permute the collection

$$\mathfrak{C} = \{\Gamma_i\}_{i=1}^{2^{n-1}} \tag{3.7}$$

of subsets of $\mathcal{R}_n$. It follows that $\mathfrak{C}$ is a CBS for G_n. In the case $n = 2$, $\mathfrak{C}$ is the CBS given in Example 3.12.

Definition 3.13. *Let G be a group acting transitively on a set S, and let $\mathfrak{D}$ be a CBS for G. We call a permutation of S whose orbits are precisely the subsets belonging to $\mathfrak{D}$ a* permutation associated to $\mathfrak{D}$.

We wish to analyze permutation groups G (acting on a set S) that possess a CBS of two-element subsets. An important tool in this analysis is a one-to-one correspondence between such CBSs and certain permutation of S.

Note that a permutation associated to $\mathfrak{D}$ must act as a single cycle on each subset belonging to $\mathfrak{D}$. In the case where $\mathfrak{D}$ is composed of two-element subsets, there is only one such cycle for each subset (the one interchanging the two elements). Thus there is only one permutation associated to $\mathfrak{D}$, and we call this *the* permutation associated to $\mathfrak{D}$.

Proposition 3.14. *Let G be a group acting transitively on a set S, and let $\mathfrak{D}$ be a CBS for G composed of two-element subsets. Let δ be the permutation associated to $\mathfrak{D}$. Then*

1. *δ fixes no elements of S.*
2. *$|\delta| = 2$*
3. *$\sigma\delta = \delta\sigma$ for every $\sigma \in G$.*
4. *$\tau\delta = \delta\tau$ for every permutation τ of S (not necessarily in G) that permutes the subsets belonging to $\mathfrak{D}$.*

Moreover, if δ is a permutation of S satisfying 1, 2, and 3 above, then there is a unique CBS $\mathfrak{D}$ for G composed of two-element subsets such that δ is the permutation associated to $\mathfrak{D}$.

Proof: Property 1 follows from the definition of δ and property 2 follows from the fact that $\mathfrak{D}$ is composed of two-element sets. We now prove property 4, which implies property 3. Let δ be a permutation of S that permutes $\mathfrak{D}$, and consider the permutation $\tau\delta\tau^{-1}$. Since $\delta(\Delta) = \Delta$ for each $\Delta \in \mathfrak{D}$ and τ permutes $\mathfrak{D}$, $\tau\delta\tau^{-1}$ maps Δ to itself for all $\Delta \in \mathfrak{D}$. Moreover, $\tau\delta\tau^{-1}$ has no fixed points in S because δ has no fixed points in S. Thus the orbits of $\tau\delta\tau^{-1}$ are precisely the subsets belonging to $\mathfrak{D}$. By the uniqueness of δ we then have $\tau\delta\tau^{-1} = \delta$.

To prove the final statement of the proposition, let $\mathfrak{D}$ be the partition of S given by the orbits of δ. Properties 1 and 2 imply that each orbit of δ has two elements. Let $\Delta = \{s, \delta(s)\}$ be a subset belonging to $\mathfrak{D}$, and let $\sigma \in G$. By property 3, we have

$$\sigma(\Delta) = \{\sigma(s), \sigma\delta(s)\} = \{\sigma(s), \delta\sigma(s)\},$$

which is again an orbit of δ. Thus every element of G permutes $\mathfrak{D}$, whence $\mathfrak{D}$ is a CBS for G. Clearly δ is the permutation associated to $\mathfrak{D}$. ■

Proposition 3.14 shows that there is a one-to-one correspondence between CBSs of G consisting of two-element sets and permutations of S that satisfy conditions 1, 2, and 3 in the Proposition. To denote this correspondence we use the phrase "associated to"; thus we speak of the CBS associated to a permutation as well as the permutation associated to a CBS.

We now turn our attention to the study of the center of a permutation group. The following proposition gives us some information about the cycle decomposition of elements in the center of such a group.

Proposition 3.15. *Let G act transitively on a set S and let $\sigma \in Z(G)$, the center of G. Then σ is a product of k disjoint m/k-cycles, where $k \mid m$.*

Proof: Write $\sigma = \sigma_1\sigma_2\cdots\sigma_k$, where the σ_i are disjoint cycles. Choose $i_1 \neq i_2$ with $1 \leq i_1, i_2 \leq k$, and let $s_1 \in \sigma_{i_1}$ and $s_2 \in \sigma_{i_2}$. Let $|\sigma_{i_1}| = m_1$ and $|\sigma_{i_2}| = m_2$. Note that $\sigma^{m_1}(s_1) = s_1$ and $\sigma^{m_2}(s_2) = s_2$.

Since G is transitive, we may choose $\tau \in G$ such that $\tau(s_2) = s_1$. Then

$$(\tau\sigma\tau^{-1})^{m_2}(s_1) = \tau\sigma^{m_2}\tau^{-1}(\tau(s_2)) = \tau\sigma^{m_2}(s_2) = \tau(s_2) = s_1. \tag{3.8}$$

However, since $\sigma \in Z(G)$ we have $\sigma^{m_2} = (\tau\sigma\tau^{-1})^{m_2}$, so (3.8) gives $\sigma^{m_2}(s_1) = s_1$. The only powers of σ that fix s_1 are multiples of m_1, so we conlcude that $m_1 \mid m_2$. By choosing $\tau \in G$ with $\tau(s_1) = s_2$ and adapting slightly the above argument, we get $m_2 \mid m_1$. We thus have that $m_2 = m_1$, and since i_1 and i_2 were arbitrary it follows that all the σ_i have the same length. Thus $k \mid m$ and the proposition is proved. ■

The next Proposition uses Proposition 3.15 to glean information about $Z(G_n)$.

Proposition 3.16. *There exists $\omega \in Z(G_n)$ that is a product of 2^{n-1} disjoint transpositions. In particular, ω satisfies conditions 1, 2, and 3 of Proposition 3.14.*

Proof: By Proposition 3.4, we have that G_n is a 2-group. It is a standard result in elementary group theory that the center of a p-group is nontrivial (see e.g. [18, page 122]). Thus we may choose $z \in Z(G_n)$ different from the identity.

Since G_n acts on the set $\mathcal{R}_n$, which has 2^n elements, Proposition 3.15 shows that z is a product of 2^m disjoint 2^{n-m}-cycles for some m. Since z is not the identity, $m < n$. Note that squaring a $2k$-cycle gives a product of two disjoint k-cycles. It follows that z^{n-m-1} is a product of 2^{n-1} transpositions. The second sentence of the Proposition is obvious. ■

Propositions 3.14 and 3.16 together show that there is a CBS $\mathcal{Z}$ of G_n whose associated permutation is ω, which lies in G_n. Recall the CBS $\mathfrak{C}$ of G_n, defined in (3.7). Let γ be the permutation

associated to $\mathfrak{C}$. We wish to show that in fact $\mathcal{Z} = \mathfrak{C}$, which is equivalent to $\omega = \gamma$. This implies that $\gamma \in G_n$, which is a valuable piece of information.

Suppose that $\omega \neq \gamma$. Our approach is to show that this implies G_n is alternating, contradicting Corollary 3.10. Doing so requires several intermediary steps and occupies the remainder of this section.

Consider a permutation group G with a CBS $\mathfrak{D}$ composed of two-element subsets. Our first step is to find a simple criterion for an element of G to have even sign. Suppose $\sigma \in G$ induces a permutation $\overline{\sigma}$ on $\mathfrak{D}$, and write $\overline{\sigma}$ as a product of disjoint cycles:

$$\overline{\sigma} = C_1 C_2 \cdots C_k.$$

We may write $\sigma = \tau_1 \tau_2 \cdots \tau_k$, where τ_i is a permutation of S (not necessarily in G) that induces C_i on $\mathfrak{D}$ and leaves fixed all points of S not contained in some element of C_i. We study under what conditions the τ_i have even sign, and from this derive a criterion for σ to have even sign.

Lemma 3.17. *Let $\mathfrak{D}$ be a partition of S into two-element subsets. Let*

$$C = (\Delta_{i_1} \quad \Delta_{i_2} \quad \cdots \quad \Delta_{i_m})$$

be a cycle of elements of $\mathfrak{D}$, and set $\Sigma = \Delta_{i_1} \cup \Delta_{i_2} \cup \cdots \cup \Delta_{i_m}$. Let τ be a permutation of S that induces the permutation C on $\mathfrak{D}$ and satisfies $\tau(u) = u$ for all $u \notin \Sigma$. Let $s \in \Sigma$. Then either

1. *τ is a product of two disjoint m-cycles (if $\tau^m(s) = s$) or*
2. *τ is a $2m$-cycle (if $\tau^m(s) \neq s$)*

In particular, τ is even if and only if $\tau^m(s) = s$.

Proof: Since τ fixes all $u \notin \Sigma$, we need only consider its action on Σ. For any j and any $k < m$, we have $\tau^k(\Delta_{i_j}) \neq \Delta_{i_j}$. Thus the orbit of any element of Σ contains at least m elements. Since Σ contains only $2m$ elements, we have that either τ is two disjoint m-cycles or a $2m$-cycle. The former is the case precisely when every element of Σ has an orbit of length m, and the latter when every element has an orbit of length $2m$. But $\tau^m(s) = s$ if and only if the orbit of s has m elements, and $\tau^m(s) \neq s$ if and only if the orbit of s has $2r$ elements. ■

We now introduce a set that indexes the two-element blocks and examine how the action of a permutation σ on this set reflects the signature of σ.

Proposition 3.18. *Let G be a group acting transitively on a set S with $2m$ elements. Suppose that m is even and $\mathfrak{D} = \{\Delta_1, \Delta_2, \ldots, \Delta_m\}$ is a CBS for G with $\#\Delta_i = 2$ for $1 \leq i \leq m$. Let E be any*

set consisting of exactly one element from each Δ_i. Then $\sigma \in G$ is even if and only if $\#\{E \cap \sigma(E)\}$ is even.

Remark: The following proof can easily be adapted to show that for m arbitrary, σ is even if and only if $m + \#\{E \cap \sigma(E)\}$ is even.

Proof: Write $E = \{e_i\}_{i=1,\ldots,m}$. Let δ be the permutation associated to $\mathfrak{D}$, and note that

$$e \in E \Longleftrightarrow \delta(e) \notin E$$

.

Let $\sigma \in G$. Denote by $\bar{\sigma}$ the induced permutation of σ on $\mathfrak{D}$, and write $\bar{\sigma}$ as a product of disjoint cycles: $\bar{\sigma} = C_1 C_2 \cdots C_l$. Each C_k can be written $(\Delta_{i_1} \ \Delta_{i_2} \ \cdots \ \Delta_{i_{m_k}})$ for some m_k. Clearly $\sum_k m_k = m$.

Let $\Sigma_k = \Delta_{i_1} \cup \Delta_{i_2} \cup \cdots \cup \Delta_{i_{m_k}}$, and let $E_k = E \cap \Sigma_k = \{e_{i_j}\}_{j=1,\ldots,m_k}$. Let τ_k be the permutation of S satisfying $\tau_k|_{\Sigma_k} = \sigma|_{\Sigma_k}$ and $\tau(u) = u$ if $u \notin \Sigma_k$. The transitivity of G on S implies the Σ_k partition S, and from this it follows that $\sigma = \tau_1 \tau_2 \cdots \tau_l$.

I claim that

$$\#\{E \cap \sigma(E)\} = \sum_k \#\{E_k \cap \tau_k(E_k)\}. \tag{3.9}$$

To see why this holds, assume that $e \in E \cap \sigma(E)$, and note that this is equivalent to $\sigma^{-1}(e) \in E$. Suppose that $e \in E_k$. Since σ and τ_k give identical permutations of Σ_k, we have $\tau_k^{-1}(e) = \sigma^{-1}(e) \in \Sigma_k \cap E = E_k$. Hence $e \in E_k \cap \tau_k(E_k)$. This is true for exactly one k, and the claim follows.

For each $1 \leq j \leq m_k$, we have

$$\tau_k(e_{i_j}) = \delta^{\epsilon_j}(e_{i_{j+1}}),$$

where $\epsilon_j = 0$ or 1. Since τ_k permutes $\mathfrak{D}$, we have by Proposition 3.14 that δ commutes with τ_k. This gives

$$\tau_k^m(e_{i_1}) = \delta^{\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{m_k}}(e_{i_1}).$$

Since $|\delta| = 2$, it follows from Lemma 3.17 that τ_k is even if and only if $\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{m_k}$ is even. But this sum is simply $\#\{j : 1 \leq j \leq m_k \text{ and } \tau_k(e_{i_j}) \notin E_k\}$, which is the same as

$$m_k - \#\{j : 1 \leq j \leq m_k \text{ and } \tau_k(e_{i_j}) \in E_k\}.$$

This last expression may be restated as $m_k - \#\{E_k \cap \tau_k(E_k)\}$. We have thus established that σ is even if and only if

$$\sum_k \big(m_k - \#\{E_k \cap \tau_k(E_k)\}\big) \equiv 0 \pmod 2. \tag{3.10}$$

Recall that $\sum_k m_k = m$ and m is even. Moreover, since we are working modulo 2 we need only count the k such that $\#\{E_k \cap \tau_k(E_k)\}$ is odd. Hence (3.10) is equivalent to

$$\#\{k : \#\{E_k \cap \tau_k(E_k)\} \text{ is odd}\} \equiv 0 \pmod 2.$$

By equation (3.9), this is equivalent to $\#\{E \cap \sigma(E)\} \equiv 0 \pmod 2$. Thus σ is even if and only if $\#\{E \cap \sigma(E)\}$ is even. ■

Example 3.19. To illustrate Proposition 3.18, consider the group S_{12} acting on the set $\{1, 2, \ldots, 12\}$, with

$$\mathfrak{D} = \{\{1,2\},\{3,4\},\ldots,\{11,12\}\}.$$

Let $\sigma = (1\ \ 9\ \ 2\ \ 10)(3\ \ 4)(5\ \ 8\ \ 12)(6\ \ 7\ \ 11)$. Note that σ is even and induces the permutation $(\Delta_1\ \ \Delta_5)(\Delta_3\ \ \Delta_4\ \ \Delta_6)$ on $\mathfrak{D}$. Choose $E = \{1,4,6,7,9,12\}$. Then $\sigma(E) \cap E = \{7,9\}$, so Proposition 3.18 is borne out. If we choose $E = \{2,4,5,8,9,12\}$, then $\sigma(E) \cap E = \{2,5,8,12\}$, bearing out the Proposition once again. ■

We now take our second step towards showing that $\gamma \neq \omega$ implies that G_n is alternating. The assumption $\gamma \neq \omega$ is equivalent to $\mathfrak{C} \neq \mathcal{Z}$. In order to apply Proposition 3.18 to this case, we need to have a single subset of S that has precisely one representative from each $\Gamma \in \mathfrak{C}$ and from each $\Omega \in \mathcal{Z}$. The following Lemma shows that such a set exists.

Lemma 3.20. *Let S be a set of $2m$ elements, and let*

$$\mathfrak{D} = \{\Delta_1, \ldots \Delta_m\} \qquad \textit{and} \qquad \mathfrak{L} = \{\Lambda_1, \ldots \Lambda_m\}$$

be two partitions of S with $\#\Delta_i = \#\Lambda_i = 2$ for $i = 1, \ldots, m$. There exists $E \subset S$ such that $\#(E \cap \Delta_i) = \#(E \cap \Lambda_i) = 1$ for $i = 1, \ldots, m$.

Proof: Let δ be the permutation associated to $\mathfrak{D}$, and let λ be the permutation associated to $\mathfrak{L}$. Note that because $|\delta| = |\lambda| = 2$, we have $(\delta\lambda)^{-1} = \lambda\delta$. This implies that for any m, $\delta(\delta\lambda)^m = (\delta\lambda)^{-m}\delta$ and $\lambda(\delta\lambda)^m = (\delta\lambda)^{-m}\lambda$. Also note that every Δ_i may be written $\{s, \delta(s)\}$ for some $s \in S$, and every Λ_i may be written $\{s, \lambda(s)\}$ for some $s \in S$.

For reasons that become clear in a moment, we wish E to have the property that $\delta\lambda(E) = E$, so we form E as a union of $\delta\lambda$-orbits:

First, choose $s_0 \in S$, and let $E_0 = \{(\delta\lambda)^j(s_0) : j \in \mathbb{Z}\}$.

Choose, if possible, s_1 such that $\{s_1, \delta(s_1), \lambda(s_1)\} \cap E_0 = \emptyset$, and let $E_1 = \{(\delta\lambda)^j(s_1) : j \in \mathbb{Z}\}$.

Choose, if possible, s_2 such that $\{s_2, \delta(s_2), \lambda(s_2)\} \cap (E_0 \cup E_1) = \emptyset$, and let

$$E_2 = \{(\delta\lambda)^j(s_2) : j \in \mathbb{Z}\}.$$

Continue until

$$\{s, \delta(s), \lambda(s)\} \cap \bigcup_{i=0}^{l} E_i \neq \emptyset \tag{3.11}$$

for all $s \in S$. Let $E = \bigcup_{i=0}^{l} E_i$.

By our remarks in the first paragraph, to establish the lemma it is enough to show that for any $s \in S$, we have $\#(\{s, \delta(s)\} \cap E) = \#(\{s, \lambda(s)\} \cap E) = 1$. Thus we must show that for any $s \in S$, either $s \in E$ and $\{\delta(s), \lambda(s)\} \cap E = \emptyset$ or $s \notin E$ and $\{\delta(s), \lambda(s)\} \subseteq E$.

First note that if $\lambda(s) \in E$ for some $s \in S$, then $\delta\lambda^2(s) \in E$, whence $\delta(s) \in E$. Conversely, if $\delta(s) \in E$ then $(\delta\lambda)^{-1}\delta(s) \in E$, whence $\lambda(s) \in E$. We thus have that either

$$\{\delta(s), \lambda(s)\} \subseteq E \qquad \text{or} \qquad \{\delta(s), \lambda(s)\} \cap E = \emptyset. \tag{3.12}$$

Thus to prove the lemma it is enough to establish the following claim:

Claim: $\{\delta(s), \lambda(s)\} \subseteq E$ if and only if $s \notin E$.

Proof: One direction is easy: suppose $s \notin E$. By (3.11) either $\delta(s) \in E$ or $\lambda(s) \in E$. Thus by (3.12) we have $\{\delta(s), \lambda(s)\} \subseteq E$.

Conversely, suppose that $\{\delta(s), \lambda(s)\} \subseteq E$. Then $\delta(s) \in E_k$ for some k, so that $\delta(s) = (\delta\lambda)^j(s_k)$. Assume that $s \in E$, so that $s = (\delta\lambda)^{j'}(s_{k'})$. Putting these together gives

$$\delta\left((\delta\lambda)^{j'}(s_{k'})\right) = (\delta\lambda)^j(s_k). \tag{3.13}$$

We now derive a contradiction. The left-hand side of (3.13) is the same as $(\lambda\delta)^{j'}(\delta(s_{k'}))$, and since $(\lambda\delta) = (\delta\lambda)^{-1}$ and E_k is $\delta\lambda$-invariant, we have $\delta(s_{k'}) \in E_k$. By construction $\delta(s_{k'})$ is not contained in any E_k with $k < k'$, so we conclude $k \geq k'$. On the other hand, from (3.13) we have

$$(\delta\lambda)^{j'}(s_{k'}) = \delta\left((\delta\lambda)^j(s_k)\right) = (\lambda\delta)^{j-1}(\lambda(s_k)),$$

and from this we conclude $\lambda(s_k) \in E_{k'}$. Thus $k' \geq k$, which shows that $k' = k$.

Equation (3.13) now becomes

$$\delta\left((\delta\lambda)^{j'}(s_k)\right) = (\delta\lambda)^j(s_k),$$

which is the same as

$$\delta(s_k) = (\delta\lambda)^{j+j'}(s_k). \tag{3.14}$$

If $\frac{1}{2}(j + j') = b$ is an integer, then (3.14) yields $\delta((\delta\lambda)^b(s_k)) = (\delta\lambda)^b(s_k)$, which contradicts the fact that δ has no fixed points in S. If $\frac{1}{2}(j + j' - 1) = c$ is an integer, then we can rewrite (3.14) as

$\lambda(s_k) = (\delta\lambda)^{j+j'-1}(s_k)$. This yields $\lambda((\delta\lambda)^c(s_k)) = (\delta\lambda)^c(s_k)$, which contradicts the fact that λ has no fixed points in S. ■

Example 3.21. To illustrate Lemma 3.20, consider the following two partitions of $\{1, 2, \ldots, 14\}$:

$$\mathfrak{D} = \{\{1,2\},\{3,4\},\{5,6\},\{7,8\},\{9,10\},\{11,12\}\{13,14\}\}$$
$$\mathfrak{L} = \{\{1,8\},\{2,6\},\{5,7\},\{3,11\},\{9,14\},\{4,12\},\{10,13\}\}$$

This gives $\delta\lambda = (1\ \ 7\ \ 6)(2\ \ 5\ \ 8)(3\ \ 12)(4\ \ 11)(5\ \ 8)(9\ \ 13)(10\ \ 14)$. Using the notation of the proof of Lemma 3.20, let $s_0 = 1$, giving $E_0 = \{1, 6, 7\}$. Next we choose $s_1 = 3$, which works because $\{3, \delta(3), \lambda(3)\} \cap E_0 = \emptyset$. This gives $E_1 = \{3, 12\}$. Finally we choose $s_2 = 9$, which works because $\{9, \delta(9), \lambda(9)\} \cap (E_0 \cup E_1) = \emptyset$. This gives $E_2 = \{9, 13\}$, and we have $E = \{1, 3, 6, 7, 9, 12, 13\}$. One can quickly verify that $\#(E \cap \Delta_i) = \#(E \cap \Lambda_i) = 1$ for $i = 1, \ldots, 7$. ■

Now all the pieces are in place for the main result of this section.

Theorem 3.22. *Let S be a set of $2m$ elements, let G be a group acting transitively on S, and let*

$$\mathfrak{D} = \{\Delta_1, \ldots \Delta_m\} \qquad \textit{and} \qquad \mathfrak{L} = \{\Lambda_1, \ldots \Lambda_m\}$$

be two CBSs of G with $\#\Delta_i = \#\Lambda_i = 2$ for $i = 1, \ldots, m$. Let δ be the permutation associated to $\mathfrak{D}$, and let λ be the permutation associated to $\mathfrak{L}$. Suppose that $\lambda \in G$ and $\lambda \neq \delta$. Then G is alternating, i.e. comprised entirely of even permutations.

Proof: By Lemma 3.20, we may choose $E \subset S$ such that $\#(E \cap \Delta_i) = \#(E \cap \Lambda_i) = 1$ for $i = 1, \ldots, m$. Let $\sigma \in G$. We want to show that

$$\delta\lambda(E \cap \sigma(E)) = E \cap \sigma(E). \tag{3.15}$$

First we show $\delta\lambda(E) = E$. Let $e \in E$. Then $\#(E \cap \{e, \lambda(e)\}) = 1$, so $\lambda(e) \notin E$. But $\{\lambda(e), \delta\lambda(e)\} \in \mathfrak{D}$, so $\#(E \cap \{\lambda(e), \delta\lambda(e)\}) = 1$. Thus $\delta\lambda(e) \in E$. Next note that by Proposition 3.14, δ and λ commute with σ. Hence

$$\delta\lambda(\sigma(E)) = \sigma(\delta\lambda(E)) = \sigma(E).$$

This establishes (3.15), from which it follows that $E \cap \sigma(E)$ is a union of $\delta\lambda$-orbits.

We claim that all $\delta\lambda$-orbits are composed of exactly two elements. We use the fact that $\lambda \in G$, which by Proposition 3.14 implies that $\lambda\delta = \delta\lambda$. Since $|\delta| = |\lambda| = 2$, this implies that $|\delta\lambda| = 2$. Thus any $\delta\lambda$-orbit has at most two elements. Suppose that $\delta\lambda$ has a one-element orbit:

$$\delta\lambda(s) = s \tag{3.16}$$

for some $s \in S$. By Proposition 3.14, δ and λ commute with any $\sigma \in G$, so applying σ to both sides of (3.16) yields $\delta\lambda(\sigma(s)) = \sigma(s)$. Since G acts transitively on S it follows that $\delta\lambda$ is the identity, which contradicts $\lambda \neq \delta$.

Thus $E \cap \sigma(E)$ is a union of two-element sets, which shows $\#(E \cap \sigma(E))$ is even. By Proposition 3.18 this implies that σ is even. Hence G is alternating. ■ .

The following corollary is a crucial ingredient in the proof of Theorem 1.7.

Corollary 3.23. *Let $\mathfrak{C}$ be the CBS for G_n defined in* (3.7)*, and let γ be the permutation associated to $\mathfrak{C}$. Then $\gamma \in Z(G_n) \cap H_n$. In particular, H_n is nontrivial.*

Proof: Let $\omega \in Z(G_n)$ be as in Proposition 3.16, and let $\mathcal{Z}$ be the CBS associated to ω. Suppose that $\omega \neq \gamma$. By Theorem 3.22 it follows that G_n is alternating, which contradicts Corollary 3.10. Thus $\gamma \in Z(G_n)$. By the definition of $\mathfrak{C}$, a permutation maps to itself every subset belonging to $\mathfrak{C}$ if and only if it fixes every root of f^{n-1}. Thus γ fixes K_{n-1}, whence $\gamma \in H_n$. ■

Example 3.24. To illustrate Corollary 3.23, consider the case $F = \mathbb{F}_p, p \neq 2$, and $n = 2$, as in Example 2.19. We have

$$\mathcal{R}_2 = \left\{ \sqrt{-x+\sqrt{x}}, -\sqrt{-x+\sqrt{x}}, \sqrt{-x-\sqrt{x}}, -\sqrt{-x-\sqrt{x}} \right\},$$

which we label $1, 2, 3$, and 4 respectively. The partition $\mathfrak{C}$ consists of the sets $\{1, 2\}$ and $\{3, 4\}$. Thus $\gamma = (1\ \ 2)(3\ \ 4)$. This limits the possibilities for H_2 to two. Either $H_2 = \{e, (1\ \ 2)(3\ \ 4)\}$ or $H_2 = \{e, (1\ \ 2), (3\ \ 4), (1\ \ 2)(3\ \ 4)\}$. In the latter case H_2 is maximal. ■

3.3 A First Characterization of Maximal H_n

In Section 3.2, we explored the implications of the CBS $\mathfrak{C}$ of G_n defined in (3.7). We now show this is just one of many CBSs for G_n. Fix m with $0 \leq m \leq n$, and let $\{\beta_1, \beta_2, \ldots, \beta_{2^{n-m}}\}$ be the roots of $f^{n-m}(y)$ (by convention we set $f^0(y) = y$). Since $f^n(y) = f^{n-m}(f^m(y))$, we have

$$f^n(y) = \prod_{i=1}^{2^{n-m}} \left(f^m(y) - \beta_i\right).$$

We define the partition $\mathfrak{C}_m$ of $\mathcal{R}_n$ to be

$$\left\{\{\text{roots of } (f^m(y) - \beta_1)\}, \{\text{roots of } (f^m(y) - \beta_2)\}, \ldots, \{\text{roots of } (f^m(y) - \beta_{2^{n-m}})\}\right\}.$$

Note that for $1 \leq i \leq 2^{n-m-1}$, we have $\beta_{2i-1} = \sqrt{-x+\alpha}$ and $\beta_{2i} = -\sqrt{-x+\alpha}$ for some root α of f^{n-m-1}. Thus $(f^m(y) - \beta_{2i-1})(f^m(y) - \beta_{2i}) = (f^m(y))^2 + x - \alpha = f^{m+1}(y) - \alpha$. It follows that

$$\{\text{roots of } (f^m(y) - \beta_{2i-1})\} \cup \{\text{roots of } (f^m(y) - \beta_{2i})\} = \{\text{roots of } (f^{m+1}(y) - \alpha)\} \tag{3.17}$$

The most important aspect of the partitions $\mathfrak{C}_m$ is that any $\sigma \in G_n$ permutes the roots of f^{n-m}, and thus permutes the subsets belonging to $\mathfrak{C}_m$. Thus $\mathfrak{C}_m$ is a CBS for every m with $0 \leq m \leq n$. Moreover, σ induces the same permutation on $\mathfrak{C}_m$ and $\mathcal{R}_{n-m}$. Note that $\mathfrak{C}_1$ is the same as the CBS $\mathfrak{C}$ of Section 3.2. Clearly $\mathfrak{C}_0$ is the CBS consisting of one-element subsets of $\mathcal{R}_n$, while $\mathfrak{C}_n$ is the CBS consisting of only the subset $\mathcal{R}_n$.

Now we will use some abelian Kummer Theory. Let $\mathcal{R}_{n-1}$ be the roots of f^{n-1}. The extension K_n/K_{n-1} is obtained by adjoining square roots of the elements $-x+\beta \in K_{n-1}$ for each $\beta \in \mathcal{R}_{n-1}$. The maximal degree of this extension is thus $2^{2^{n-1}}$, because $|\mathcal{R}_{n-1}| = 2^{n-1}$. Thus the order of H_n is at most $2^{2^{n-1}}$. Consider the multiplicative group K_{n-1}^*, and let $V = \{-x+\beta : \beta \in \mathcal{R}_{n-1}\}$. Let $\langle V, {K_{n-1}^*}^2 \rangle$ be the subgroup of K_{n-1}^* generated by ${K_{n-1}^*}^2$ and V. Abelian Kummer Theory (see e.g. [21]) gives us an isomorphism

$$H_n \cong \langle V, {K_{n-1}^*}^2 \rangle / {K_{n-1}^*}^2.$$

Every coset representative of the group $\langle V, {K_{n-1}^*}^2 \rangle / {K_{n-1}^*}^2$ is a product of elements belonging to V. Thus all coset representatives are contained in the set of all products

$$\prod_{\beta \in S} (\beta - x), \tag{3.18}$$

where S varies over the subsets of $\mathcal{R}_{n-1}$. These products are in one-to-one correspondence with the subsets of $\mathcal{R}_{n-1}$ and thus there are $2^{2^{n-1}}$ of them. When H_n is maximal they are all distinct cosets.

Now suppose that H_n is not maximal. Then the cosets whose representatives are given in (3.18) are not all distinct, so we have

$$\left(\prod_{\beta \in T} (\beta - x) \right) \left(\prod_{\beta \in T'} (\beta - x) \right) \in {K_{n-1}^*}^2$$

for two distinct subsets T, T' of $\mathcal{R}_{n-1}$. Hence $S = (T \cup T') - (T \cap T')$ is nonempty and we have

$$\prod_{\beta \in S} (\beta - x) \in {K_{n-1}^*}^2. \tag{3.19}$$

Given (3.19), we wish to use elements of G_{n-1} to force other such products to be in ${K_{n-1}^*}^2$. We use the notation $A \ominus B$ to denote the symmetric difference of A and B, that is, $A \ominus B = (A \cup B) - (A \cap B)$.

Proposition 3.25. *Suppose that for some $S \subseteq \mathcal{R}_{n-1}$,*

$$\prod_{\beta \in S} (\beta - x) \in {K_{n-1}^*}^2. \tag{3.20}$$

Let $\sigma \in G_{n-1}$. Then

$$\prod_{\beta \in \sigma(S) \ominus S} (\beta - x) \in {K_{n-1}^*}^2.$$

Proof: Applying σ to (3.20) yields

$$\prod_{\beta \in S} (\sigma(\beta) - x) \in {K_{n-1}^*}^2,$$

which is the same as

$$\prod_{\beta \in \sigma(S)} (\beta - x) \in {K_{n-1}^*}^2.$$

Hence

$$\left(\prod_{\beta \in S} (\beta - x) \right) \left(\prod_{\beta \in \sigma(S)} (\beta - x) \right) \in {K_{n-1}^*}^2,$$

and so

$$\left(\prod_{\beta \in \sigma(S) \ominus S} (\beta - x) \right) \left(\prod_{\beta \in \sigma(S) \cap S} (\beta - x)^2 \right) \in {K_{n-1}^*}^2.$$

The proposition then follows. ■

Supposing that a product of the form $\prod_{\beta \in S}(\beta - x)$ is in ${K_{n-1}^*}^2$, the next proposition gives more information about what other such products we can force to be in ${K_{n-1}^*}^2$.

Proposition 3.26. *Let m satisfy $0 \leq m \leq n-2$, and let $S \subset \mathcal{R}_{n-1}$ be a nonempty union of subsets belonging to $\mathfrak{C}_m$ but not a union of subsets belonging to $\mathfrak{C}_{m+1}$. Then there exists $\sigma \in G_{n-1}$ such that $\sigma(S) \ominus S$ is a nonempty union of subsets belonging to $\mathfrak{C}_{m+1}$*

Proof: Write

$$\mathfrak{C}_m = \{\Gamma_{m,1}, \Gamma_{m,2}, \ldots \Gamma_{m,2^{n-m}}\},$$

and note that by (3.17) we have

$$\mathfrak{C}_{m+1} = \left\{ \{\Gamma_{m,1} \cup \Gamma_{m,2}\}, \ldots, \{\Gamma_{m,2^{n-m}-1} \cup \Gamma_{m,2^{n-m}}\} \right\}.$$

Let $\gamma \in H_{n-m}$ be the element mentioned in Corollary 3.23. Let σ be any element of G_{n-1} extending γ (by general Galois theory, e.g., as in [10], such an element exists). The action of γ

on the subsets belonging to $\mathfrak{C}_m$ is the same as the action of γ on $\mathcal{R}_{n-m}$. Recall that if $\mathcal{R}_{n-m} = \{\beta_1, \ldots, \beta_{2^{n-m}}\}$ then γ exchanges the elements β_{2i-1} and β_{2i} for $1 \leq i \leq 2^{n-m-1}$. Since σ extends γ, we have

$$\sigma(\Gamma_{m,2i-1}) = \Gamma_{m,2i} \qquad \text{and} \qquad \sigma(\Gamma_{m,2i}) = \Gamma_{m,2i-1} \tag{3.21}$$

for $1 \leq i \leq 2^{n-m-1}$. Thus for any $\Gamma \in \mathfrak{C}_m$, the oribt of Γ under σ is a subset belonging to $\mathfrak{C}_{m+1}$. Note that because S is a union of subsets belonging to $\mathfrak{C}_m$ and σ permutes $\mathfrak{C}_m$, $\sigma(S) \ominus S$ is also a union of subsets belonging to $\mathfrak{C}_m$. To prove the proposition, it is therefore enough to show that $\sigma(S) \ominus S$ is nonempty and a union of σ- orbits, i.e. $\sigma(\sigma(S) \ominus S) = \sigma(S) \ominus S$.

Since S is nonempty and not a union of elements of $\mathfrak{C}_{m+1}$, there is a pair $\{\Gamma_{m,2k-1}, \Gamma_{m,2k}\}$ such that one is in S and the other is not. Thus

$$\{\Gamma_{m,2k-1} \cup \Gamma_{m,2k}\} \subset \sigma(S) \ominus S,$$

implying that $\sigma(S) \ominus S$ is not empty.

If T, T' are subsets of $\mathcal{R}_n$ then $\sigma(T - T') = \sigma(T) - \sigma(T')$ because σ permutes $\mathcal{R}_n$. Similar identities hold for $\sigma(T \cup T')$ and $\sigma(T \cap T')$. Thus we have

$$\begin{aligned}\sigma(\sigma(S) \ominus S) &= \sigma(\sigma(S) \cup S) - \sigma(\sigma(S) \cap S) \\ &= (\sigma^2(S) \cup \sigma(S)) - (\sigma^2(S) \cap \sigma(S)) = \sigma(S) \ominus \sigma^2(S). \end{aligned} \tag{3.22}$$

By (3.21), σ^2 maps to itself each subset belonging to $\mathfrak{C}_m$. Since S is a union of such subsets we have $\sigma^2(S) = S$. Thus (3.22) shows $\sigma(\sigma(S) \ominus S) = \sigma(S) \ominus S$. ■

We now apply Propositions 3.25 and 3.26 to show the main result of this section.

Theorem 3.27. *H_n is maximal if and only if $p_n \notin {K_{n-1}^*}^2$.*

Proof: We show the contrapositive of both directions. First assume that H_n is not maximal, whence $[K_n : K_{n-1}]$ is not maximal. Let $\mathcal{R}_{n-1}$ be the set of roots of f^{n-1}. By (3.19),

$$\prod_{\beta \in S} (\beta - x) \in {K_{n-1}^*}^2$$

for some $S \subseteq \mathcal{R}_{n-1}$. We wish to show

$$\prod_{\beta \in \mathcal{R}_{n-1}} (\beta - x) \in {K_{n-1}^*}^2. \tag{3.23}$$

If $S = \mathcal{R}_{n-1}$ we are done. If $S \neq \mathcal{R}_{n-1}$, there is some m with $0 \leq m \leq n-2$ such that S is a union of subsets belonging to $\mathfrak{C}_m$ but not a union of subsets belonging to $\mathfrak{C}_{m+1}$. In this case, applying Propositions 3.25 and 3.26 a finite number of times yields (3.23).

The roots α of f^n are the same as the roots of the equations $y^2 = \beta - x$ for $\beta \in \mathcal{R}_{n-1}$. Thus the product of all roots of f^n is

$$\prod_{\beta \in \mathcal{R}_{n-1}} -(\beta - x)\,.$$

Since $\#\mathcal{R}_{n-1}$ is even, this product is the same as the one in (3.23). But the product of all roots of f^n is $f^n(0)$ (there is no minus sign because f^n has even degree), and by definition $f^n(0) = p_n$. Thus (3.23) gives $p_n \in {K^*_{n-1}}^2$.

Now assume $p_n \in {K^*_{n-1}}^2$. By the remarks of the preceding paragraph, we have that (3.23) holds. Thus the cosets defined in (3.18) are not all distinct, whence H_n is not maximal by Kummer theory. ■

3.4 Primitive Mandelbrot Periods and Maximal H_n

We begin this section by returning to the study of p_n left off in Section 3.1. To apply the results of Section 3.1 to H_n, we examine the discriminant $\mathrm{D}_{K_{n-1}/K}$ of the extension K_{n-1}/K and compare it to $\mathrm{Disc}\, f_n$. Finally, we combine results from every section of this chapter to prove Theorem 3.2.

We noted in section 3.1 that the roots of p_n consist of all $c \in \overline{F}$ such that 0 is periodic under iteration of $f_c = x^2 + c$ with period dividing n. We now wish to examine the values of c for which 0 is periodic with primitive period n (some authors refer to this as exact period n). We say that such c have *primitive Mandelbrot period n*. To this end, we define the following element of K:

$$\Phi_n = \prod_{d|n} (p_d)^{\mu(n/d)}. \tag{3.24}$$

Note that it is not immediately clear that the Φ_n are polynomials in x. For if $\mu(n/d) = -1$, there will be a factor of p_d in the denominator of Φ_n. We will take care of this difficulty in a moment.

We say that $c \in \overline{F}$ has *formal Mandelbrot period n* if $\Phi_n(c) = 0$. Suppose that c has primitive Mandelbrot period n. Then $p_n(c) = 0$ but $p_m(c) \neq 0$ for each $m < n$. It is then clear from the definition of Φ_n that $\Phi_n(c) = 0$, since there is a factor of p_n in the numerator but no such factors in the denominator. Thus any point of primitive Mandelbrot period n must have formal Mandelbrot period n. The next proposition shows the converse holds as well. It is interesting to contrast this result with those in [25] regarding the primitive part of cyclotomic polynomials.

Proposition 3.28. *For each n, Φ_n is a polynomial. Moreover, the roots of Φ_n are precisely those $c \in \overline{F}$ with primitive Mandelbrot period n.*

Proof: Our argument is modeled after the one in [25], Lemma 2.3. Let $c \in \overline{F}$ be a root of p_n, and let m be such that c is a primitive root of p_m, i.e. $p_m(c) = 0$ but $p_i(c) \neq 0$ for all $i < m$. Clearly we must have $m|n$. Then for any d, $p_d(c) = 0$ if and only if $m|d$. Thus in the product (3.24), the only terms we need to consider are the ones corresponding to the divisors d of n such that $m|d$. Hence if we write $n = mn_1$, we are interested in terms corresponding to all the products mk that divide mn_1. Thus we consider

$$\prod_{mk|mn_1} (p_{mk})^{\mu(n/mk)} = \prod_{k|n_1} (p_{mk})^{\mu(n_1/k)}.$$

Suppose now that $\mathrm{ord}_{(x-c)}(p_m) = e$; by assumption $e \geq 1$. Proposition 3.7 then tells us that for all k, $\mathrm{ord}_{(x-c)}(p_{mk}) = e$. Hence we have

$$\begin{aligned} \mathrm{ord}_{(x-c)} \left(\prod_{k|n_1} (p_{mk})^{\mu(n_1/k)} \right) &= \sum_{k|n_1} \mu\left(\frac{n_1}{k}\right) \mathrm{ord}_{(x-c)}(p_{mk}) \\ &= e \sum_{k|n_1} \mu\left(\frac{n_1}{k}\right) \\ &= \begin{cases} e & \text{if } n_1 = 1, \text{ that is, } m = n \\ 0 & \text{if } n_1 > 1, \text{ that is, } m < n \end{cases} \end{aligned}$$

In either case $\mathrm{ord}_{(x-c)}(\Phi_n) \geq 0$. By the definition of Φ_n, the set of possible poles of Φ_n is contained in the roots of p_n (the roots of p_d are also roots of p_n for any $d|n$), so we've shown that Φ_n is a polynomial. Moreover, any root c of Φ_n is a root of p_n, because a root of the product (3.24) must be a root of p_d for some $d|n$. Thus the above argument shows that the primitive Mandelbrot period of c cannot be less than n. Hence every root of Φ_n has primitive Mandelbrot period n. The converse was shown in the discussion immediately preceding this proposition. ■

Corollary 3.29. *The Φ_n are pairwise relatively prime.*

Proof: Immediate from Proposition 3.28. ■

We now consider $\deg(\Phi_n)$. From (3.24), we see that

$$\deg(\Phi_n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) \deg(p_d) = \sum_{d|n} \mu\left(\frac{n}{d}\right) 2^{d-1}. \tag{3.25}$$

Corollary 3.30. $\deg(\Phi_n)$ *is odd if and only if n is squarefree.*

Proof: From (3.25) we see that $\deg(\Phi_n)$ is odd if and only if $\mu(n) \neq 0$, i.e. if and only if n is squarefree. ■

Note also that the term of the rightmost sum in (3.25) corresponding to $d = n$ contributes 2^{n-1}. Since

$$\sum_{i=1}^{\lfloor n/2 \rfloor} 2^{i-1} \leq 2^{n/2} - 1$$

for $n \geq 2$, we conclude that $\deg(\Phi_n) \geq 2^{n-1} - 2^{n/2}$. For $n \geq 4$, this last expression is at least 2^{n-2}. Since $\deg(\Phi_1) = \deg(\Phi_2) = 1$ and $\deg(\Phi_3) = 3$, we have that $\deg(\Phi_n) \geq 1$ for all n. This shows:

Corollary 3.31. *For each $n \geq 1$ there exists $c \in \overline{F}$ with primitive Mandelbrot period n.*

We now describe our method for proving Theorem 3.2. From Corollary 3.25 we see that Φ_n cannot be a square in K when n is squarefree. We wish to show that $\Phi_n \notin K^{*2}$ implies H_n maximal. We use proof by contradiction, and the following lemma provides a big step towards a contradiction.

Lemma 3.32. *Suppose that $\Phi_n \notin K^{*2}$ and H_n is not maximal. Then there is an irreducible factor q of Φ_n such that $q \mid \mathrm{D}_{K_{n-1}/K}$.*

Proof: Since Φ_n is not a square in K, the squarefree part of Φ_n must include at least one irreducible $q \in \mathbb{F}_p[x]$, i.e. $\mathrm{ord}_q(\Phi_n)$ is odd. By Proposition 3.28, all roots of q in $\overline{F}$ have primitive Mandelbrot period n, while all roots of $\frac{p_n}{\Phi_n}$ have primitive Mandelbrot period less than n. Thus $q \nmid \frac{p_n}{\Phi_n}$, meaning that $\mathrm{ord}_q(p_n)$ is odd.

Since H_n is not maximal, Theorem 3.27 implies $p_n \in K_{n-1}^{*2}$. It follows that q is a square in K_{n-1} (because $\mathrm{ord}_q(p_n)$ is odd). Thus the ideal (q) ramifies in K_{n-1}, so $q \mid \mathrm{D}_{K_{n-1}/K}$. ■

To complete the proof of Theorem 3.2, we wish to show that if $q \in F[x]$ is irreducible and $q \mid \mathrm{D}_{K_{n-1}/K}$, then $q \mid \mathrm{Disc}\, f_n$. We begin by estimating $\mathrm{D}_{K_n/K_{n-1}}$, which requires finding a basis of K_n/K_{n-1}. Let $\mathcal{R}_{n-1} = \{\beta_1, \ldots, \beta_{2^{n-1}}\}$, and for $i = 1, \ldots, 2^{n-1}$ let α_i satisfy $\alpha_i^2 = -x + \beta_i$. Note that

$$K_n = K_{n-1}(\alpha_1, \ldots, \alpha_{2^{n-1}}).$$

Choose $S \subseteq \{\alpha_1, \ldots, \alpha_{2^{n-1}}\}$ of minimal size such that $K_n = K_{n-1}(S)$. Thus we have

$$[K_n : K_{n-1}] = 2^{|S|}$$

. Define $B \subset K_n$ as the set of all products of the form $\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_m}$ where $0 \leq m \leq |S|$ and $\alpha_{i_j} \in S$.

Proposition 3.33. *B is a basis for K_n/K_{n-1}.*

Proof: The elements of B are in one-to-one correspondence with the subsets of the set S above, so $|B| = 2^{|S|} = [K_n : K_{n-1}]$. Hence it suffices to show that B spans K_n/K_{n-1}. Let $\gamma \in K_n$. Since $K_n = K_{n-1}(S)$, we can write γ as a K_{n-1}-linear combination of products of powers of the $\alpha \in S$.

Since $\alpha^2 \in K_{n-1}$ for all $\alpha \in S$, γ can in fact be written as a linear combination of products of elements in S. Therefore B spans K_n/K_{n-1}.

Recall from earlier that we denote the Galois group of K_n/K_{n-1} by H_n. By Corollary 3.23 we know that $[K_n : K_{n-1}] > 1$, which implies that S is not empty. Fix an i with $\alpha_i \in S$, and let $S_i = S - \{\alpha_i\}$. Since $|S_i| < |S|$, we cannot have $K_n = K_{n-1}(S_i)$, so we must have $[K_n : K_{n-1}(S_i)] = 2$. The unique non-trivial automorphism of this extension is the one that sends α_i to $-\alpha_i$ and fixes α_j for each $\alpha_j \in S$ with $j \neq i$. Call this automorphism σ_i; clearly it is an element of H_n.

Proposition 3.34. *Let $b, b' \in B$. Then*

$$\mathrm{Tr}_{K_n/K_{n-1}}(bb') = \begin{cases} [K_n : K_{n-1}]b^2 & \text{if } b = b' \\ 0 & \text{if } b \neq b' \end{cases}$$

Proof: First suppose $b = b'$. Then either both are 1 and we are done or bb' is a product of some number of α_i^2. Since each α_i is the square root of an element of K_{n-1}, we immediately have $bb' \in K_{n-1}$. Hence $\mathrm{Tr}_{K_n/K_{n-1}}(bb') = [K_n : K_{n-1}]bb' \neq 0$ (since we are not in characteristic 2).

Now suppose that $b \neq b'$, and let S be the set defined just before Proposition 3.33. Without loss of generality there is an $\alpha_i \in S$ such that $\alpha_i \mid b$ but $\alpha_i \nmid b'$. Hence $bb' = \alpha_i\gamma$, where α_i does not divide γ. Consider the extension $K_n/K_{n-1}(S_i)$. As noted above, the Galois group of this extension consists of the identity and σ_i. But $\sigma_i(bb') = -bb'$, so we have $\mathrm{Tr}_{K_n/K_{n-1}(S_i)}(bb') = bb' - bb' = 0$. Now by the transitivity of the trace with respect to extensions ([13, page 15]),

$$\begin{aligned} \mathrm{Tr}_{K_n/K_{n-1}}(bb') &= \mathrm{Tr}_{K_{n-1}(S_i)/K_{n-1}}(\mathrm{Tr}_{K_n/K_{n-1}(S_i)}(bb')) \\ &= \mathrm{Tr}_{K_{n-1}(S_i)/K_{n-1}}(0) \\ &= 0. \end{aligned}$$

■

Corollary 3.35. *For each n there exist positive integers m_1 and m_2 such that*

$$\mathrm{D}_{K_n/K_{n-1}} \mid 2^{m_1}\left(\prod_{\alpha_i \in S} \alpha_i^2\right)^{m_2}.$$

Proof: Let $A = \mathbb{F}_p[x]$, and denote by A_{n-1} and A_n the integral closures of A in K_{n-1} and K_n respectively. Then A_n is an A_{n-1}-module of rank $[K_n : K_{n-1}]$. Also, every α_i is integral over K since its minimal polynomial over K is $f^n \in A[y]$, which is monic. Hence $B \subseteq A_n$. Since the elements of B generate K_n as a K_{n-1}-vector space, they must generate a free R_{n-1}-subsmodule of R_n that has rank $[K_n : K_{n-1}]$. Thus we have

$$\mathrm{D}_{K_n/K_{n-1}} \mid \mathrm{D}_{K_n/K_{n-1}}(B),$$

where $\mathrm{D}_{K_n/K_{n-1}}(B) = \det(\mathrm{Tr}_{K_n/K_{n-1}}(b_i b_j))$ (see e.g. [20, page 65]).

However, Proposition 3.34 gives

$$\mathrm{Tr}_{K_n/K_{n-1}}(b_i b_j) = \begin{cases} [K_n : K_{n-1}]b_i^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

So the matrix in question is diagonal, with determinant $\prod_i [K_n : K_{n-1}]b_i^2$. For each k, the number of $b_i \in B$ with $\alpha_k \mid b_i$ is $2^{|S|-1}$ (one for each subset of $S - \{\alpha_k\}$). Hence we have

$$\mathrm{D}_{K_n/K_{n-1}}(B) = [K_n : K_{n-1}]^{|B|} \left(\prod_{\alpha_i \in S} \alpha_i^2 \right)^{2^{|A|-1}}$$

Since $[K_n : K_{n-1}]$ is a power of 2, the first factor is also a power of 2. This proves the corollary. ■

We use Corollary 3.35 to get information about $\mathrm{D}_{K_n/K}$. For this we need the tower formula for discriminants (see e.g. [13, page 126]). In our case it gives

$$\mathrm{D}_{K_n/K} = \mathrm{N}_{K_{n-1}/K}(\mathrm{D}_{K_n/K_{n-1}})(\mathrm{D}_{K_{n-1}/K})^{[K_n:K_{n-1}]}.$$

Applying Corollary 3.35 to this equation yields

$$\mathrm{D}_{K_n/K} \mid 2^{m_3} \left(\prod_{\alpha_i \in S} \mathrm{N}_{K_{n-1}/K}(\alpha_i^2) \right)^{m_2} (\mathrm{D}_{K_{n-1}/K})^{[K_n:K_{n-1}]}. \tag{3.26}$$

The Galois conjugates of α_i^2 are simply the α_j^2, where $j = 1, \ldots, 2^{n-1}$. We noted in the discussion preceding Proposition 3.9 that $\prod_{j=1}^{2^{n-1}} \alpha_j^2 = p_n$. Thus for each i, $\mathrm{N}_{K_{n-1}/K}(\alpha_i^2)$ is a power of p_n, and so (3.26) becomes

$$\mathrm{D}_{K_n/K} \mid 2^{m_3}(p_n)^{m_4}(\mathrm{D}_{K_{n-1}/K})^{[K_n:K_{n-1}]}.$$

for suitable numbers m, m_n depending on n.

An easy inductive argument now shows:

Proposition 3.36. *There exist positive integers $j_0, j_1, j_2, \ldots, j_n$, such that*

$$\mathrm{D}_{K_n/K} \mid 2^{j_0} \prod_{i=1}^{n} (p_i)^{j_i}.$$

With Proposition (3.36) and Lemma 3.32 we are near to being able to prove an important theorem relating the maximality of H_n to the factorization of Φ_n in the prime subfield of K. This theorem leads directly to the main results of this chapter. However, first we need the following lemma, which shows that to prove $\Phi_n \notin K^{*2}$ it is enough to verify that Φ_n is not a square in $F'(x)$, where either $F' = \mathbb{Q}$ or else $F' = \mathbb{F}_p$ is the prime subfield of F.

Lemma 3.37. *Let* $\operatorname{char} F \neq 2$, $K = F(x)$, F' *the prime subfield of* F, *and let* Φ_n *be as in* (3.24). *Suppose that* $\Phi_n \notin F'(x)^{*2}$. *Then* $\Phi_n \notin K^{*2}$.

Proof: The main thing to note is that p_n has coefficients in F', and it follows from (3.24) and Proposition 3.28 that Φ_n does as well. We prove the contrapositive of the Lemma. Suppose that $\Phi_n = q^2$ for some $q \in K$. Then all roots of q in $\overline{K}$ are also roots of Φ_n, and hence lie in $\overline{F'}$. The coefficients of q, being symmetric polynomials of the roots, thus also lie in $\overline{F'}$. The fact that F' is perfect implies that irreducible factors of Φ_n in $F'[x]$ cannot become squares in $\overline{F'}(x)$. Therefore if Φ_n is a square in $\overline{F'}(x)$ it must already be a square in $F'(x)$. ■

We now give a theorem that quickly leads to the main results of the chapter.

Theorem 3.38. *Let* $\operatorname{char} F \neq 2$, $K = F(x)$, *and* $f = y^2 + x \in K[y]$. *Let* $F' \subseteq F$ *be the prime subfield of* F, *and let* K_n *be the splitting field of* f^n *over* K. *Finally, put* $H_n = \operatorname{Gal}(K_n/K_{n-1})$ *and let* $\Phi_n \in K$ *be as defined in* (3.24). *Then* H_n *is maximal (i.e.* $\#H_n = 2^{2^{n-1}}$*) if and only if* $\Phi_n \notin F'(x)^{*2}$.

Proof: By Lemma 3.37, $\Phi_n \notin F'(x)^{*2}$ implies that $\Phi_n \notin K^{*2}$. By Lemma 3.32 it is then enough to derive a contradiction from the assertion that there is an irreducible factor q of Φ_n such that $q \mid \mathrm{D}_{K_{n-1}/K}$. From Proposition 3.36 we see that all roots of $\mathrm{D}_{K_{n-1}/K}$ have primitive Mandelbrot period at most $n-1$. Yet by Proposition 3.28 roots of q have primitive Mandelbrot period n.

To show the converse, we remark that if Φ_n is a square in $F'(x)$ then p_n is a square in K_{n-1}. To see this, note that $p_n = \prod_{d|n} \Phi_d$ (use Mobius inversion on (3.24)), and one can easily show that $\Phi_m \in K_{n-1}^{*2}$ for any $m \leq n-1$. That H_n is not maximal then follows from Theorem 3.27. ■

Theorem 3.38 shows that to determine whether H_n is maximal we need only prove something about the factorization of Φ_n in the prime subfield of F. This handy criterion leads directly to a proof of the first two statements of Theorem 3.2 (we give a proof of the third in section 3.5). Recall that Theorem 3.2 states that:

1. If $\operatorname{char} F \neq 2$, H_n is maximal for all squarefree n.

2. If F has characteristic 0, then H_n is maximal for all n.

3. If F has characteristic $p \equiv 3 \bmod 4$, then H_n is maximal for all n.

Proof of statements 1 and 2: Proposition 3.30 shows that $\deg \Phi_n$ is odd when n squarefree. By Theorem 3.38, it follows that H_n is maximal when n is squarefree. Suppose now that F has

characteristic 0. Reducing modulo 2 we have

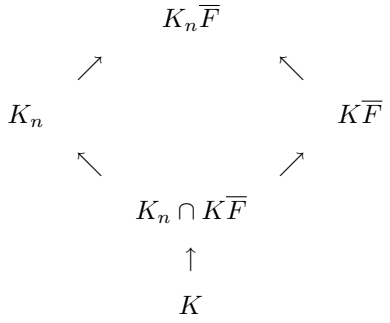
$$(p_n)' = 2p_{n-1}(p_{n-1})' - 1 = 1$$

for any n. Therefore p_n is separable over $\mathbb{Q}$ for all n. Theorem 3.38 then gives H_n maximal for all n. ■

One of the salient features of Theorem 3.38 is that it reduces the maximality of H_n (and therefore G_n) to information contained entirely in the prime subfield of F. In particular, it shows that if G_n is maximal, then it remains maximal when F is replaced by any algebraic extension of F. Since constant field extensions (see Definition 2.14) in K_n/K are subextensions that are algebraic over F, it follows that they vanish when we replace F by $\overline{F}$. Thus the fact that G_n remains maximal when we replace F by $\overline{F}$ implies that K_n/K is geometric. This is the essence of the proof of the following corollary:

Corollary 3.39. *Suppose that H_m is maximal for all $m \leq n$. Then K_n/K is geometric.*

Proof: Let $\overline{F}$ be the algebraic closure of F, and recall that $K = F(x)$ and K_n/K is geometric if $\overline{F} \cap K_n = F$. Consider the diagram



where the arrows denote inclusion. By a basic theorem of Galois theory [10, page 505], we have

$$\mathrm{Gal}\,(K_n\overline{F}/K\overline{F}) \cong \mathrm{Gal}\,(K_n/K_n \cap K\overline{F}). \tag{3.27}$$

We clearly have $K\overline{F} = \overline{F}F(x) = \overline{F}(x)$, so the left-hand side of (3.27) is $\mathrm{Gal}\,(K_n\overline{F}/\overline{F}(x))$. However, this is just the splitting field of of f^n over $\overline{F}(x)$. By Theorem 3.38 the assumption that H_m maximal for all $m \leq n$ implies that Φ_m is not a square in K for all $m \leq n$. Using Theorem 3.38 again gives that G_n remains maximal when we replace F by any algebraic extension of F. Therefore

$$\#\mathrm{Gal}\,(K_n\overline{F}/\overline{F}(x)) = \#\mathrm{Gal}\,(K_n/K).$$

Combining this with (3.27) gives $\#\mathrm{Gal}\,(K_n/K_n \cap \overline{F}(x)) = \#\mathrm{Gal}\,(K_n/K)$. Therefore the two groups are in fact equal, which implies $K_n \cap \overline{F}(x) = K$. Intersecting both sides of this with $\overline{F}$ then gives $K_n \cap \overline{F} = F$, as desired. ■

We now have the following immediate consequence of Theorem 3.2 and Corollary 3.39.

Corollary 3.40. *If* $\operatorname{char} F = 0$ *or* $\operatorname{char} F = p \equiv 3 \bmod 4$ *then* K_n/K *is geometric for all* n.

We end this section with some evidence for Conjecture 3.1, which states that for $\operatorname{char} F \neq 2$, H_n is maximal for all n, and Conjecture 2.17, which states that K_n/K is geometric for all n. By Theorem 3.38 and Corollary 3.39, both of these conjectures follow from showing that Φ_n is not a square in $\mathbb{F}_p(x)$ for all $n \geq 1$ and for all $p \neq 2$ (the charactertic zero case is proven in Theorem 3.2).

When n is not squarefree the degree of Φ_n is even, making it a candidate to be a square in $\mathbb{F}_p[x]$. On the other hand, as shown in the proof of statements 1 and 2 of Theorem 3.2 (page 54), Φ_n is separable over $\mathbb{Q}$, meaning that there are only finitely many primes p (those dividing $\operatorname{Disc} \Phi_n$) such that Φ_n is not separable over $\mathbb{F}_p$. Theorem 3.38 then shows that for any fixed n, H_n is maximal for any F whose characteristic does not belong to some finite set of primes. Moreover, the degree of Φ_n grows very quickly, on the order of 2^n. Therefore it has a large number of distinct roots in $\overline{\mathbb{Q}}$. Intuitively it seems unlikely that all these roots would overlap when reduced modulo one of a finite set of primes.

We can use Theorem 3.38 and the observations of the previous paragraph to give numerical evidence for Conjectures 3.1 and 2.17. In particular, we can settle the case $n = 4$ in general. Over $\mathbb{Q}$, we have $\operatorname{Disc} \Phi_4 = 58673 = 23 \times 2551$, and one can verify directly that modulo both these primes Φ_4 is not a square. Thus H_4 is maximal for all F of characteristic $\neq 2$. Combining this with Theorem 3.2 shows (still assuming $\operatorname{char} F \neq 2$) that H_n is maximal for $n \leq 7$, whence G_n is maximal and K_n/K is geometric (by Corollary 3.39) for $n \leq 7$.

Even the second non-squarefree case of $n = 8$ is too big to tackle in this manner: the discriminant over $\mathbb{Q}$ of Φ_8 is approximately 10^{250}, large enough that factorization is no longer feasible. For fixed $p \neq 2$, we may choose $c \in \mathbb{F}_{p^k}$ for some $k \geq 1$ and compute $\Phi_n(c)$ quickly. If this is not a square in $\mathbb{F}_{p^k}$ for some c, then the maximality of H_n is established. Using this technique, one can verify that for $p = 5$, a case not covered by Theorem 3.2, H_n is maximal (and thus K_n/K is geometric) for all $n \leq 2000$. Note that this justifies the comment made on page 28.

We have one bit of unfinished business in this chapter, which is to give the proof of Theorem 3.2 in the case where the characteristic of F is a prime of the form $4k + 3$.

3.5 The case $\operatorname{char} F = p \equiv 3 \bmod 4$ in Theorem 3.2

An interesting question related to our considerations in this chapter is to determine the Galois groups over $\mathbb{Q}$ of the iterates of $x^2 + 1$. First posed by J. McKay, it was addressed by Odoni [26] who reduced the question to whether certain elements b_n were squares in $\mathbb{Q}$, analogously to the Φ_n

in our results of Chapter 3. In a 1992 paper [35], Stoll devised a clever method that allowed him to show that the b_n are never squares. He extended his result to show that the Galois groups of iterates of $x^2 + a$ are all maximal for an infinite set of $a \in \mathbb{Z}$. Here we use a minor modification of Stoll's argument to show that Φ_n is not a square in $\mathbb{F}_p(x)$ when $p \equiv 3 \bmod 4$ (Theorem 3.44). Using Theorem 3.38, this proves the final assertion of Theorem 3.2.

We make some modifications to our setup as follows: let A be the ring of integers in a Dedekind domain K, suppose A is a principal ideal domain, and let $g \in A[y^2]$ be an even polynomial. Put $q_1 = \pm g(0)$, and $q_n = g(q_{n-1})$ for $n \geq 1$. Assume all $q_n \neq 0$ and put $\Psi_n = \prod_{d|n} q_d^{\mu(n/d)}$.

Lemma 3.41. *For each n, $\Psi_n \in A$.*

Proof: : A straightforward generalization of Proposition 3.28. ■

Lemma 3.42. *Suppose that for all $n \geq 1$ there is an $m_n \in A$ such that:*

1. *m_n and q_n are relatively prime,*
2. *$m_n \mid q_n + q_{2n}$, and*
3. *-1 is not a square in $A/(m_n)$.*

Then for all $n \geq 2$, Ψ_n is not a square in K.

Proof: Let $n \geq 2$ have prime decomposition $p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r}$, where $e_j \geq 1$, $r \geq 1$, and the p_j are distinct primes. Set $n' = p_1 \cdots p_r > 1$, and $k = n/n'$; note that if $d \mid n$, then $\mu(n/d) \neq 0$ precisely when d is a multiple of k.

Note that m_k and q_{2k} are relatively prime: any common factor of m_k and q_{2k} must also be a common factor of m_n and q_k by property 2 above, and thus this common factor is 1 by property 1 above. Clearly $q_k \equiv -q_{2k} \bmod m_k$. We show that $q_{lk} \equiv q_{2k} \bmod m_k$ for all $l \geq 2$. We have

$$q_{3k} = g^k(q_{2k}) \equiv g^k(-q_k) \bmod m_k. \tag{3.28}$$

Since $g \in A[y^2]$, $g^k(-q_k) = g^k(q_k)$, and $g^k(q_k) = q_{2k}$. Thus (3.28) yields

$$q_{3k} \equiv q_{2k} \bmod m_k \tag{3.29}$$

and by induction $q_{lk} \equiv q_{2k} \bmod m_k$ for all $l \geq 2$, as desired.

We now have

$$\Psi_n = \prod_{d|n} q_d^{\mu(n/d)} = \prod_{t|n'} q_{kt}^{\mu(n'/t)} \equiv (-1)^{\mu(n')} \prod_{t|n'} q_{2k}^{\mu(n'/t)} \bmod m_k.$$

Since $\sum_{t|n'} \mu(n'/t) = 0$ and $\mu(n') = \pm 1$, the rightmost expression above is just $-1 \bmod m_k$. But -1 is not a square modm_k, so Ψ_n is not a square in A. Hence Ψ_n is not a square in K. ■

To make the line of reasoning begun by Lemma 3.42 work, we naturally need to find elements m_n. We propose $m_n = q_n + q_{n+1}$. This choice automatically satisfies property 2 of Lemma 3.42. Indeed, if a prime $\mathfrak{p}$ divides $q_n + q_{n+1}$, then clearly we have $q_{n+1} \equiv -q_n \bmod \mathfrak{p}$. Moreover, as in the proof of Lemma 3.42,

$$q_{n+2} = g(q_{n+1}) \equiv g(-q_n) = g(q_n) = q_{n+1} \equiv -q_n \bmod \mathfrak{p},$$

so $\mathfrak{p}$ divides $q_n + q_{n+2}$. It follows that $q_n + q_{n+1} \mid q_n + q_{n+2}$. Repeating this argument $n-2$ times yields $q_n + q_{n+1} \mid q_n + q_{2n}$.

We now specialize to the case $A = \mathbb{F}_p[x]$. The next result gives a condition under which the choice $m_n = q_n + q_{n+1}$ satisfies properties 1 and 3 of Lemma 3.42.

Lemma 3.43. *Suppose that $g(0) = \pm 1$, $\deg q_n$ is odd for $n \geq 2$, and -1 is not a square in $\mathbb{F}_p$. Then Ψ_n is not a square in $\mathbb{F}_p(x)$ for all $n \geq 2$.*

Proof: Let $m_n = q_n + q_{n+1}$. Note that modulo q_n we have $q_{n+1} = g(q_n) \equiv g(0) = \pm 1$. Therefore q_n and q_{n+1} are relatively prime, and thus q_n and m_n are relatively prime as well. We now need only establish that -1 is not a square mod m_n. Since $\deg q_n$ is odd for $n \geq 2$, we have $\deg q_2 > 0$, and thus for $n \geq 1$ we have $\deg m_n = \deg q_{n+1}$, which is odd. Thus m_n has at least one irreducible factor s of odd degree. By assumption -1 is not a square in $\mathbb{F}_p$, and it follows from [32, Proposition 1.10] that -1 is not a square mod s. Hence -1 cannot be a square mod m_n. Lemma 3.42 then shows that Ψ_n is not a square in $\mathbb{F}_p(x)$ for all $n \geq 2$. ■

Theorem 3.44. *If $p \equiv 3 \bmod 4$, then Φ_n is not a square in $\mathbb{F}_p(x)$ for each $n \geq 1$*

Proof: Recall that we defined $p_1 = x$, $p_n = p_{n-1}^2 + x$, and $\Phi_n = \prod_{d|n} p_d^{\mu(n/d)}$. Let $g = xy^2 + 1$. Clearly $\deg g(a)$ is odd for any $a \in \mathbb{F}_p[x]$, so the hypotheses of Lemma 3.43 are satisfied. Now $xq_n = (xq_{n-1})^2 + x$, and $xq_1 = x$, so we have $xq_n = p_n$ for all n. Thus

$$\Psi_n = \prod_{d|n} q_d^{\mu(n/d)} = \prod_{d|n} x^{\mu(n/d)} \prod_{d|n} q_d^{\mu(n/d)} = \prod_{d|n} p_d^{\mu(n/d)} = \Phi_n.$$

Lemma 3.43 then gives that Φ_n is not a square in $\mathbb{F}_p(x)$ for $n \geq 2$. Clearly $\Phi_1 = x$ is not a square in $\mathbb{F}_p(x)$. ■

The final assertion of Theorem 3.2 now follows immediately from Theorems 3.44 and 3.38.

Chapter 4

Construction of the Galois Process for f and Probabilistic Background

The group-theoretic results presented in Chapter 3 are in some sense the engine of this thesis. However, unless it is attached to some surrounding apparatus – imagine gears, axles, and wheels if you wish – this engine does not result in forward motion. In this chapter we construct a surrounding apparatus, and in the next chapter we show that forward motion indeed occurs: enough motion to prove our main result.

Recall the setup of Chapter 3: let F be a field of characteristic $\neq 2$, let $K = F(x)$, and set $f = y^2 + x \in K[y]$. Let K_n be the splitting field over K of the nth iterate f^n, and put $G_n = \mathrm{Gal}\,(K_n/K)$ and $H_n = \mathrm{Gal}\,(K_n/K_{n-1})$. The surrounding apparatus mentioned in the previous paragraph emerges from probability theory. To understand why this is natural, recall that G_n acts on the set $\mathcal{R}_n$ of roots of f^n, and we seek information about the number of $g \in G_n$ that fix at least one element of $\mathcal{R}_n$ (see Theorem 2.18). The results of Chapter 3 lead directly to information about the fixed points in $\mathcal{R}_n$ of elements in H_n. To use this we need to consider the partitioning of G_n into cosets of H_n. Specifically, if $g_0 \in G_n$ fixes t elements of R_{n-1}, we consider the function

$$\begin{aligned} \phi : g_0 H_n &\rightarrow \mathbb{Z}^+ \\ g &\mapsto \text{number of fixed points of } g \text{ in } \mathcal{R}_n \end{aligned} \tag{4.1}$$

If n is squarefree, we use Theorem 3.2 to determine $\#\phi^{-1}(u)$ explicitly for any u (see Proposition 5.6). For general n, Corollary 3.23 allows us to establish the average value of ϕ (see Theorem 5.3). Both of these statements are more natural if we think of ϕ as a random variable and $g_0 H_n$ as a probability space with the uniform distribution. The first statement gives the distribution of ϕ, while the

second gives the expectation of ϕ. The framework of probability theory also gives us powerful tools for extending results on the behavior of ϕ.

Before we can bring to bear the tools of probability, we must describe the probability space in which we are working. We wish to find a space with a random variable X_n whose distribution resembles that of ϕ above:

$$\mathbf{P}(X_n = t) = \frac{1}{\#G_n}\#\{g \in G_n : g \text{ fixes } t \text{ elements of } \mathcal{R}_n\}. \tag{4.2}$$

One method of doing this is to let G_n be the underlying space, give it the uniform probability distribution, and let X_n be defined similarly to ϕ. Then if we condition X_n on the event g_0H_n we obtain the same probability distribution as that of ϕ.

We are interested, however, in limiting behavior as n grows, so this approach is insufficient. We take this opportunity to recall that a sequence $X_0, X_1, X_2, \ldots$ of random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is known as a *discrete-time stochastic process*, which we abbreviate simply to *process*. To accomodate the behavior of G_n for all n, we need a process that satisfies (4.2) for each n. Indeed, for the applications we have in mind, we need a process satisfying a stronger property, namely

$$P(X_0 = t_0, \ldots, X_n = t_n) = \frac{1}{\#G_n}\#\left\{g \in G_n : g \text{ fixes } t_i \text{ elements of } \mathcal{R}_i \text{ for } i = 0, 1, \ldots, n\right\} \tag{4.3}$$

for any $n \geq 0$ and any nonnegative integers $t_0, \ldots, t_n$.

A process can be thought of as a game of chance, with X_n denoting a gambler's score at turn n. Consider a process that satisfies (4.3); thus our gambler is playing a "G_n game" in that her chances of gaining or losing points at each turn are determined by the structure of G_n. By Theorem 2.18 we have

$$\delta(\mathcal{I}_n) = \mathbf{P}(X_n > 0), \tag{4.4}$$

where $\mathcal{I}_n$ is the set of $\alpha \in \overline{\mathbb{F}}_p$ such that 0 has an nth preimage in $\mathbb{F}_p(\alpha)$ under iteration of $x^2 + \alpha$. By Corollary 2.6, to achieve our main goal of showing $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$ we need only show that $\lim_{n\to\infty} \delta(\mathcal{I}_n) = 0$. Thus we wish to find $\lim_{n\to\infty} \mathbf{P}(X_n > 0)$, namely the probability that the gambler never goes bankrupt. And we are rooting against the gambler, as we wish to show that bankruptcy is certain.

The structure of H_n enters into this process via conditional distributions. The computation of a conditional probability such as $\mathbf{P}(X_n = t_n \mid X_{n-1} = t_{n-1})$ amounts to computing the proportion of elements of the set $S = \{g \in G_n : g \text{ fixes } t_{n-1} \text{ elements of } \mathcal{R}_{n-1}\}$ that fix t_n elements of $\mathcal{R}_n$. Elements of H_n by definition induce the identity permutation on $\mathcal{R}_{n-1}$, so $g \in S$ if and only if

$gh \in S$. It follows that S is a union of cosets of H_n. We may then apply analyses like the one described for the function ϕ defined in (4.1).

In this chapter we construct a process that satisfies (4.3); this occupies the first two sections. In sections 3 and 4, we give some definitions and basic properties of certain well-known kinds of processes that play a role in Chapter 5.

4.1 Construction of the Galois Process for f

In this section we construct a process that satisifes equation (4.3). We refer to such a process as the *Galois process associated to iterates of* f (we use *the* instead of *a* because uniqueness is immaterial as long as it satisfies (4.3)). We abbreviate this to the Galois process for f, or simply GP(f).

Constructing probability spaces with specified properties can be highly non-trivial, and it is often impossible to do so explicitly. However, the Daniell-Kolmogorov Extension theorem allows one to show that many useful spaces exist without having to construct them. This is the tool that I originally used to show the Galois process exists, and it remains the "official" method in that the proofs of Chapter 5 are all written to conform to it. Late in the process of doing my thesis research, however, I discovered a much nicer and more intuitive explicit construction, which is briefly described. If the reader is willing to accept that the proofs of Chapter 5 may be safely translated into the language of the explicit construction, then he/she can skip over the "official" method.

We begin with the explicit construction, which allows one to get more of a flavor for what is going on and provides an intuition helpful for understanding the results of Chapter 5. Since K_{n-1} is a Galois sub-extension of K_n, there is a natural quotient homomorphism $\psi_n : G_n \to G_{n-1}$. We choose as our underlying space the inverse limit G of the G_n with respect to these maps. Thus $g \in G$ looks like

$$(g_1, g_2, g_3, \ldots),$$

where $g_n \in G_n$ and the sequence is coherent in that $\psi_n(g_n) = g_{n-1}$. We have obvious nth factor projections $\pi_n : G \to G_n$. We give G the standard pro-finite topology, where the open sets are generated by the collection $\{\pi_n^{-1}(g_n) : g_n \in G_n, n \in \mathbb{N}\}$. One can easily show that G is compact with this topology. We take as our σ-algebra the Borel sets $\mathcal{B}$, and for our probability measure we take the Harr measure $\mathbf{P}$, which by definition is invariant under translation (i.e. left-multiplication). Since G is compact, we may normalize $\mathbf{P}$ so that $\mathbf{P}(G) = 1$. The projections π_n are continuous in our topology, so the sets $\pi_n^{-1}(g_n)$ are measurable for any $g_n \in G_n$. Note also that

$$G = \bigsqcup_{g_n \in G_n} \pi_n^{-1}(g_n), \tag{4.5}$$

where the union is disjoint. Clearly if $\pi_n(g) = g_n$ then $\pi_n^{-1}(g_n) = g\pi_n^{-1}(e)$, so the sets $\pi_n^{-1}(g_n)$ are translates of one another. Thus (4.5) and the translation invariance of $\mathbf{P}$ give

$$1 = \sum_{g_n \in G_n} \mathbf{P}(\pi_n^{-1}(g_n)),$$

and from this we get

$$\mathbf{P}(\pi_n^{-1}(g_n)) = 1/\#G_n \tag{4.6}$$

for all $g_n \in G_n$.

Finally, for each n we define

$$X_n(g) = \text{number of fixed points of } \pi_n(g) \text{ in } \mathcal{R}_n.$$

Recall that $\psi(\pi_n(g)) = \pi_{n-1}(g)$, where $\psi : G_n \to G_{n-1}$ is the quotient homomorphism with kernel H_n. Thus $\pi_n(g)$ induces the same permutation on $\mathcal{R}_{n-1}$ as $\pi_{n-1}(g)$. By induction we have that $\pi_n(g)$ and $\pi_i(g)$ have the same action on $\mathcal{R}_i$ for any $i \leq n$. Thus for any $i \leq n$, $X_i(g)$ is the same as the number of fixed points in $\mathcal{R}_i$ of $\pi_n(g)$. It follows that $X_0(g) = t_0, \ldots, X_n(g) = t_n$ if and only if $\pi_n(g)$ fixes t_i elements of $\mathcal{R}_i$ for $0 \leq i \leq n$. From (4.6) we then immediately get that the process $(G, \mathcal{B}, \mathbf{P}, (X_n)_{n\geq 0})$ satisfies (4.3).

We now give the "official" method of proving the existence of the Galois process, which makes use of the well-known Daniell-Kolmogorov Extension theorem. Suppose that we have a σ-algebra $\mathcal{E}$ on a set E and for each n a probability measure μ_n defined on the n-fold product of $(E, \mathcal{E})$. The Daniell-Kolmogorov Theorem says that, provided the μ_n meet an obvious consistency condition, there is a probability μ defined on an infinite product of $(E, \mathcal{E})$ that extends each of the μ_n. The Theorem also covers continuous-time processes, but for our purposes it is enough to stick with the discrete-time case. In this section we state the discrete-time version of the theorem in detail, while in the next section we show how it can be used to construct the process we seek.

Let E be a set and $\mathcal{E}$ a σ-algebra on E. Suppose that for each n we have a probability measure μ_n defined on the product space $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$, and suppose that the μ_n are compatible in the following way: for any n and any $A_i \subset E, i = 0, 1, 2, \ldots, n-1$, we have

$$\mu_{n-1}\left(A_0 \times A_1 \times \cdots \times A_{n-1}\right) = \mu_n\left(A_0 \times A_1 \times \cdots \times A_{n-1} \times E\right). \tag{4.7}$$

Now let us introduce the *canonical space* $(\Omega, \mathcal{F}, (X_n)_{n\geq 0})$ defined as follows:

$$\Omega = E^{\mathbb{N}}, \qquad \omega \in \Omega \text{ is denoted } (\omega_n)_{n\geq 0}, \qquad X_n(\omega) = \omega_n \text{ for every } n.$$

For each $n \geq 0$, define $\mathcal{F}_n$ to be the σ-algebra generated by all sets of the the form $X_i^{-1}(A)$, where $A \in \mathcal{E}$ and $0 \leq i \leq n$. We denote this $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Note that $X_i^{-1}(A)$ is simply

$E \times \cdots \times E \times A \times E \times \cdots$, with the A being in the ith position. Thus $\mathcal{F}_n$ consists of all sets of the form $B \times E \times E \times \cdots$, where $B \in \mathcal{E}^{\otimes(n+1)}$. Finally, set $\mathcal{F} = \sigma(X_0, X_1, X_2, \ldots) = \sigma(\bigcup_{n=0}^{\infty} \mathcal{F}_n)$.

We may use μ_n to define a probability $\mathbf{P}_n$ on $(\Omega, \mathcal{F}_n)$ by taking, for every $B \in \mathcal{E}^{\otimes(n+1)}$,

$$\mathbf{P}_n(B \times E \times E \times \ldots) = \mu_n(B). \tag{4.8}$$

Given this setup, it is natural to ask if there is a probability measure $\mathbf{P}$ defined on all of $\mathcal{F}$ that agrees with every $\mathbf{P}_n$. The Daniell-Kolmogorov Theorem provides an affirmative answer:

Theorem 4.1. *Using the notation above, and supposing that the μ_n satisfy (4.7), there exists a unique probability measure $\mathbf{P}$ on the canonical space $(\Omega, \mathcal{F})$ such that for any $F \in \mathcal{F}_n$, $\mathbf{P}(F) = \mathbf{P}_n(F)$.*

For a proof of Theorem 4.1, see [39, page 81]. Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ be the process whose existence is guaranteed in Theorem 4.1. Note that because $\mathbf{P}$ extends $\mathbf{P}_n$, we have by (4.8) that for any $A_i \in \mathcal{E}$, $i = 0, 1, 2, \ldots, n$,

$$\begin{aligned} \mu_n(A_0 \times A_1 \times \cdots \times A_n) &= \mathbf{P}(A_0 \times \cdots \times A_n \times E \times E \times \cdots) \\ &= \mathbf{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n), \end{aligned} \tag{4.9}$$

where the last expression is the usual shorthand for

$$\mathbf{P}\left(X_0^{-1}(A_0) \cap X_1^{-1}(A_1) \cap \cdots \cap X_n^{-1}(A_n)\right).$$

We now use Theorem 4.1 to construct the process we seek. Unfortunately, this construction utilises the structure of G only implicitly, as its underlying space is $\mathbb{N}^{\mathbb{N}}$. Take $E = \mathbb{N}$ and $\mathcal{E} = \mathcal{P}(\mathbb{N})$, the power set of $\mathbb{N}$. We now define the probability measures μ_n for $n \geq 0$.

Let n be given. For each $i \leq n$ we define

$$\begin{aligned} v_{i,n} : G_n &\to \mathbb{N} \\ g &\mapsto \text{number of fixed points of } g \text{ in } \mathcal{R}(f_i). \end{aligned}$$

We make the convention that $\mathcal{R}(f_0) = \{0\}$, so that $v_{0,n}(g) = 1$ for all $g \in G_n$. We now use these maps to define

$$\begin{aligned} v_n : G_n &\longrightarrow \mathbb{N}^{n+1} \\ g &\longmapsto (v_{0,n}(g), v_{1,n}(g), \ldots, v_{n,n}(g)). \end{aligned}$$

Example 4.2. Consider the case $n = 2$; we use the labeling given in Example 2.19. From Theorem 3.2 we know that H_1 and H_2 are maximal, meaning

$$H_2 = \{e, (1\ \ 2), (3\ \ 4), (1\ \ 2)(3\ \ 4)\}$$

and $\#G_2 = 8$. By the definition of H_n, these are the only elements that restict to the identity permutation on $\mathcal{R}(f_1)$. The four elements of G_2 not contained in H_2 must therefore interchange the sets $\{1,2\}$ and $\{3,4\}$ and must have squares that preserve these sets. One can easily show these elements are $(1\ \ 3\ \ 2\ \ 4), (1\ \ 4\ \ 2\ \ 3), (1\ \ 3)(2\ \ 4)$, and $(1\ \ 4)(2\ \ 3)$. ■

We have, for instance,

$$\begin{aligned} v_2\left((1\ \ 2)\right) &= (1,2,2) \\ v_2\left((1\ \ 2)(3\ \ 4)\right) &= (1,2,0) \\ v_2\left(e\right) &= (1,2,4) \\ v_2\left((1\ \ 3)(2\ \ 4)\right) &= (1,0,0) \end{aligned}$$

We pause to enumerate two elementary properties of the v_n: First, since the domain of v_n is finite, we have for all $B_1, B_2 \in \mathcal{P}(\mathbb{N})^{\otimes(n+1)}$,

$$\#v_n^{-1}(B_1 \cup B_2) = \#v_n^{-1}(B_1) + \#v_n^{-1}(B_2) - \#v_n^{-1}(B_1 \cap B_2). \tag{4.10}$$

Second, the finiteness of the G_n also gives us that for each n

$$v_n(G_n) = F_0 \times F_1 \times \cdots \times F_n,$$

where each F_i is a finite set.

At last we are in a position to define μ_n:

$$\begin{aligned} \mu_n : \quad \mathcal{P}(\mathbb{N})^{\otimes(n+1)} &\longrightarrow \mathbb{R}^+ \\ B &\longmapsto \frac{\#{v_n}^{-1}(B)}{\#G_n} \end{aligned} \tag{4.11}$$

Example 4.3. We have

$$\begin{aligned} \mu_2\left(\{1\}\times\{2\}\times\{4\}\right) &= \frac{\#\{e\}}{8} = \frac{1}{8} \\ \mu_2\left(\{1\}\times\{2\}\times\{2\}\right) &= \frac{\#\{(12),(34)\}}{8} = \frac{1}{4} \\ \mu_2\left(\{1\}\times\{0\}\times\{0\}\right) &= \frac{\#\{(1324),(1423),(13)(24),(14)(23)\}}{8} = \frac{1}{2} \\ \mu_2\left(\{1\}\times\{2\}\times\mathbb{N}\right) &= \frac{\#\{e,(12),(34),(12)(34)\}}{8} = \frac{1}{2} \end{aligned}$$

We now show that the μ_n are indeed probability measures. As usual, the condition of countable additivity requires a small bit of work. We give two easy Lemmas that simplify the proof.

Lemma 4.4. *Let B be a subset of $\mathbb{N}^{n+1}$ belonging to $\mathcal{P}(\mathbb{N})^{\otimes(n+1)}$. Then $\mu_n(B) = 0$ if and only if $B \cap v_n(G_n) = \emptyset$.*

Proof: Taking v_n^{-1} of both sides of the equation $B \cap v_n(G_n) = \emptyset$ gives

$$v_n^{-1}(B) \cap G_n = \emptyset. \tag{4.12}$$

(Note that $v_n^{-1}(v_n(G_n)) = G_n$ because $f^{-1}(f(S)) \supseteq S$ in general and $v_n^{-1}(v_n(G_n)) \subseteq G_n$.) But $v_n^{-1}(B) \subseteq G_n$, so (4.12) is equivalent to $v_n^{-1}(B) = \emptyset$. This holds if and only if $\mu_n(B) = 0$. ■

Lemma 4.5. *If B_1 and B_2 are disjoint subsets of $\mathbb{N}^{n+1}$ belonging to $\mathcal{P}(\mathbb{N})^{\otimes(n+1)}$, then $\mu_n(B_1 \cup B_2) = \mu_n(B_1) + \mu_n(B_2)$.*

Proof: Applying v_n^{-1} to both sides of $B_1 \cap B_2 = \emptyset$ gives $v_n^{-1}(B_1 \cap B_2) = \emptyset$. The lemma then follows from equation (4.10). ■

Proposition 4.6. *For each n, μ_n is a probability measure on $(\mathbb{N}^{n+1}, \mathcal{P}(\mathbb{N})^{\otimes(n+1)})$.*

Proof: It is clear from the definition that $0 \leq \mu_n \leq 1$ for all n. Moreover, we obviously have $\mu_n(\mathbb{N}^{n+1}) = \frac{\#G_n}{\#G_n} = 1$. It remains to verify that each μ_n is countably additive: if $B_1, B_2, \ldots$ are pairwise disjoint subsets belonging to $\mathcal{P}(\mathbb{N})^{\otimes(n+1)}$, then

$$\mu_n\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu_n(B_i).$$

Since the B_i are pairwise disjoint and $v_n(G_n)$ is a product of finite sets, there must exist an i_0 such that $\left(\bigcup_{i=i_0+1}^{\infty} B_i\right) \cap v_n(G_n) = \emptyset$. Thus by Lemma 4.4, we have $\mu_n\left(\bigcup_{i=i_0+1}^{\infty} B_i\right) = 0$. This gives us

$$\mu_n\left(\bigcup_{i=1}^{\infty} B_i\right) = \mu_n\left(C \cup \bigcup_{i=1}^{i_0} B_i\right),$$

where $\mu_n(C) = 0$ and $C, B_1, B_2, \ldots, B_{i_0}$ are pairwise disjoint. Using induction and Lemma 4.5, we have

$$\mu_n\left(C \cup \bigcup_{i=1}^{i_0} B_i\right) = \mu_n(C) + \sum_{i=1}^{i_0} \mu_n(B_i) = \sum_{i=1}^{i_0} \mu_n(B_i) = \sum_{i=1}^{\infty} \mu_n(B_i),$$

where the last equality holds since $\mu_n(B_i) = 0$ when $i > i_0$. Thus μ_n is a probability measure. ■

Now that we have established that each μ_n is a probability measure, the only remaining obstacle to applying the Daniell-Kolmogorov Theorem is to show that the μ_n satisfy (4.7).

Proposition 4.7. *The μ_n satisfy the compatibility condition (4.7).*

Proof: We must show that for all $n \geq 1$, we have

$$\mu_{n-1}\left(A_0 \times A_1 \times \cdots \times A_{n-1}\right) = \mu_n\left(A_0 \times A_1 \times \cdots \times A_{n-1} \times \mathbb{N}\right),$$

where $A_i \subseteq \mathbb{N}$ for each i. Recall that H_n is the Galois group of K_n/K_{n-1}, and the quotient map $\psi_n : G_n \to G_{n-1}$ gives an isomorphism between G_n/H_n and G_{n-1}. Now fix $g \in G_n$. We have $v_n(g) \in A_0 \times A_1 \times \cdots \times A_{n-1} \times \mathbb{N}$ if and only if $v_{n,i}(g) \in A_i$ for $i = 0, 1, 2, \ldots, n-1$ (note that the value of $v_{n,n}$ is irrelevant). This last condition is equivalent to $v_{n-1,i}(\psi_n(g)) \in A_i$ for $i = 0, 1, 2, \ldots, n-1$, i.e. $v_{n-1}(\psi_n(g)) \in A_0 \times A_1 \times \cdots \times A_{n-1}$. Thus we may rewrite $\mu_n\left(A_0 \times A_1 \times \cdots \times A_{n-1} \times \mathbb{N}\right)$ as

$$\frac{1}{\#G_n} \#\{g \in G_n \mid \psi_n(g) \in v_{n-1}^{-1}\left(A_0 \times A_1 \times \cdots \times A_{n-1}\right)\} \tag{4.13}$$

By definition we have $\psi_n(g') = \psi_n(g)$ if and only if $g' \in gH_n$, so (4.13) is equal to

$$\frac{\#H_n}{\#G_n} \#\{gH_n \in G_n/H_n \mid (\psi_n(g)) \in v_{n-1}^{-1}(A_0 \times A_1 \times \cdots \times A_{n-1})\},$$

and using $\#G_n = \#H_n \cdot \#G_{n-1}$, we have that this is equal to

$$\frac{1}{\#G_{n-1}} \#\{g \in G_{n-1} \mid g \in v_{n-1}^{-1}(A_0 \times A_1 \times \cdots \times A_{n-1})\},$$

which is $\mu_{n-1}\left(A_0 \times A_1 \times \cdots \times A_{n-1}\right)$ ■

Thanks to Propositions 4.6 and 4.7, we may apply Theorem 4.1 to show that there exists a process $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n \geq 0})$ such that

$$\mathbf{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n) = \mu_n(A_0 \times A_1 \times \cdots \times A_n) \tag{4.14}$$

for any $A_i \in \mathbb{N}$ (see discussion following Theorem 4.1). In particular, for any n and any nonnegative integers $t_0, \ldots, t_n$, we have

$$\begin{aligned} \mathbf{P}(X_0 = t_0, X_1 = t_1, \ldots, X_n = t_n) &= \mu_n(\{t_0\} \times \{t_1\} \times \cdots \times \{t_n\}) \\ &= \frac{1}{\#G_n} \#\left\{g \in G_n \mid v_n(g) \in \{t_0\} \times \{t_1\} \times \cdots \times \{t_n\}\right\} \end{aligned}$$

and this last expression, by the definition of v_n, is simply

$$\frac{1}{\#G_n} \#\left\{g \in G_n : g \text{ fixes } t_i \text{ elements of } \mathcal{R}_i \text{ for } i = 0, 1, \ldots, n\right\}.$$

This establishes (4.3), which is the principal property we desired our process to have.

We close with the remark, which is helpful in the proofs of Chapter 5, that because $\mathcal{R}_0 = \{0\}$, we have $\mathbf{P}(X_0 = 1) = 1$.

4.2 Martingales and Markov Chains

Since GP(f) has random variables that take their values in $\mathbb{Z}$, we now restrict our attention mainly to such processes. This section is devoted to an important class of processes known as martingales.

We give a somewhat restricted definition of these, which is sufficient for our purposes. The concept of martingale draws its inspiration from a gambler playing a fair game (i.e. one with expectation 0) repeatedly. Let X_n denote the gambler's gain at game n. The gambler chooses her stakes according to some rule involving the outcomes of previous games, so the X_n are not independent. However, the knowledge of past games should not affect the fairness of future games. Thus regardless of the values of $X_0, X_1, \ldots, X_n$, the expectation of X_{n+1} should be 0. To formulate this rigorously, we give a brief review of the notion of conditional expectation.

First recall that if $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space and X a random variable taking values in $\mathbb{Z}$, then the *expectation* of X is

$$E(X) = \sum_{k \in \mathbb{Z}} k \cdot \mathbf{P}(X = k),$$

where $X = k$ denotes the set $X^{-1}(k)$. We say X is *integrable* if $E(X)$ is finite.

Consider now two random variables X and Y taking values in $\mathbb{Z}$. We define the familiar conditional probability

$$\mathbf{P}(Y = k \mid X = t) = \mathbf{P}(Y = k \cap X = t)/\mathbf{P}(X = t).$$

The conditional expectation $E(Y \mid X = t)$ is then just the expectation defined above but using the conditional probability:

$$E(Y \mid X = t) \;=\; \sum_{k \in \mathbb{Z}} k \cdot \mathbf{P}(Y = k \mid X = t) \;=\; \frac{1}{\mathbf{P}(X = t)} \sum_{k \in \mathbb{Z}} k \cdot \mathbf{P}(Y = k \cap X = t).$$

We set $h(t) = E(Y \mid X = t)$ for $t \in \mathbb{Z}$. We may then compose h with X to get a random variable $E(Y \mid X)$ that maps ω to $E(Y \mid X = X(\omega))$.

We can repeat this construction with multiple variables as follows: let $Y, X_1, \ldots, X_n$ be $\mathbb{Z}$-valued, and define

$$\mathbf{P}(Y = k \mid X_1 = t_1, \ldots, X_n = t_n) = \frac{\mathbf{P}\left(Y = k \cap \bigcap_{i=1}^n X_i = t_i\right)}{\mathbf{P}\left(\bigcap_{i=1}^n X_i = t_i\right)}.$$

Analogously to the single-variable case, we define $E(Y \mid X_1 = t_1, \ldots, X_n = t_n)$ to be

$$\sum_{k \in \mathbb{Z}} k \cdot \mathbf{P}(Y = k \mid X_1 = t_1, \ldots, X_n = t_n). \tag{4.15}$$

Then we define the random variable $E(Y \mid X_1, \ldots, X_n)$ that takes ω to

$$E(Y \mid X_1 = X_1(\omega), \ldots, X_n = X_n(\omega)).$$

We now give the definition of a martingale. In the discussion at the beginning of this section of a gambler playing repeatedly a fair game, we let the X_n be the gambler's gain at turn n. The random variables in the following definition may be thought of as the gambler's *accumulated* gains through turn n.

Definition 4.8. *Let $S = (\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ be a stochastic process where each X_n takes values in $\mathbb{Z}$ and is integrable. We say S is a* martingale *if for every $n \geq 0$ we have*

$$E(X_{n+1} \mid X_0, \ldots, X_n) = X_n.$$

We note that this definition is less general than the one presented in many probability texts[1] (see e.g. [4]).

We say that a martingale $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ *converges* if

$$\mathbf{P}\left(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) \text{ exists}\}\right) = 1.$$

Martingales have achieved much of their usefulness because they often converge. The following is a standard theorem (see e.g. [7, page 71])

Theorem 4.9. *Let $M = (\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ be a martingale with $X_n \geq 0$ for each n (X_n need not be integer valued). Then M converges. We set $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$, which is defined except on a set of probability* 0.

We end this section with a statement of the definition of a Markov chain, a concept which we use less than martingales but which nonetheless plays a role in Chapter 5 in the case where H_n is maximal for all n. Loosely, a Markov chain is an infinite sequence of random variables, indexed by what can be thought of as a time parameter, where only information from the previous variable may affect the values of the next variable. Any sequence of independent variables certainly qualifies; Markov chains can be thought of as one level more complicated than such sequences.

Definition 4.10. *Let $S = (\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ be an integer-valued stochastic process. We say S is a* Markov chain *if for any natural numbers $m_1 < m_2 < \cdots < m_k$ and integers $s_1, s_2, \ldots, s_k$ such that $\mathbf{P}(X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1}) > 0$, we have*

$$\mathbf{P}(X_{m_k} = s_k \mid X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1}) = \mathbf{P}(X_{m_k} = s_k \mid X_{m_{k-1}} = s_{k-1}).$$

[1] The conditioning $E(X_{n+1} \mid X_0, \ldots, X_n)$ takes into account only the values of previous X_n, that is, the gambler's past gains. Thus in using this form of conditioning in our definition, we are assuming the gambler has knowledge only of her past gains. (Another way of saying this is that her knowledge of the past consists only of events of the form $X_0^{-1}(A_0) \cap \cdots \cap X_n^{-1}(A_n)$ for Borel sets $A_0, \ldots, A_n$, the collection of which is known as the σ-algebra generated by $X_0, \ldots, X_n$.) But there may be other information available about the past. For instance, if the gambler skips game 1, then $X_1 = 0$ with certainty. Yet knowledge of what happened in game 1 could in principle be useful in future predictions. The most general definition of a martingale allows us to stipulate the sophistication of the gambler's knowledge of the past; in our definition above, we only allow the gambler the barebones knowledge of the past gains at each game. To state the more general version, we must expand our notion of conditional expectation to arbitrary σ-algebras, not merely the one generated by $X_0, \ldots, X_n$. Once this is done, the general definition of a martingale is with respect to an increasing sequence of σ- algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$, with each X_n being $\mathcal{F}_n$-measurable and $E(X_{n+1} \mid \mathcal{F}_n) = X_n$.

4.3 Branching Processes

In this section, we discuss a certain kind of Markov chain that lends itself to a particularly neat analysis. We assume throughout the section that all Markov chains have $X_0 = 1$ with probability 1. This assumption is easily dispensed with, but improves clarity and is sufficient for all of our applications.

An important characteristic of a Markov chain is its *transition probabilites*, namely

$$p_{ij,n} = \mathbf{P}(X_n = j \mid X_{n-1} = i).$$

Note that $p_{ij,n}$ is undefined when $\mathbf{P}(X_{n-1} = i) = 0$. In the case where $p_{ij,n}$ depends only on i and j, we call the Markov chain *time homogeneous*. We refer to a Markov chain as *non-negative* when its random varibles take on only non-negative integer values. The following definition is adapted from [16]. We give a slighly more complicated definition that suits our purposes in Chapter 5.

Definition 4.11. *Suppose $M = (\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ is a non-negative, time-homogeneous Markov chain. We call M a* branching process *if for each $i \geq 0$ such that*

$$\mathbf{P}(X_{n-1} = i) > 0,$$

the transition probabilities $\{p_{ij}\}_{j\in\mathbb{N}}$ have the same distribution as the sum of i independent variables, each with the same distribution as X_1. We take the sum of 0 random variables to have value 0 with probability 1.

The study of such processes began with an 1874 paper by Francis Galton and the Reverend H.W. Watson [38] (see [16] for a summary), and some authors use "Galton-Watson processes" instead of "branching processes". Galton proposed the problem of determining the likelihood of eventual extinction of a family line, as measured by the number of males in each generation. He assumed a single progenitor at generation 0, and also made the crucuial assumption that the probability distribution of sons is identical for every man in every generation. Such a scenario is modeled by a branching process, with X_n denoted the size of the nth generation.

As another illustration, we can think of a branching process as a game with the following rules. The house has an infinite supply of tickets, each having a number k written on the back, with k chosen at random according to the probability distribution of X_1. To begin the game you are given one ticket (this corresponds to $X_0 = 1$). You then trade in your ticket to the house; in return you receive a number of tickets equal to the number on the back of your original ticket. At the next round you trade in all your tickets and receive a number of new tickets equal to the sum of the numbers on the backs of all your previous tickets. Continue in this fashion.

As a side note, one can turn this into an amusing parlor game by giving each player the option of stopping at any given round and keeping all her tickets. Subsequent players then try to beat her score. I have tried this with X_1 such that $\mathbf{P}(X_1 = 3) = 1/6$, $\mathbf{P}(X_1 = 1) = 1/2$, and $\mathbf{P}(X_1 = 0) = 1/3$; the results were highly entertaining.

An important general property of a branching process is that 0 is an absorbing state provided that $\mathbf{P}(X_1 = 0) > 0$. By this we mean that for any m, $\mathbf{P}(X_{m+1} = 0, X_{m+2} = 0, \ldots \mid X_m = 0) = 1$. This is suggested by the above discussion: if you run out of tickets at any round, you have no hope of ever getting any more. From the definition and the assumption that $\mathbf{P}(X_1 = 0) > 0$ we have by an easy induction that $\mathbf{P}(X_n = 0) > 0$ for all $n \geq 1$, and $\mathbf{P}(X_n = 0 \mid X_{n-1} = 0) = 1$ for all $n \geq 2$. Since a Galton- Watson process is a Markov chain, this means that for any $1 \leq m < n$,

$$\mathbf{P}(X_n = 0 \mid X_m = 0, \ldots, X_{n-1} = 0) = 1,$$

which is the same as

$$\mathbf{P}(X_m = 0, \ldots, X_{n-1} = 0, X_n = 0) = \mathbf{P}(X_m = 0, \ldots, X_{n-1} = 0).$$

Repeated application of this gives

$$\mathbf{P}(X_m = 0, \ldots, X_{n-1} = 0, X_n = 0) = \mathbf{P}(X_m = 0).$$

Therefore $\mathbf{P}(X_{m+1} = 0, \ldots, X_n = 0 \mid X_m = 0) = 1$. Note that $\mathbf{P}(\cdot \mid X_m = 0)$ is a probability measure on Ω. We need now only show that the intersection of an increasing sequence of probability 1 events again has probability 1. To do this we employ the very useful property of continuity of probability measures (see e.g. [15, pages 8,9]):

Proposition 4.12. *Suppose we have events $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ in a probability space. Then*

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Similarly, if $B_1 \subseteq B_2 \subseteq B_3 \subseteq \cdots$, then we have

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbf{P}(B_n).$$

We now have

$$\mathbf{P}(X_{m+1} = 0, X_{m+2} = 0, \ldots \mid X_m = 0) = 1, \tag{4.16}$$

as desired.

From now on we make the assumption that $\mathbf{P}(X_1 = 0) > 0$, as this is sufficient for our applications. In Watson and Galton's original paper, Watson proposed an ingenious solution to Galton's

problem on family lines. The solution centered on the idea of probability generating functions. The *probability generating function* (or pgf for short) of a random variable X taking non-negative integer values is defined as the complex function

$$g(z) = \sum_{k=0}^{\infty} \mathbf{P}(X = k) z^k.$$

We often write $p_k = \mathbf{P}(X = k)$, so $g(z) = \sum p_k z^k$. Note that $g(z)$ converges for $|z| \leq 1$ because $\sum p_k = 1$.

The following Proposition, taken from [15], is quite useful:

Proposition 4.13. *Suppose $Y_1, \ldots, Y_n$ are independent with respective pgfs $g_1, \ldots, g_n$. Then the random variable $Y_1 + Y_2 + \cdots + Y_n$ has pgf $g_1 g_2 \cdots g_n$.*

Now let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n \geq 0})$ be a branching process. Let g be the pgf of X_1. The pgf of a random variable depends only on its distribution. Thus under the condition $X_{n-1} = i$, Proposition 4.13 and Definition 4.11 tell us X_n has the pgf $(g(z))^i$. Note that this holds even for $i = 0$. This leads to Watson's main result:

Proposition 4.14. *Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n \geq 0})$ be a branching process, let g be the pgf of X_1, and let g_n be the pgf for X_n (without conditioning). Then g_n is the nth iterate of g.*

Proof: We make use of the fact that the events $\{X_{n-1} = i\}$ partition Ω, so that if A is any event, $\mathbf{P}(A) = \sum_i \mathbf{P}(A \cap \{X_{n-1} = i\})$. Obviously we may restrict this sum to the i such that $\mathbf{P}(X_{n-1} = i) > 0$. We do this in the following computation:

$$\begin{aligned}
g_n &= \sum_k \mathbf{P}(X_n = k) z^k \\
&= \sum_k \sum_i \mathbf{P}(X_n = k, X_{n-1} = i) z^k \\
&= \sum_k \sum_i \mathbf{P}(X_{n-1} = i) \mathbf{P}(X_n = k \mid X_{n-1} = i) z^k \\
&= \sum_i \mathbf{P}(X_{n-1} = i) \sum_k \mathbf{P}(X_n = k \mid X_{n-1} = i) z^k
\end{aligned}$$

We remarked just prior to this Proposition that $\sum_k \mathbf{P}(X_n = k \mid X_{n-1} = i) z^k = (g(z))^i$. Thus we have

$$g_n = \sum_i \mathbf{P}(X_{n-1} = i)(g(z))^i = g_{n-1}(g(z)).$$

The Proposition then follows by induction. ■

We now turn our attention to the probability e_n of extinction by generation n, which we define as

$$\mathbf{P}(X_n = 0, X_{n+1} = 0, X_{n+2} = 0, \ldots).$$

Clearly $\{e_n\}_{n\in\mathbb{N}}$ is an increasing sequence. By (4.16), we have $e_n = \mathbf{P}(X_n = 0)$. We can thus apply Proposition 4.14, which shows that

$$e_n = g^n(0), \tag{4.17}$$

where the superscript indicates iteration. Note that $e_n = g(e_{n-1})$, and also $e_0 = 0$. One may easily show using Proposition 4.12 that the probability of eventual extinction is $e = \lim_{n\to\infty} e_n$, which exists because $e_n \leq 1$ for all n. We clearly have $g(e) = e$ (because g is continuous on $0 \leq z \leq 1$). Moreover, e is the limit of an increasing sequence whose first element is 0. A short inductive argument (see [15]) shows:

Proposition 4.15. *Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ be a branching process, and g the pgf of X_1. Then e is the smallest non-negative solution of the equation $g(x) = x$.*

One can also show (see [15] again) that $e = 1$ if and only if $E(X_1) \leq 1$. In the case $E(X_1) < 1$, we have the following trivial estimate for the rate at which e_n approaches 1:

$$e^n \geq 1 - (E(X_1))^n .$$

When $E(X_1) = 1$, matters are more complicated. The following result is originally due to Kolmogorov and can be found in [16, page 21].

Theorem 4.16. *Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ be a branching process with $E(X_1) = 1$, and let g be the pgf for X_1. Then*

$$\lim_{n\to\infty} \left(\mathbf{P}(X_n > 0) - \frac{2}{ng''(1)} \right) = 0.$$

Thus when $E(X_1) = 1$, we have $e_n \sim 1 - 2/(ng''(1))$.

We close our examination of branching processes with a link to martingales.

Proposition 4.17. *Let $B = (\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$ be a branching process, and suppose that $E(X_1) = 1$. Then W is a martingale.*

Proof: The expectation of a random variable depends only on its distribution. By the additivity of expectations (see [15]), if $Z_1, \ldots, Z_m$ are identically distributed random variables then

$$E(Z_1 + \cdots + Z_m) = mE(Z_1).$$

Hence by definition 4.11 we have

$$E(X_n | X_{n-1} = i) = iE(X_1) = i.$$

Therefore $E(X_n | X_{n-1}) = X_{n-1}$.

Now by definition B is a Markov chain, so

$$\mathbf{P}(X_n = s_n \mid X_0 = s_0, \ldots, X_{n-1} = s_{n-1}) = \mathbf{P}(X_n = s_n \mid X_{n-1} = s_{n-1})$$

for any $s_0, s_1, \ldots, s_n$ in $\mathbb{Z}$. By the definition of conditional expectation (4.15), this implies

$$E(X_n \mid X_0, \ldots, X_{n-1}) = E(X_n \mid X_{n-1}),$$

and the proposition follows. ■

Chapter 5

The Threads Come Together

In this chapter, we use the group-theoretic insight gained in Chapter 3 to determine the long-run behavior of the Galois process for f, which we refer to as $\mathrm{GP}(f)$ (see Section 4.1 for definitions). Specifically, we show the behavior of this process is intimately linked to the structure of the groups H_n, and the main results of Chapter 3 give us information about these. The most important result of the present chapter is Corollary 5.4, which draws on Corollary 3.23 to establish that $\mathrm{GP}(f)$ is a martingale. It then follows (see Theorem 4.9) that $\mathrm{GP}(f)$ converges. We next use Theorem 3.2 to establish explicitly the behavior of X_n for all n such that H_n is maximal. This allows us to show that $\mathrm{GP}(f)$ converges to 0 with probability 1.

Applying the convergence to 0 of $\mathrm{GP}(f)$ in the case $F = \mathbb{F}_p, p \neq 2$, and using Theorem 2.18 allows us to show that $\lim_{n\to\infty} \delta(\mathcal{I}_n) = 0$. Recall that δ is Dirichlet density (see (1.6)) and

$$\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset\},$$

namely the set of $\alpha \in \overline{\mathbb{F}}_p$ such that 0 has an nth preimage in $\mathbb{F}_p(\alpha)$ under iteration of $f_\alpha = x^2 + \alpha$. Corollary 2.6 then shows that $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$, which is Theorem 1.7, our main result. We end the chapter with an exploration of some of the nice consequences one can derive from assuming Conjecture 3.1, namely that H_n is maximal for all n.

We recall that $K = F(x), \operatorname{char} F \neq 2$. Let $f = y^2 + x$, K_n be the splitting field of f^n over $K, G_n = \mathrm{Gal}\,(K_n/K)$, and $H_n = \mathrm{Gal}\,(K_n/K_{n-1})$. Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0}))$ be $\mathrm{GP}(f)$ as defined in Section 4.1.

5.1 The Galois Process for f is a Martingale

This section builds up to Theorem 5.3, which gives us the information we need to show the GP(f) is a martingale. Before we arrive there, we establish some basic results. The proofs of these results have the same flavor as that of the more complicated arguments of this chapter, including the proof of Theorem 5.3.

We first recall a fact from Chapter 3 that makes frequent appearances in the proofs in this chapter. First, we denote the roots of f^{n-1} by $\mathcal{R}_{n-1}$ and the roots of f^n by $\mathcal{R}_n$. If $\mathcal{R}_{n-1} = \{\beta_1, \ldots \beta_{2^{n-1}}\}$ then

$$\mathcal{R}_n = \left\{\pm\sqrt{-x+\beta_1}, \ldots, \pm\sqrt{-x+\beta_{2^{n-1}}}\right\}. \tag{5.1}$$

Note that in the language of Chapter 3, the partition $\mathfrak{C}$ is simply

$$\left\{\left\{\pm\sqrt{-x+\beta_1}\right\}, \ldots, \left\{\pm\sqrt{-x+\beta_{2^{n-1}}}\right\}\right\} = \{\{\pm r\} : r \in \mathcal{R}_n\}. \tag{5.2}$$

Recall that each $g \in G_n$ permutes the subsets belonging to $\mathfrak{C}$. We make frequent use of the fact that the permutation g induces on $\mathcal{R}_{n-1}$ is the same as the permutation g induces on $\mathfrak{C}$. In particular, the number of distinct subsets $\{\pm r\}_{r\in\mathcal{R}_n}$ mapped to themselves by g is equal to the number of elements of $\mathcal{R}_{n-1}$ that g fixes.

We begin with a simple result that shows 0 is an absorbing state of GP(f).

Proposition 5.1. *Let* $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0}) = \mathrm{GP}(f)$*, and* $m < n$*. Then*

$$\mathbf{P}(X_n = 0 \mid X_m = 0) = 1.$$

Proof: Using the explicit construction of GP(f) given in Section 4.1, the proposition follows from this observation: if $g \in G_n$ fixes no element of $\mathcal{R}_m$, then there is a complete block system of $\mathcal{R}_n$ such that g maps no block to itself. Hence g fixes no elements of $\mathcal{R}_n$.

In order to be consistent with the "official" construction given in Chapter 4, i.e., the one using the Daniell-Kolmogorov extension theorem, we must jump through a few more hoops. By the definition of conditional probability, we must compute

$$\frac{\mathbf{P}(X_m = 0, X_n = 0)}{\mathbf{P}(X_m = 0)}. \tag{5.3}$$

Let $S = A_0 \times A_1 \times \cdots \times A_n \subset \mathbb{N}^{n+1}$, where

$$A_i = \begin{cases} \{0\} & \text{if } i = m \\ \mathbb{N} & \text{otherwise} \end{cases}$$

Let $S_0 \subset \mathbb{N}^{n+1}$ be defined similarly, except we set $A_n = \{0\}$. Clearly $S_0 \subset S$. From the basic property of GP(f) given in (4.14), the expression in (5.3) is equal to

$$\frac{\mu_n(S_0)}{\mu_n(S)}$$

where μ_n is the function defined in (4.11). From the definition of μ_n this is the same as

$$\frac{\#v_n^{-1}(S_0)}{\#v_n^{-1}(S)}. \tag{5.4}$$

One easily sees that $v_n^{-1}(S)$ is non-empty. Indeed, $G_1 \cong S_2$, so any $g \in G_n$ that extends $(1\ \ 2) \in G_1$ must interchange the blocks of a two-block complete block system of G_i, $i \geq 2$. Thus g has no fixed points in $\mathcal{R}_i$ and it follows that $v_n(g) = (1, 0, 0, \ldots, 0) \in S$.

Now $g \in v_n^{-1}(S)$ if and only if g has no fixed points in $\mathcal{R}_m$, and $v_n^{-1}(S_0)$ consists of those $g \in v_n^{-1}(S)$ that also have no fixed points in $\mathcal{R}_n$. The Proposition then follows from the remark made in the first paragraph of the proof. ■

Since $\mathbf{P}(\,\cdot \mid X_m = 0)$ defines a probability measure on Ω and a countable intersection of probability 1 events again has probability 1, it follows from Proposition 5.1 that

$$\mathbf{P}(X_{m+1} = 0, X_{m+2} = 0, \ldots \mid X_m = 0) = 1.$$

Thus 0 is an absorbing state of GP(f)

The main results of this chapter deal with conditional probabilities of GP(f), as does Proposition 5.1. In our main results, the equivalents of the set $v_n^{-1}(S)$ in the proof of Proposition 5.1 are unions of cosets of H_n. Thus we warm up for the main events with results analyzing a single coset g_0H_n.

Proposition 5.2. *Let $n \geq 1$, and suppose $g_0 \in G_n$ fixes t elements of $\mathcal{R}_{n-1}$. Then the number of elements of $\mathcal{R}_n$ fixed by any $g \in g_0H_n$ is an even integer $2w$ with $0 \leq w \leq t$.*

Proof: By (5.1), $\mathcal{R}_n$ may be partitioned into sets of the form $\{\pm r\}$. Thus the number of elements of $\mathcal{R}_n$ fixed by any $g \in G_n$ is an even integer between 0 and 2^n. Moreover, if $g(r) = r$ for some $r \in \mathcal{R}_n$, then $g(r^2 + x) = r^2 + x$. But from (5.1) one sees that $r^2 + x$ is in R_{n-1}. Thus the number of elements of $\mathcal{R}_n$ fixed by g is at most twice the number of elements of $\mathcal{R}_{n-1}$ fixed by g.

Note that an element of H_n induces the identity permutation on $\mathcal{R}_{n-1}$. Since elements of g_0H_n differ from g_0 only by an element of H_n, they all induce the same permutation on $\mathcal{R}_{n-1}$ as g_0. In particular, they all fix t elements of $\mathcal{R}_{n-1}$. Hence the number of elements of $\mathcal{R}_n$ fixed by any $g \in g_0H_n$ is an even integer $2w$ with $0 \leq w \leq t$. ■

We now give a meatier result on the structure of a single coset g_0H_n, which paves the way for our

result that GP(f) is a martingale. The proof makes fundamental use of Corollary 3.23, providing a reward for the hard work we did in Chapter 3.

Theorem 5.3. *Suppose that $g_0 \in G_n$ fixes t elements of $\mathcal{R}_{n-1}$, and consider $g_0H_n \in G_n/H_n$. For $g \in G_n$, let $s_n(g)$ denote the number of fixed points of g in $\mathcal{R}_n$. Then*

$$\frac{1}{\#H_n} \sum_{g \in g_0 H_n} s_n(g) = t,$$

so that on average $g \in g_0H_n$ fixes t elements of $\mathcal{R}_n$.

Proof: Consider $\gamma \in H_n$ (cf. Corollary 3.23), which we recall maps r to $-r$ for each $r \in \mathcal{R}_n$. Clearly γ has order 2. The group $\{e, \gamma\} = \langle\gamma\rangle$ acts by right multiplication on the set g_0H_n, dividing it into disjoint two-element orbits. We show that for each $g \in g_0H_n$,

$$s_n(g) + s_n(g\gamma) = 2t. \tag{5.5}$$

From this it follows that

$$\sum_{g \in g_0 H_n} s_n(g) = t \cdot \#H_n,$$

which proves the Theorem.

Fix $g \in g_0H_n$, and let $r \in \mathcal{R}_n$. Suppose first that $g(r^2) \neq r^2$. This obviously implies that g fixes no element of $\{\pm r\}$. Since $\gamma(r^2) = r^2$, we have

$$g\gamma(r^2) = g(r^2) \neq r^2.$$

Therefore $g\gamma$ has no fixed points in $\{\pm r\}$.

Now suppose that $g(r^2) = r^2$. Clearly g either fixes both elements of $\{\pm r\}$ or exchanges them. On the other hand, γ exchanges the elements of $\{\pm r\}$. Therefore if g fixes the elements of $\{\pm r\}$ then $g\gamma$ exchanges them, while if g exchanges the elements of $\{\pm r\}$ then $g\gamma$ fixes them. In either case we have

$$\text{(number of fixed points of } g \text{ in } \{\pm r\}) + \text{(number of fixed points of } g\gamma \text{ in } \{\pm r\}) = 2. \tag{5.6}$$

Finally we note that because $g = g_0h$ for some $h \in H_n$, g and g_0 induce the same permutation on $\mathcal{R}_{n-1}$. The number of subsets $\{\pm r\} \subset \mathcal{R}_n$ that are mapped to themselves by g is thus the same as the number mapped to themselves by g_0. By (5.1) this is clearly equal to the number of elements of $\mathcal{R}_{n-1}$ fixed by g_0, namely t. Using this fact and (5.6) we obtain (5.5). ■

The next result uses Theorem 5.3 to establish that GP(f) is a martingale. The proof uses reasoning similar to that of Lemma 5.1.

Corollary 5.4. *GP(f) is a martingale.*

Proof: Denote GP(f) by $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})$. By Definition 4.8 we must show that

$$E(X_n \mid X_0, \ldots, X_{n-1}) = X_{n-1} \text{for all } n.$$

Thus for any integers $t_0, \ldots, t_{n-1}$ such that

$$\mathbf{P}(X_0 = t_0, \ldots, X_{n-1} = t_{n-1}) > 0, \tag{5.7}$$

we must show that

$$E(X_n \mid X_0 = t_0, \ldots, X_{n-1} = t_{n-1}) = t_{n-1}. \tag{5.8}$$

The left-hand side of (5.8) is simply $\sum_k k \cdot \mathbf{P}(X_n = k \mid X_0 = t_0, \ldots, X_{n-1} = t_{n-1})$. Therefore by the definition of conditional probability, (5.8) becomes

$$\sum_k k \cdot \frac{\mathbf{P}(X_0 = t_0, \ldots, X_{n-1} = t_{n-1}, X_n = k)}{\mathbf{P}(X_0 = t_0, \ldots, X_{n-1} = t_{n-1})}. \tag{5.9}$$

Let $S = \{t_0\} \times \{t_1\} \times \cdots \times \{t_{n-1}\} \times \mathbb{N} \subset \mathbb{N}^{n+1}$, and let $S_k = \{t_0\} \times \{t_1\} \times \cdots \times \{t_{n-1}\} \times \{k\} \subset \mathbb{N}^{n+1}$. From the basic property of GP(f) given in (4.14), the expression in (5.9) is equal to

$$\sum_k k \cdot \frac{\mu_n(S_k)}{\mu_n(S)}$$

where μ_n is the function defined in (4.11). From the definition of μ_n this is the same as

$$\sum_k k \cdot \frac{\#v_n^{-1}(S_k)}{\#v_n^{-1}(S)}.$$

This in turn may be rewritten as

$$\frac{1}{\#v_n^{-1}(S)} \sum_k k \cdot \#v_n^{-1}(S_k) \tag{5.10}$$

From (5.7) it follows that $v_n^{-1}(S)$ is nonempty, so the denominators in the above two expressions are non-zero. Now $g \in v_n^{-1}(S)$ if and only if g fixes t_{n-1} elements of $\mathcal{R}_{n-1}$, t_{n-2} elements of $\mathcal{R}_{n-2}$, and so on down to t_0 elements of $\mathcal{R}_0$. Clearly $v_n^{-1}(S_k)$ consists of those $g \in v_n^{-1}(S)$ that also fix k elements of $\mathcal{R}_n$. Thus (5.10) becomes

$$\frac{1}{\#v_n^{-1}(S)} \sum_k k \cdot \#\left\{g \in v_n^{-1}(S) : g \text{ fixes } k \text{ elements of } \mathcal{R}_n\right\}$$

and this in turn is simply

$$\frac{1}{\#v_n^{-1}(S)} \sum_{g \in v_n^{-1}(S)} (\text{number of fixed points of } g \text{ in } \mathcal{R}_n). \tag{5.11}$$

Expression (5.11) is just the number of elements of $\mathcal{R}_n$ that $g \in v_n^{-1}(S)$ fixes on average.

It is important to note that if $h \in H_n$, then h induces the identity permutation on each $\mathcal{R}_i$ for $0 \leq i \leq n-1$. Thus if $g \in v_n^{-1}(S)$ then so is gh. It follows that $v_n^{-1}(S)$ is a union of cosets of H_n. We may now apply Theorem 5.3. By definition, each $g \in v_n^{-1}(S)$ fixes t_{n-1} elements of $\mathcal{R}_{n-1}$. Theorem 5.3 then proves that (5.11) is equal to t_{n-1}. ■

In view of Corollary 5.4, we call GP(f) a *Galois martingale*. One can repeat our construction of GP(f) for an arbitrary Galois tower over K and obtain a perfectly good process that one could fairly call a Galois process. However, not all Galois processes are martingales; we refer the ones that are as Galois martingales. Our GP(f) is the first known member of this new class of processes. This gives some justification for the first two words of the title of this thesis.

The fact that GP(f) is a martingale gives us much information about its eventual behavior. Theorem 4.9 shows that

$$\mathbf{P}\left(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) \text{ exists}\}\right) = 1. \tag{5.12}$$

In GP(f), X_n is integer valued, so (5.12) implies that with probability 1 the sequence $(X_n(\omega))_{n\geq 0}$ is eventually constant. This strong statement plays a central role in proving our main results.

5.2 The Galois Process for f Converges to 0

In this section, we establish

$$\mathbf{P}\left(X_n = 0 \text{ for all } n \text{ sufficiently large}\right) = 1. \tag{5.13}$$

As noted in the discussion at the beginning of Chapter 4 (page 60) our main result (Theorem 1.7) is equivalent to $\lim_{n\to\infty} \mathbf{P}(X_n > 0) = 0$, so (5.13) quickly leads to a proof of our main result. Since GP(f) converges (5.12), it is enough to show that for any $t \geq 1$,

$$\mathbf{P}\left(X_n = t \text{ for all } n \text{ sufficiently large}\right) = 0.$$

Consider for a moment a fixed $t \geq 1$. Note that

$$\{X_n = t \text{ for all } n \text{ sufficiently large }\} = \bigcup_{m=1}^{\infty} \{X_n = t \text{ for all } n \geq m\}.$$

Thus to prove (5.13) it is enough to show that for each $m \geq 0$,

$$\mathbf{P}\left(X_n = t \text{ for all } n \geq m\right) = 0. \tag{5.14}$$

To accomplish this, we must draw on the results of Chapter 3. Specifically, Theorem 3.2 proves that H_n is maximal for n squarefree. When H_n is maximal we may explicitly compute

$$\mathbf{P}(X_n = t \mid X_{n-1} = t, X_{n-2} = t, \ldots, X_m = t).$$

We do this in the following Lemma and Proposition.

Lemma 5.5. *Let $t \geq 1$ and suppose that $g_0 \in G_n$ fixes t elements of $\mathcal{R}_{n-1}$. Suppose further that H_n is maximal. Then we have*

$$\#\{g \in g_0 H_n : g \text{ fixes } u \text{ elements of } \mathcal{R}_n\} = \begin{cases} \binom{t}{w} 2^{\left(2^{n-1}-t\right)} & \text{if } u = 2w \text{ for some } 0 \leq w \leq t \\ 0 & \text{otherwise} \end{cases}$$

Proof: By Proposition 5.2, the number of elements of $\mathcal{R}_n$ fixed by any $g \in g_0 H_n$ is an even integer $2w$ with $0 \leq w \leq t$. Thus we suppose u is of this form.

The maximality of H_n implies its order is $2^{2^{n-1}}$. By (5.1), this implies that for each pair $\{\pm r\} \subset \mathcal{R}_n$, there exists a unique $h_r \in H_n$ that maps r to $-r$ and fixes all other elements of $\mathcal{R}_n$.

Denote by M the set

$$\{r \in R_n : g_0(r^2) = r^2\}.$$

Note that $r^2 + x$ is an element of $\mathcal{R}_{n-1}$, and $r \in M$ if and only if $g_0(r^2 + x) = r^2 + x$. Since g_0 fixes t elements of R_{n-1}, we have $M = \{\pm r_1\} \cup \cdots \cup \{\pm r_t\}$ for some elements $r_1, \ldots, r_t$ of $\mathcal{R}_n$. In particular, $\#M = 2t$.

Now let J be the subgroup of H_n that fixes each element of M. There are $2^{n-1} - t$ elements of $\mathcal{R}_{n-1}$ that are not fixed by g_0, and this implies that there are $2^{n-1} - t$ pairs $\{\pm r\} \subset \mathcal{R}_n$ not contained in M. The maximality of H_n therefore shows

$$\#J = 2^{(2^{n-1}-t)}. \tag{5.15}$$

Consider $h \in H_n$. Since h fixes r^2 for each $r \in \mathcal{R}_n$, we have $g_0 h(r^2) \neq r^2$ for any $r \in \mathcal{R}_n - M$ (by the definition of M). Therefore $g_0 h$ cannot fix any element of $\mathcal{R}_n - M$. Moreover, any $j \in J$ fixes all elements of M. It follows from these two observations that for any $j \in J$, $g_0 hj$ and $g_0 h$ fix the same number of elements of $\mathcal{R}_n$. Thus every element of a set of the form $g_0 hJ$ has the same number of fixed points in $\mathcal{R}_n$.

Recalling that $M = \{\pm r_1\} \cup \cdots \cup \{\pm r_t\}$, we can write any $h \in H_n$ as

$$(h_{r_1})^{e_1} (h_{r_2})^{e_2} \cdots (h_{r_t})^{e_t} j,$$

where $e_i = 0$ or 1 for each i and $j \in J$. Thus any coset $g_0 hJ$ may be written uniquely as $g_0 (h_{r_1})^{e_1} (h_{r_2})^{e_2} \cdots (h_{r_t})^{e_t} J$. Moreover, all elements of this coset will have exactly

$$2t - (2e_1 + \cdots + 2e_t) \tag{5.16}$$

fixed points in $\mathcal{R}_n$. The number of ways (5.16) can equal $2w$ is precisely $\binom{t}{w}$. The Lemma then follows from (5.15). ■

The next Proposition gives, for n with H_n maximal, an explicit expression of the probability distribution of X_n given past behavior. However, the Proposition does not hold for all possible past behaviors: we must assume that the value of X_{n-1} is known. Under the additional hypothesis that H_n is maximal for all n, we can dispense with this assumption and GP(f) becomes a Markov chain (see Proposition 5.9).

Proposition 5.6. *Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0}) = \mathrm{GP}(f)$, and suppose that H_n is maximal. Suppose also that $m_1 < m_2 < \cdots < m_k$, with $m_k = n-1$. Then for any positive integers $t_1, \ldots, t_k$ with*

$$\mathbf{P}(X_{m_1} = t_1, \ldots, X_{m_k} = t_k) > 0$$

we have

$$\mathbf{P}(X_n = u \mid X_{m_1} = t_1, \ldots, X_{m_k} = t_k) = \begin{cases} \binom{t_k}{w} \frac{1}{2^{t_k}} & \text{if } u = 2w \text{ for some } 0 \leq w \leq t_k \\ 0 & \text{otherwise} \end{cases} \tag{5.17}$$

Proof: The argument is similar to the proof of Corollary 5.4. To prove the Proposition, we must compute

$$\frac{\mathbf{P}(X_{m_1} = t_1, \ldots, X_{m_k} = t_k, X_n = u)}{\mathbf{P}(X_{m_1} = t_1, \ldots, X_{m_k} = t_k)}. \tag{5.18}$$

Let $S = A_0 \times A_1 \times \cdots \times A_n \subset \mathbb{N}^{n+1}$, where

$$A_i = \begin{cases} \{t_{m_j}\} & \text{if } i = m_j \text{ for some } j \\ \mathbb{N} & \text{otherwise} \end{cases}$$

Let S_u be defined similarly to S, except we set $A_n = \{u\}$. Clearly $S_u \subset S$. From the basic property of GP(f) given in (4.14), the expression in (5.18) is equal to

$$\frac{\mu_n(S_u)}{\mu_n(S)}$$

where μ_n is the function defined in (4.11). From the definition of μ_n this is the same as

$$\frac{\#v_n^{-1}(S_u)}{\#v_n^{-1}(S)}. \tag{5.19}$$

Note that $g \in v_n^{-1}(S)$ if and only if g fixes t_k elements of $\mathcal{R}_{m_k} = \mathcal{R}_{n-1}$, t_{k-1} elements of $\mathcal{R}_{m_{k-1}}$, and so on down to t_1 elements of $\mathcal{R}_{m_1}$. Also, $v_n^{-1}(S_u)$ consists of those $g \in v_n^{-1}(S)$ that fix u elements of $\mathcal{R}_n$. By Proposition 5.2, $v_n^{-1}(S_u) = \emptyset$ if u is not of the form $2w$ for $0 \leq w \leq t_k$. Thus (5.19) equals 0 unless u is of this form. We now assume $u = 2w$ for some $0 \leq w \leq t_k$, and we compute (5.19).

Note that if $h \in H_n$, then h induces the identity permutation on each of $\mathcal{R}_{m_k}$, $\mathcal{R}_{m_{k-1}}, \ldots, \mathcal{R}_{m_1}$. Thus if $g \in v_n^{-1}(S)$ then so is gh. This shows that $v_n^{-1}(S)$ is a union of cosets of H_n. By Lemma 5.5, in each coset contained in $v_n^{-1}(S)$ there are $\binom{t_k}{w} 2^{\left(2^{n-1}-t_k\right)}$ elements of $v_n^{-1}(S_u)$. On the other hand, since H_n is maximal there are $2^{2^{n-1}}$ elements in each coset of H_n. It follows that (5.19) is equal to

$$\binom{t_k}{w} \frac{2^{\left(2^{n-1}-t_k\right)}}{2^{2^{n-1}}},$$

which proves the Proposition. ■

A consequence of Proposition 5.6 is that when H_n is maximal, for any $m < n$ and $1 \le w \le 2^{m-1}$ we have

$$\mathbf{P}(X_n = 2w \mid X_m = 2w, \ldots, X_{n-1} = 2w) = \binom{2w}{w} \frac{1}{4^w}, \tag{5.20}$$

provided of course that $\mathbf{P}(X_m = 2w, \ldots, X_{n-1} = 2w) > 0$.

This is the crucial ingredient in establishing equation (5.14). Denote by c_w the right-hand side of equation (5.20). We wish to give an upper bound for c_w when $w \ge 1$. To do this, we note that

$$\frac{c_{w+1}}{c_w} = \frac{1}{4} \frac{(2w+2)(2w+1)}{(w+1)^2} = \frac{4w^2+6w+2}{4w^2+8w+4},$$

and the right-hand side of this equation is less than 1 for $w \ge 1$. Since $c_1 = 1/2$, we have $c_w \le 1/2$ for $1 \le w \le 2^{m-1}$. Thus we have proved the following Lemma:

Lemma 5.7. *Suppose H_n is maximal. Then for any $m < n$ and $0 \le w \le 2^{m-1}$ we have*

$$\mathbf{P}(X_n = 2w \mid X_m = 2w, \ldots, X_{n-1} = 2w) \le \frac{1}{2}.$$

Now, armed with Lemma 5.7, we present a theorem that proves equation (5.14). Note that Theorem 3.2 is instrumental in the proof.

Theorem 5.8. *Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\ge 0}) = \mathrm{GP}(f)$. Then for any $t \ge 1$ and $m \ge 1$,*

$$\mathbf{P}\left(X_n = t \text{ for all } n \ge m\right) = 0.$$

Proof: If t is odd, then it follows from Proposition 5.2 that $\mathbf{P}(X_n = t) = 0$ for all $n \ge 1$. Also, if $t > 2^m$ then $\mathbf{P}(X_m = t) = 0$, and the Theorem is true. We now assume $t = 2w$ for some $0 < w \le 2^{m-1}$.

Let A_m be the event $\{X_n = 2w \text{ for all } n > m\}$. For each $j > m$, let $B_{m,j}$ be the event

$$\{X_n = 2w \text{ for } n = m, m+1, \ldots, j\}.$$

We note that for all $j > m$, $B_{m,j} \supseteq B_{m,j+1}$. Also, we have

$$A_m = \bigcap_{j=m+1}^{\infty} B_{m,j}.$$

By Proposition 4.12 this gives $\mathbf{P}(A_m) = \lim_{j\to\infty} \mathbf{P}(B_{m,j})$.

To give an upper bound for $\mathbf{P}(B_{m,j})$, we let C_n be the event $\{X_n = 2w\}$. We now have

$$B_{m,j} = \bigcap_{n=m}^{j} C_n.$$

If $\mathbf{P}(C_{j-1} \cap \cdots \cap C_m) = 0$ then clearly $\mathbf{P}(B_{m,j}) = 0$. Assuming $\mathbf{P}(C_{j-1} \cap \cdots \cap C_m) \neq 0$, we have

$$\mathbf{P}\left(\bigcap_{n=m}^{j} C_n\right) = \mathbf{P}(C_m)\mathbf{P}(C_{m+1} \mid C_m) \cdots \mathbf{P}(C_j \mid C_m \cap \cdots \cap C_{j-1}) \tag{5.21}$$

by the definition of conditional probability.

We now apply Lemma 5.7, which states that if H_n is maximal then

$$\mathbf{P}(C_n \mid C_m \cap \cdots \cap C_{n-1}) \leq 1/2.$$

Thanks to Theorem 3.2 we know that H_n is maximal when n is squarefree. Denoting by S the set of squarefree positive integers, (5.21) yields

$$\mathbf{P}(B_{m,j}) \leq \left(\frac{1}{2}\right)^{\#(S\cap\{m,\ldots,j\})}$$

The infinitude of S now gives us $\lim_{j\to\infty} \mathbf{P}(B_{m,j}) = 0$, which completes the proof. ■

Proof of Theorem 1.7: Theorem 5.8 implies that for any $t \geq 1$,

$$\mathbf{P}\left(\bigcup_{m=1}^{\infty} \{X_n = t \text{ for all } n \geq m\}\right) = 0. \tag{5.22}$$

GP(f) is a martingale by Corollary 5.4, and thus converges. Its random varibles are integer valued, so each sequence $(X_n(\omega))_{n\in\mathbb{N}}$ is therefore eventually constant with probability 1. Hence (5.22) shows

$$\mathbf{P}\left(\bigcup_{m=1}^{\infty} \{X_n = 0 \text{ for all } n \geq m\}\right) = 1.$$

Clearly for any $m \geq 1$,

$$\{X_n = 0 \text{ for all } n \geq m\} \subseteq \{X_n = 0 \text{ for all } n \geq m+1\}.$$

Therefore by Proposition 4.12 we have

$$\lim_{m\to\infty} \mathbf{P}(\{X_n = 0 \text{ for all } n \geq m\}) = 1.$$

Note that $\{X_m = 0\} \supseteq \{X_n = 0 \text{ for all } n \geq m\}$, giving

$$\lim_{m\to\infty} \mathbf{P}(\{X_m = 0\}) = 1.$$

Taking the complement of the events $\{X_m = 0\}$ now yields $\lim_{m\to\infty} \mathbf{P}(\{X_m > 0\}) = 0$. By equation (4.4) this implies

$$\lim_{n\to\infty} \frac{1}{\#G_n} \#\{g \in G_n \mid g \text{ has a fixed point in } \mathcal{R}_n\} = 0. \tag{5.23}$$

Finally, we specialize to the case $F = \mathbb{F}_p, p \neq 2$. By Theorem 2.18 the equation (5.23) in this case implies that $\lim_{n\to\infty} \delta(\mathcal{I}_n) = 0$. Corollary 2.6 then gives our main result, namely $\delta(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$. The proof of Theorem 1.7 is complete. ■

We note that by Theorem 2.18, equation (5.23) also shows that $\lim_{n\to\infty} D(\mathcal{I}_n) = 0$, provided that K_n/K is geometric for all n. By Corollary 3.40 we know this is the case when $p \equiv 3 \bmod 4$. Therefore by Corollary 2.6, $D(\mathcal{H}(\overline{\mathbb{F}}_p)) = 0$ when $p \equiv 3 \bmod 4$. Assuming Conjecture 2.17 this holds for all $p \neq 2$. These remarks prove the statement following Theorem 1.7.

5.3 Consequences of Conjecture 3.1

We now examine some consequences of Conjecture 3.1, which states that H_n is maximal for all n. With this assumption we show that $\mathrm{GP}(f)$ is a particularly simple branching process (see Section 4.3).

Proposition 5.9. *Suppose that Conjecture 3.1 holds, i.e. H_n is maximal for all n. Then GP(f) is a Markov chain.*

Proof: Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0}) = \mathrm{GP}(f)$. We must show that for any nonnegative integers k, $m_1 < m_2 < \cdots < m_k$, and $s_1, \ldots, s_k$,

$$\mathbf{P}(X_{m_k} = s_k \mid X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1}) = \mathbf{P}(X_{m_k} = s_k \mid X_{m_{k-1}} = s_{k-1}). \tag{5.24}$$

When $m_{k-1} = m_k - 1$, (5.24) follows from Proposition 5.6 (note that the expression on the right-hand side of (5.17) depends only on u and t_k). To show (5.24) follows from this is a standard exercise in elementary probability. We give it here for completeness.

We proceed by induction on the quantity $m_k - m_{k-1}$. The base case was established above. Suppose (5.24) holds for all $m_1 < m_2 < \cdots < m_k$ with $m_k - m_{k-1} = i$, and choose $m_1 < m_2 < \cdots < m_k$ with $m_k - m_{k-1} = i + 1$. We note that the events $\{X_{m_{k-1}+1} = j\}$ partition Ω, and it follows that if $A \subset \Omega$ is an event then

$$\mathbf{P}(A) = \sum_{j\in J} \mathbf{P}(A \cap \{X_{m_{k-1}+1} = j\}). \tag{5.25}$$

We may thus rewrite the left-hand side of (5.24) as follows:

$$\sum_{j\in J} \frac{\mathbf{P}(X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1}, X_{m_{k-1}+1} = j, X_{m_k} = s_k)}{\mathbf{P}(X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1})}.$$

Multiplying numerator and denominator of this expression by

$$\mathbf{P}(X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1}, X_{m_{k-1}+1} = j)$$

yields

$$\sum_{j \in J} \big(\mathbf{P}(X_{m_k} = s_k \mid X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1}, X_{m_{k-1}+1} = j) \cdot \mathbf{P}(X_{m_{k-1}+1} = j \mid X_{m_1} = s_1, \ldots, X_{m_{k-1}} = s_{k-1})\big) \quad (5.26)$$

By our inductive hypothesis, the first part of this product is equal to

$$\mathbf{P}(X_{m_k} = s_k \mid X_{m_{k-1}+1} = j),$$

and this in turn is equal to $\mathbf{P}(X_{m_k} = s_k \mid X_{m_{k-1}} = s_{k-1}, X_{m_{k-1}+1} = j)$. Also by the inductive hypothesis, the second part of the product is equal to $\mathbf{P}(X_{m_{k-1}+1} = j \mid X_{m_{k-1}} = s_{k-1})$. Hence (5.26) is equal to

$$\sum_{j \in J} \big(\mathbf{P}(X_{m_k} = s_k \mid X_{m_{k-1}} = s_{k-1}, X_{m_{k-1}+1} = j) \cdot \mathbf{P}(X_{m_{k-1}+1} = j \mid X_{m_{k-1}} = s_{k-1})\big).$$

Using the definition of conditional probability, this is easily seen to be

$$\sum_{j \in J} \mathbf{P}(X_{m_k} = s_k, X_{m_{k-1}+1} = j \mid X_{m_{k-1}} = s_{k-1}),$$

which is equal to $\mathbf{P}(X_{m_k} = s_k \mid X_{m_{k-1}} = s_{k-1})$ by (5.25). ■

Proposition 5.10. *Suppose that Conjecture 3.1 holds. Then GP(f) is a branching process.*

Proof: Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n \geq 0})) = \mathrm{GP}(f)$. We begin by noting that $G_1 = S_2$, so that $\mathbf{P}(X_1 = 0) = 1/2 = \mathbf{P}(X_1 = 2)$. Thus X_1 may be considered the result of flipping a fair coin, with two points awarded for heads and zero for tails. Letting g be the probability generating function of X_1, we clearly have

$$g(z) = \frac{1}{2} + \frac{1}{2}z^2.$$

Now let Z be the sum of i independent random variables with the same distribution as X_1. It is not hard to see that for $0 \leq j \leq 2i$,

$$\mathbf{P}(Z = j) = \begin{cases} \binom{i}{k}\frac{1}{2^i} & \text{if } j = 2k \text{ for some } 0 \leq k \leq i \\ 0 & \text{otherwise} \end{cases} \quad (5.27)$$

By Proposition 5.9, Conjecture 3.1 implies that GP(f) is a Markov chain. We now draw on Proposition 5.6, which holds for all n by our assumption of Conjecture 3.1. The right-hand side of equation

(5.17) does not depend on n, which shows that GP(f) is time-homogenous. It is clearly non-negative. Finally, Proposition 5.6 gives

$$p_{ij} = \mathbf{P}(X_n = j \mid X_{n-1} = i) = \begin{cases} \binom{i}{k}\frac{1}{2^i} & \text{if } j = 2k \text{ for some } 0 \le k \le i \\ 0 & \text{otherwise} \end{cases}$$

for all $i > 0$ such that $\mathbf{P}(X_{n-1} = i) > 0$. This is the same as the distribution in (5.27), and we know from Proposition 5.1 that 0 is an absorbing state. Thus GP(f) satisfies Definition 4.11. ■

One immediate consequence of Proposition 5.10 is that we can determine the probability e of eventual extinction of GP(f) under the assumption that H_n is maximal for all n. The probability generating function of X_1 is $g(z) = \frac{1}{2}(1 + z^2)$, and by Proposition 4.15, e is the smallest non-zero root of $g(z) - z = \frac{1}{2}(1 - 2z + z^2)$. This clearly gives $e = 1$, which implies Theorem 1.7. Thus establishing that H_n is maximal for all n would allow us to use the above reasoning to arrive at a proof of Theorem 1.7 more expeditiously. However, a proof of the maximality of all H_n has been elusive; see the discussion on page 56.

We close with an examination of the values $\mathbf{P}(X_n > 0)$ under the assumption of Conjecture 3.1. First we give some reminders of the algebraic significance of $\mathbf{P}(X_n > 0)$. Recall that by definition $\mathcal{I}_n$ is the set of $\alpha \in \overline{\mathbb{F}}_p$ such that 0 has an nth preimage in $\mathbb{F}_p(\alpha)$ under iteration of $f_\alpha = x^2 + \alpha$:

$$\mathcal{I}_n = \{\alpha \in \overline{\mathbb{F}}_p : f_\alpha^{-n}(0) \cap \mathbb{F}_p(\alpha) \neq \emptyset\}.$$

By the remark on page 85 following the proof of Theorem 1.7 and the fact that Conjecture 3.1 implies K_n/K geometric for all n (Corollary 3.39), we have that $D(\mathcal{I}_n)$ exists for all n. By Theorem 2.18 and equation (4.4) we then have

$$D(\mathcal{I}_n) = \frac{1}{\#G_n}\#\{g \in G_n \mid g \text{ has a fixed point in } \mathcal{R}_n\} = \mathbf{P}(X_n > 0). \tag{5.28}$$

The definition of natural density says that

$$D(\mathcal{I}_n) = \lim_{k\to\infty} \frac{\#(\mathcal{I}_n \cap \mathbb{F}_{p^k})}{p^k}.$$

By the prime number theorem for polynomials in $\mathbb{F}_p[x]$ [32, Proposition 2.2], for large k almost all elements of $\mathbb{F}_{p^k}$ have degree k. Hence

$$\#(\mathcal{I}_n \cap \mathbb{F}_{p^k}) \approx \#\{\alpha \in \mathbb{F}_{p^k} : f_\alpha^{-n}(0) \cap \mathbb{F}_{p^k} \neq \emptyset\}.$$

We sum up this discussion in the following statement: $\mathbf{P}(X_n > 0)$ is approximately equal to the proportion, for large k, of $\alpha \in \mathbb{F}_{p^k}$ such that 0 has an nth preimage in $\mathbb{F}_{p^k}$ under iteration of $x^2 + \alpha$.

This proportion is easily computed as long as k is not too large. For notational convenience we define:

$$h(p,k,n) = \frac{\#\{\alpha \in \mathbb{F}_{p^k} : f_\alpha^{-n}(0) \cap \mathbb{F}_{p^k} \neq \emptyset\}}{p^k}. \tag{5.29}$$

We thus have shown $\mathbf{P}(X_n > 0) = \lim_{k\to\infty} h(p,k,n)$.

The following corollary of Proposition 5.10 gives a simple formula for $\mathbf{P}(X_n > 0)$.

Corollary 5.11. *If $p \equiv 3 \bmod 4$, then $1 - \mathbf{P}(X_n > 0)$ is equal to the nth iterate of $\frac{1}{2} + \frac{1}{2}z^2$ evaluated at $z = 0$. More generally, this is true if Conjecture 3.1 holds.*

Remark: Corollary 5.11 may be modified to hold in the case where it is only known that H_n is maximal for all $n \leq N$. Naturally in this case the conclusion only holds for $n \leq N$.

Proof of Corollary 5.11: Let $(\Omega, \mathcal{F}, \mathbf{P}, (X_n)_{n\geq 0})) = \mathrm{GP}(f)$. Under the assumption of Conjecture 3.1, Proposition 5.10 shows that $\mathrm{GP}(f)$ is a branching process. Therefore by Proposition 4.14 the probability generating function of X_n is the nth iterate of the pgf $g(z) = \frac{1}{2} + \frac{1}{2}z^2$ of X_1. Thus we have that $\mathbf{P}(X_n = 0)$ is $g^n(0)$. ■

In light of Corollary 5.11 and the discussion preceding it, assuming H_m is maximal for $m \leq n$ we can measure for various values of p the quality of the approximation of $\lim_{k\to\infty} h(p,k,n)$ given by $\mathbf{P}(X_n > 0)$. The assumption that H_m is maximal for $m \leq n$ is not hard to verify in many cases. Indeed, as noted in the discussion on page 56, H_m is maximal for $m \leq 7$, and for specific p one can computationally verify up to $m = 16$ thanks to Theorem 3.38. Also, Theorem 3.2 shows that H_m is maximal for all m as long as $p \equiv 3 \bmod 4$.

The table on the following page presents data on $h(p,k,n)$ for $n \leq 10$ over three finite fields of different characteristic and different degree. One can show that H_n is maximal for $n \leq 10$ in all three of these characteristics, so $\mathbf{P}(X_n > 0)$ does indeed approximate $\lim_{k\to\infty} h(p,k,n)$ for $n \leq 10$. The values in the second column are generated using Corollary 5.11.

The rightmost column of the table is perhaps surprising. Even though the extension has degree 1, the values of $h(p,n,k)$ are for the most part *closer* to $\lim_{k\to\infty} h(p,k,n)$ than are the values of $h(p,k,n)$ for the other extensions, of degree 6 and 9. This suggests that, assuming Conjecture 3.1, the values $\mathbf{P}(X_n > 0)$ also equal $\lim_{p\to\infty} h(p,n,1)$, where the limit is taken through primes. A proof of this statement would require a kind of "horizontal" version of the second form of the Tchebotarev Density theorem. Such a theorem may be achievable by paying careful attention to the dependencies of the error term in the proof of Tchebotarev (see [32, Theorem 9.13B]).

n	$\lim_{k\to\infty} h(p,k,n)$	$h(3,9,n)$	$h(5,6,n)$	$h(22307,1,n)$
1	0.5000	0.5000	0.5000	0.5000
2	0.3750	0.3750	0.3750	0.3750
3	0.3047	0.3060	0.2983	0.3044
4	0.2583	0.2520	0.2515	0.2595
5	0.2249	0.2199	0.2239	0.2265
6	0.1996	0.1947	0.1997	0.2006
7	0.1797	0.1751	0.1777	0.1807
8	0.1636	0.1576	0.1602	0.1649
9	0.1502	0.1470	0.1493	0.1512
10	0.1389	0.1370	0.1391	0.1393

Assuming H_n maximal for all n, we can use Theorem 4.16 to determine the rate at which $\lim_{n\to\infty}\lim_{k\to\infty} h(p,n,k)$ approaches 0 as $n\to\infty$. The probability generating function of X_1 is $g(z) = \frac{1}{2}+\frac{1}{2}z^2$, and obviously $g''(1)=1$. Theorem 4.16 then gives

$$\lim_{n\to\infty}\left(\mathbf{P}(X_n>0)-\frac{2}{n}\right)=0,$$

and by the above remarks $\mathbf{P}(X_n>0)=\lim_{k\to\infty} h(p,n,k)$.

We close with a comment on $\mathcal{H}(\overline{\mathbb{F}}_p)$. Assuming Conjecture 3.1 (and unconditionally for $p\equiv 3 \bmod 4$), we have shown that $D(\mathcal{H}(\overline{\mathbb{F}}_p))=0$, which is equivalent to

$$\#(\mathcal{H}(\overline{\mathbb{F}}_p)\cap\mathbb{F}_{p^k})=o(p^k),$$

where the little-o notation indicates a function whose limit is 0 when divided by p^k. It would be interesting to obtain a lower bound for $\#(\mathcal{H}(\overline{\mathbb{F}}_p)\cap\mathbb{F}_{p^k})$. It follows from our work here that $\limsup_{k\to\infty}\#(\mathcal{H}(\overline{\mathbb{F}}_p)\cap\mathbb{F}_{p^k})=\infty$. This is because the roots of Φ_n are all contained in $\mathcal{H}(\overline{\mathbb{F}}_p)$ (they are the points of primitive Mandelbrot period n in the terminology of Section 3.4), and the Φ_n are of positive degree and pairwise relatively prime by Proposition 3.28. However, it is not immediately clear that $\lim_{k\to\infty}\#(\mathcal{H}(\overline{\mathbb{F}}_p)\cap\mathbb{F}_{p^k})=\infty$, and even if this holds there seems to be no obvious way to give a lower bound for the rate of growth of $\#(\mathcal{H}(\overline{\mathbb{F}}_p)\cap\mathbb{F}_{p^k})$.

Bibliography

[1] G. Álvarez, M. Romera, G. Pastor, and F. Montoya. Determination of Mandelbrot set's hyperbolic component centres. *Chaos Solitons Fractals*, 9(12):1997–2005, 1998.

[2] Mohamed Ayad and Donald L. McQuillan. Irreducibility of the iterates of a quadratic polynomial over a field. *Acta Arith.*, 93(1):87–97, 2000.

[3] Mohamed Ayad and Donald L. McQuillan. Corrections to: "Irreducibility of the iterates of a quadratic polynomial over a field" [Acta Arith. **93** (2000), no. 1, 87–97]. *Acta Arith.*, 99(1):97, 2001.

[4] Paolo Baldi, Laurent Mazliak, and Pierre Priouret. *Martingales and Markov chains*. Chapman & Hall/CRC, Boca Raton, FL, 2002. Solved exercises and elements of theory, Translated from the 1998 French original.

[5] Alan F. Beardon. *Iteration of rational functions*, volume 132 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. Complex analytic dynamical systems.

[6] Robert L. Benedetto. p-adic dynamics and Sullivan's no wandering domains theorem. *Compositio Math.*, 122(3):281–298, 2000.

[7] Zdzisław Brzeźniak and Tomasz Zastawniak. *Basic stochastic processes*. Springer Undergraduate Mathematics Series. Springer-Verlag London Ltd., London, 1999. A course through exercises.

[8] Robert L. Devaney. *An introduction to chaotic dynamical systems*. Addison-Wesley Studies in Nonlinearity. Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, second edition, 1989.

[9] Robert L. Devaney. *A first course in chaotic dynamical systems*. Addison-Wesley Studies in Nonlinearity. Addison-Wesley Publishing Company Advanced Book Program, Reading, MA, 1992. Theory and experiment, With a separately available computer disk.

[10] David S. Dummit and Richard M. Foote. *Abstract algebra*. Prentice Hall Inc., Englewood Cliffs, NJ, 1991.

[11] Yuval Fisher and Jay Hill. Bounding the area of the mandelbrot set. Available at http://www.geocities.com/CapeCanaveral/Lab/3825/Period-Area-16.html.

[12] M. Fried. The nonregular analogue of Tchebotarev's theorem. *Pacific J. Math.*, 112(2):303–311, 1984.

[13] A. Fröhlich and M. J. Taylor. *Algebraic number theory*, volume 27 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1993.

[14] Dante Giarrusso and Yuval Fisher. A parameterization of the period 3 hyperbolic components of the Mandelbrot set. *Proc. Amer. Math. Soc.*, 123(12):3731–3737, 1995.

[15] John Haigh. *Probability models*. Springer Undergraduate Mathematics Series. Springer-Verlag London Ltd., London, 2002.

[16] Theodore E. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963.

[17] M. Herman and J.-C. Yoccoz. Generalizations of some theorems of small divisors to non-Archimedean fields. In *Geometric dynamics (Rio de Janeiro, 1981)*, volume 1007 of *Lecture Notes in Math.*, pages 408–447. Springer, Berlin, 1983.

[18] I. N. Herstein. *Abstract algebra*. Macmillan Publishing Company, New York, second edition, 1990. With a preface by David J. Winter.

[19] Virpi Kauko. Trees of visible components in the Mandelbrot set. *Fund. Math.*, 164(1):41–60, 2000.

[20] Serge Lang. *Algebraic number theory*, volume 110 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1994.

[21] Serge Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2002.

[22] M. Lutzky. Counting hyperbolic components of the Mandelbrot set. *Phys. Lett. A*, 177(4-5):338–340, 1993.

[23] Curt McMullen. Families of rational maps and iterative root-finding algorithms. *Ann. of Math. (2)*, 125(3):467–493, 1987.

[24] Curtis T. McMullen. Frontiers in complex dynamics. *Bulletin of the AMS*, 31(2):155–172, 1994.

[25] Patrick Morton and Pratiksha Patel. The Galois theory of periodic points of polynomial maps. *Proc. London Math. Soc. (3)*, 68(2):225–263, 1994.

[26] R. W. K. Odoni. Realising wreath products of cyclic groups as Galois groups. *Mathematika*, 35(1):101–113, 1988.

[27] A. Peinado, F. Montoya, J. Muñoz, and A. J. Yuste. Maximal periods of $x^2 + c$ in $\mathbb{F}_q$. In *Applied algebra, algebraic algorithms and error-correcting codes (Melbourne, 2001)*, volume 2227 of *Lecture Notes in Comput. Sci.*, pages 219–228. Springer, Berlin, 2001.

[28] Juan Rivera-Letelier. *Dynamique des fonctions rationnelles sur des corps locaux*. PhD thesis, Universite Paris 6, 2001.

[29] Juan Rivera-Letelier. Dynamique des fonctions rationnelles sur des corps locaux. *Astérisque*, (287):xv, 147–230, 2003. Geometric methods in dynamics. II.

[30] Juan Rivera-Letelier. Espace hyperbolique p-adique et dynamique des fonctions rationnelles. *Compositio Math.*, 138(2):199–231, 2003.

[31] Alain M. Robert. *A course in p-adic analysis*, volume 198 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.

[32] Michael Rosen. *Number theory in function fields*, volume 210 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[33] Mitsuhiro Shishikura. The Hausdorff dimension of the boundary of the Mandelbrot set and Julia sets. *Ann. of Math. (2)*, 147(2):225–267, 1998.

[34] Carl Ludwig Siegel. Iteration of analytic functions. *Ann. of Math. (2)*, 43:607–612, 1942.

[35] Michael Stoll. Galois groups over $\mathbf{Q}$ of some iterated polynomials. *Arch. Math. (Basel)*, 59(3):239–244, 1992.

[36] Dennis Sullivan. Quasiconformal homeomorphisms and dynamics. I. Solution of the Fatou-Julia problem on wandering domains. *Ann. of Math. (2)*, 122(3):401–418, 1985.

[37] E. Thiran, D. Verstegen, and J. Weyers. p-adic dynamics. *J. Statist. Phys.*, 54(3-4):893–913, 1989.

[38] H.W. Watson and Francis Galton. On the probability of the extinction of families. *J. Anthropol. Inst. Great Britain and Ireland*, 4:138–144, 1874.

[39] A. D. Wentzell. *A course in the theory of stochastic processes*. McGraw-Hill International Book Co., New York, 1981. Translated from the Russian by S. Chomet, With a foreword by K. L. Chung.

[40] Helmut Wielandt. *Finite permutation groups*. Translated from the German by R. Bercov. Academic Press, New York, 1964.