CHAPTER 1

A Complex Dynamics Primer

The mutability of its fundamental concepts is not an argument for rejecting Science. Each new scientific theory preserves a hard core of the knowledge provided by its predecessor and adds to it. Science progresses by replacing old theories with new. But an age as dominated by Science as ours does need a perspective from which to examine the scientific beliefs which it takes so much for granted, and history provides one important source of such perspective.

Thomas S. Kuhn, *The Copernican Revolution*, [**1976**], p. 3.

What is complex dynamics? At the risk of giving a reductive definition, at its most elemental level it concerns the character of the set

$$\{z, f(z), f^2(z), \ldots, f^n(z), \ldots\},$$

where z is a complex number and f is a given complex function. The quantity f^n, generally referred to as the nth *iterate* of f, denotes the n-fold composition of f.

Our book gives a history of complex dynamics in the first half of the 20th century. The seeds for complex dynamics, however, lie in the last third of the 19th century, when several mathematicians studied the iteration of an arbitrary analytic function f in the vicinity of a *fixed point*, that is, a point λ satisfying $f(\lambda) = \lambda$.

Our goal in the present chapter is to lay the mathematical and historical foundations for the developments of the 1900s that we discuss in later chapters by tracing the arc of the 19th century study of iteration.[1]

1.1. Signs of life

It is ultimately futile to attempt a definitive dating of the genesis of a branch of mathematics. Sometimes it is best simply to call attention to early signs of its life: In 1870 a young *Gymnasium* teacher from Pforzheim, Germany, Ernst Schröder, published a paper in *Mathematische Annalen* with the intriguing title "On infinitely many algorithms for solving equations".[2] It was followed quickly by a second paper, "On iterated functions".[3] Together, the two papers examine iterative equation solving algorithms on the complex plane.

To properly study the algorithms he developed, Schröder realized that he needed to better understand the process of iteration, and so he examined the following question: Given an analytic function f, what is the nature of the infinite

[1] There are several excellent contemporary introductions to complex dynamics. See for example, Blanchard's expository paper [**1984**] or books by Milnor [**2006a**] and Steinmetz [**1993**]. Other excellent references are Beardon [**1991**] and Carleson and Gamelin [**1993**].

[2] Translation of "*Ueber unendlich viele Algorithmen zur Auflösung der Gleichungen*" [**1870**].

[3] "*Ueber iterirte Funktionen*" [**1871**].

sequence $\{z, f(z), f^2(z), \ldots, f^n(z), \ldots\}$, where z is an arbitrary point in the complex plane?

Even though his are the first papers that we know to treat the iteration of complex functions systematically, it is difficult to say why he chose to study the subject of iterative algorithms. It is not even clear whether the idea was even his since he credited the underpinnings of his study to an obscure German mathematician, Heinrich Eggers, remarking,

> *I was inspired in this work in 1867 by reports from Dr. Heinrich Eggers of Meklenburg, who was formerly Professor at the Gymnasium at Schaffhusen and has now emigrated to America.* [**1870**, p. 371][4]

Eggers' role is mysterious. We could not find these reports, and there appear to be no papers by him prior to 1875 when he published two very brief notices[5] [**1876a, 1876b**] that treat iterative algorithms similar to those found in Schröder's papers. Perhaps because of their brevity, these works are nowhere near as deep or as rich as Schröder's study, and it is difficult to argue with Stewart's characterization of them as "inconsequential" [**1993**, p. 1]. Moreover, as we will see, one of the strengths of Schröder's paper are his attempts to understand the iteration of complex functions from a theoretical viewpoint. That perspective is absent from Eggers' work.

Perhaps Schröder felt that no other explanation for his study was necessary. Solving equations, after all, is one of the oldest activities to engage mathematicians. Given the procedures involved for solving cubic and quartic equations, and the impossibility of finding roots of general higher degree polynomials via radicals, it was natural to look towards numerical methods, not just for polynomials, but for arbitrary equations as well. One such method is what is commonly known as Newton's method, and, given the intense interest in the solutions of equations in the first half of the 19th century, it is perhaps no accident that this interest was accompanied by a resurgence of interest in Newton's method.[6]

1.2. Newton's method: The inspiration for complex dynamics?

Newton's method for finding a root of $q(z) = 0$ is commonly given by the recursion formula

$$z_{k+1} = z_k - \frac{q(z_k)}{q'(z_k)}.$$

If all goes well, an appropriately chosen initial point z_0 generates a sequence $\{z_k\}$ converging to a root of q. The *Newton's method function* for a given q is the

[4]Schröder's first paper [**1870**] was translated into English by G. W. Stewart in [**1993**], and while all translations in this work are generally our own, the translation of this quotation is Stewart's, modulo a very few minor changes. Stewart's introduction to his translation contains a very nice analysis of Schröder's first paper in the context of numerical analysis. Alexander [**1994a**, Chap. 1] discusses the role of Schröder's investigation in the development of complex analysis.

[5]These notices appeared in *The Analyst*, a well-regarded amateur mathematics journal published in Des Moines, Iowa, between 1874 and 1883 that was a precursor to the *Annals of Mathematics*. See Deborah Kent's article [**preprint**, pp. 13–14, 19] for more information about the *Analyst*.

[6]See Alexander [**1996a**], Kollerstrom [**1992**], or Ypma [**1995**] for accounts of the history of Newton's method.

functional representation of the algorithm

$$N_q(z) = z - \frac{q(z)}{q'(z)},$$

and application of Newton's method is tantamount to the iteration of N_q, that is, $\{z_k\} = \{N_q^k(z_0)\}$. When analyzing the convergence of Newton's method to the roots of q, it is convenient to use the functional representation N_q. For example, it is trivial, yet nonetheless very useful, to see that the zeros of q become fixed points of N_q.

The standard formulation of Newton's method is not due to Newton, who used a version of this algorithm to invert power series by solving equations of the form $f(y) - x = 0$, where $f(y)$ is the series to be inverted and y is its inverse.[7] Rather, it was evidently first given in a then contemporary paper—from the point of view of Schröder—by Joseph Fourier [**1831**] that appeared posthumously. Indeed, among the infinitely many algorithms referred to in the title of Schröder's first paper are Newton's method and a family of its generalizations.

One issue with Newton's method is the selection of an initial value, or seed, z_0. Determining when "all goes well", that is, when the selection of z_0 leads to a sequence $\{z_n\}$ which converges to a root of $q(z) = 0$, as well as when it does not, is a highly nontrivial question. Generally speaking, while an initial value "near" a root of q gives rise to a convergent sequence, determining precisely what "near" means can be quite difficult.

For example, the function

$$f(z) = z^3 - 4z^2 - 3z + 5 \tag{1.1}$$

has three real roots: one, z_1 near $z = -1.3$; another, z_2 near $z = 0.87$; and the third, z_3 near $z = 4.22$. Surrounding each root is a small disc where Newton's method generates sequences which converge rapidly to that particular root. However, there are regions where strange behavior occurs. For example, in the disc D of radius 0.005 centered at $z = -0.115$, an arbitrary value may converge to any of the three roots under Newton's method, and reasons for this do not emerge from routine calculation. Moreover, there are some initial values in this interval (though they may be hard to find by hand) which lead to periodic sequences of the form $\{z_0, z_1, \ldots, z_{p-1}, z_0, z_1, \ldots, z_{p-1}, \ldots\}$ where a finite string of values endlessly repeats.

1.3. Schröder: The cornerstones of a theory of iteration

Schröder evidently decided that if one is to study algorithms such as Newton's method, one needs to understand iteration; that solutions of equations are not in general real valued may well account for his choice to place iteration on the complex plane.

This desire to unravel the mysteries of iteration led him to articulate several important concepts, perhaps the most fundamental of which is the following from [**1870**, §2], one of the sections inspired by Eggers, which we state in contemporary terms:

[7] Solving for y in the series for the arcsine,

$$y + \frac{1}{6}y^3 + \frac{3}{40}y^5 + \frac{5}{112}y^7 + \frac{35}{1152}y^9 + \frac{63}{2816}y^{11} + \cdots - x = 0,$$

Newton obtained the standard series for sine.

THEOREM 1.1 (Schröder's Fixed Point Theorem). *If f is an analytic function with $f(\lambda) = \lambda$ and $|f'(\lambda)| < 1$, then there is a neighborhood D of the fixed point λ on which all points converge to λ under iteration by f; that is, for any $z \in D$,*

$$\lim_{n\to\infty} f^n(z) = \lambda.$$

Schröder's theorem gives rise to the following definition—although the terminology is not Schröder's, he certainly articulated the concepts it describes:

DEFINITION 1.2 (Fixed point classification I). If a fixed point λ satisfies $|f'(\lambda)| < 1$, it is said to be an *attracting fixed point*. If the modulus $|f'(\lambda)|$ is zero, it is *superattracting*. The quantity $f'(\lambda)$ is called the *multiplier* of the fixed point, and $|f'(\lambda)|$ is the *modulus of the multiplier*.

Schröder's proof of Theorem 1.1 is heuristic and relied on the Taylor series of f expanded about the attracting fixed point λ. In the case where $0 < |f'(\lambda)| < 1$, he chose points z that he described as "infinitely close" to λ, hence the degree one term of the series "outweighs all succeeding terms" [**1870**, p. 321]. He reasoned that he could ignore higher order terms, thus $f(z) - \lambda = f'(\lambda)(z - \lambda)$, effecting an implicit linearization of iteration, an idea we will return to.[8] Implicitly replacing z by $f(z)$, he deduced that $f^n(z) - \lambda = (f'(\lambda))^n(z - \lambda)$, from whence he concluded that $f^n(z) - \lambda$ converges to zero as $n \to \infty$.

The treatment of the superattracting case proceeded similarly with the consideration of $f(z) - \lambda = a_n(z - \lambda)^n$, where a_n is the first nonzero Taylor coefficient, in effect, reducing iteration to the power mapping $(z - \lambda) \mapsto a_n(z - \lambda)^n$.

Although Schröder did not discuss other kinds of fixed points in any detail, they are often classified as follows in the present-day literature:

DEFINITION 1.3 (Fixed point classification II). If the modulus of its multiplier is strictly greater than one, the fixed point is called *repelling*. If $|f'(\lambda)| = 1$, then λ is a *rationally neutral* (or *rationally indifferent*) fixed point if $|f'(\lambda)|$ is a root of unity and *irrationally indifferent* (or neutral) if it is not.

The heuristic Schröder used in his proof of Theorem 1.1 explains why Newton's method works: if λ is a simple root for a given function q, routine calculations reveal not only that $N_q(\lambda) = \lambda$, but also that $N_q'(\lambda) = 0$. Thus, simple roots of $q(z) = 0$ are superattracting fixed points of N_q, and Newton's method converges very quickly, quadratically[9] in fact, on a neighborhood D of λ.

If λ is a root of multiplicity r of $q(z) = 0$, then it remains an attracting fixed point, but is no longer superattracting since the derivative of N_q at λ equals $\frac{r-1}{r}$. In order to speed up convergence, Schröder used a modified version of Newton's method,

$$M_q = z - r\frac{q}{q'},$$

that preserved quadratic convergences [**1870**, §3]. This tinkering with the nature of Newton's method was at the heart of his development of a family of algorithms (one of the families referred to in the title of his paper) that increase the rate of

[8] Schröder did use the equal sign, but wrote ϵ instead of $(z - \lambda)$; however, just a few lines earlier he used ϵ^k to represent $(z - \lambda)^k$ in the Taylor series expansion of f around λ.

[9] A sequence $\{z_k\}$ converges (at least) quadratically to a point λ if there exist positive constants C and K such that if $k > K$, then $|z_{k+1} - \lambda| < C\,|z_k - \lambda|^2$. The convergence is of order r if $|z_k - \lambda|^2$ is replaced by $|z_k - \lambda|^r$.

convergence of the traditional Newton's method N_q. In the next few sections he constructed a family of Newton's method-like algorithms whose convergence is of arbitrary order r.[10]

Schröder cited Eggers' role in the genesis of the material he presented in §3, but not for the sections where he constructed algorithms of an arbitrary order of convergence. Interestingly, Eggers' article [**1876b**] describes not only the generalization of Newton's method for roots of multiplicity greater than one, but it also indicates, via an example, how one might construct an algorithm whose convergence is cubic, noting that,

> *As it would require considerable space to prove* [the generalization to arbitrary order]*, I suppress the demonstration here, but I may be permitted to add one example.* [**1876b**, p. 150]

The example he provided involved an iterative algorithm for finding the roots of the real function $f(x) = (x-2)^3(x-4)$ proceeding from the initial value $x = 2.5$. Eggers' justification for the convergence of his algorithm employed ideas very similar to the heuristic Schröder used in his proof of Theorem 1.1, but which is apparently restricted to the real line, so perhaps the idea to extend the heuristic—and the iterative methods—to the complex plane was Schröder's. That this might be the case is suggested by the fact that nowhere in his articles did Eggers explicitly treat the case of complex iteration.

Schröder's heuristic begs a fundamental question, and not just for Newton's method: How far away can an arbitrary point be from an attracting fixed point λ of a function f and still converge to it under iteration? Schröder did not know the answer to this question, and, indeed, it emerged as one of the deepest unanswered questions in the study of iteration until independently answered by Pierre Fatou and Gaston Julia near the end of World War I in works that are among the central subjects of our book.

In contemporary terms, Schröder's perspective was *local*, meaning it focused on behavior in the vicinity of fixed points, rather than *global*, where the focus is on a classification of the convergence behavior of the *iterative family* $\mathcal{I} = \{f^n\}$ on the entire Riemann sphere. In other words, while Schröder knew that points converge to an attracting fixed point under iteration on a neighborhood of an attracting fixed point, he did not know what happened globally, except in special cases.

1.4. Schröder and functional equations

Functional equations, that is, equations with a function for an unknown, have proved very useful in the study of iteration because they provide a means to reduce the iteration of a function to a simpler mapping. Schröder introduced two functional equations that have become fundamental. Our goal here is not so much to review Schröder's approach and methods (although we will do that to some extent), but rather to indicate the utility of these functional equations in the study of iteration.

1.4.1. The Schröder functional equation. The first, called the Schröder functional equation, seeks to linearize iteration by conjugating f to the mapping $z \mapsto hz$:

[10] See Stewart [**1993**, Intro.] and Alexander [**1994a**, §1.4] for more details about Schröder's algorithms.

DEFINITION 1.4 (Schröder functional equation).

$$S \circ f = h \cdot S,$$

where f is a known analytic function and h is an unspecified complex constant.

If a solution S to the Schröder equation can be found that is defined on D, then the following diagram commutes and f is conjugate to the mapping $z \mapsto hz$ on $N = S(D)$,

$$\begin{array}{ccc} D & \xrightarrow{f} & D \\ {\scriptstyle S}\downarrow & & \downarrow{\scriptstyle S} \\ N & \xrightarrow[z \mapsto hz]{} & N. \end{array} \tag{1.2}$$

Should D contain a neighborhood of a fixed point λ whose multiplier $f'(\lambda)$ is nonzero, it follows from routine calculations that $h = f'(\lambda)$ (and $S(\lambda) = 0$), in which case we arrive at what we will refer to as the *canonical Schröder equation*:

DEFINITION 1.5 (Canonical Schröder functional equation).

$$S \circ f = f'(\lambda) \cdot S, \tag{1.3}$$

where λ is a fixed point of f with $f'(\lambda) \neq 0$ and D contains a neighborhood of λ.[11]

The linearization, provided by an invertible solution S to the canonical Schröder equation on D, greatly simplifies iteration on a neighborhood of λ since it follows from (1.3) not only that iteration is conjugate to the mapping $z \mapsto f'(\lambda) \cdot z$ on a neighborhood of the origin, but that $f^n = S^{-1} \circ (f'(\lambda)^n \cdot S)$ on D. Because $S(\lambda) = 0$, we have $S^{-1}(0) = \lambda$. It therefore follows immediately from the expression for f^n that if $0 < |f'(\lambda)| < 1$, then

$$\lim_{n \to \infty} f^n(z) = \lambda$$

for any z near λ; that is, all z in D converge to the attracting fixed point under iteration by f.

The heuristic Schröder used in his proof of Theorem 1.1 naturally leads to the canonical Schröder equation. Indeed, as we will see below, about 15 years after Schröder's investigation, it was shown that the existence of an attracting, but not superattracting, fixed point is equivalent to a solution of the canonical Schröder equation. Thus, iteration near a fixed point whose multiplier has a modulus strictly between one and zero is always linearizable.

Surprisingly, Schröder did not suggest the canonical equation. His examples of solved Schröder equations were not constructed with the goal of conjugating iteration to multiplication by the derivative of an attracting fixed point. This was due in part to the fact that he was unable to solve arbitrary Schröder equations directly. More will be said about these matters shortly. Perhaps he did not bring up the canonical equation because it made little sense to him to suggest an approach he could not carry out. Moreover, he had already made the point that iteration was tantamount to multiplication by a derivative with his proof of Theorem 1.1.

[11] Solutions are generally normalized so $S'(\lambda) = 1$.

1.4.2. The Abel functional equation. The second functional equation that Schröder introduced is known as the Abel functional equation, and it reduces iteration to addition:[12]

DEFINITION 1.6 (Abel functional equation).

$$A \circ f = A + h,$$

where h is a complex constant and f is known.

The canonical Abel equation reduces iteration to a unit translation:

DEFINITION 1.7 (Canonical Abel functional equation).

$$S \circ f = S + 1. \tag{1.4}$$

Sometimes points converge to a fixed point from one direction only, and the Abel equation can be useful in modeling iteration in this situation. The effect of (1.4) is to translate the dynamics of a function f to that of the elementary mapping $g(z) = z + 1$, whose iteration takes place on the Riemann sphere, and the point at infinity is viewed as a fixed point. In the infinite strip $\Re(z) > C$ union $\Re(z) < -C$ for large C, points are both attracted to ∞ and repelled from it.

There are functions f whose iteration is modeled by g in the sense that there is a region D with a fixed point of f on its boundary such that

$$\begin{array}{ccc} D & \xrightarrow{f} & D \\ {\scriptstyle A}\big\downarrow & & \big\downarrow{\scriptstyle A} \\ \tilde{D} & \xrightarrow[z \mapsto z+1]{} & \tilde{D}, \end{array} \tag{1.5}$$

where $A(\lambda) = \infty$ and $\tilde{D}$ contains the point of infinity on its boundary. Since this diagram commutes, $A \circ f = A + 1$, and there is a solution to the Abel equation for f on D. The behavior we describe occurs if $f(\lambda) = \lambda$ and $f'(\lambda)$ is a root of unity. We will discuss the iteration of such functions further in Sections 1.7.3 and 7.2.5.

In our example of $g(z) = z+1$ the point at infinity does not act like an attracting fixed point since there are points near infinity that are repelled from infinity under iteration by g. In other situations the point at infinity can be an attracting fixed point. In fact, this occurs with any polynomial of degree greater than one and is one reason why iteration of complex functions is generally viewed as taking place on the Riemann sphere $\overline{\mathbb{C}}$.

To make the notion of iteration near infinity rigorous, one can analyze the behavior of f on the set $D = \{z : |z| > R\}$, which is said to be a *neighborhood of*

[12] Niels Abel examined this functional equation—but not in the context of iteration—in a fragment of uncertain date published posthumously which is found in an edition of his collected works [**1881**] that appeared 10 years after Schröder's investigation. Copies of this fragment were scarce prior to this edition, so it is doubtful that Schröder was aware of it. We do not know the origins of the Schröder equation but, since linearization is useful in a wide variety of contexts, we suspect its roots go back a long way.

infinity, via the coordinate change

$$
\begin{array}{ccc}
D & \xrightarrow{f} & D \\
\frac{1}{z}\Big\downarrow & & \Big\downarrow\frac{1}{z} \\
\tilde{D} & \xrightarrow[\tilde{f}]{} & \tilde{D},
\end{array}
\tag{1.6}
$$

which maps D to a neighborhood $\tilde{D}$ of zero. If zero is indeed a fixed point of $\tilde{f}$, infinity is said to be a fixed point of f and is attracting, repelling, or neutral according to the nature of $\tilde{f}'(0)$. The complete set of points which converge to infinity is called the *basin of infinity*.

1.4.3. An early application of functional equations. As he freely admitted, Schröder was unable to develop general solutions to either functional equation and thought that finding a direct solution to a functional equation would be a "difficult task" [**1871**, p. 302]. Because of this, he demonstrated their utility via the construction of examples, one of which involves an application of the Schröder equation to Newton's method, which will be discussed shortly.[13]

Due to the difficulties he found with a direct approach to the solution of functional equations, he instead suggested an "opposite approach" [**1871**, p. 302] that for the Schröder equation is as follows: rather than solving for S directly, fix h, and instead consider the family

$$\mathcal{F}_S = \{S^{-1} \circ h \cdot S\} \tag{1.7}$$

for various strategically chosen S in hopes that relatively simple functions f would emerge.

One such strategy is to let S be a periodic function. The expression $S^{-1} \circ h \cdot S$ often results in a rational identity. Although Schröder succeeded in finding formulae for the iterates of several functions in this manner, it was not an approach susceptible to a wide generalization. Nonetheless, in particular circumstances it can be quite useful, and other mathematicians followed this approach, as we will see in Section 7.5.1.

Schröder had considerable success applying this "opposite approach" to the study of Newton's method for the quadratic. One of the tasks he set for himself was the characterization of Newton's method for polynomials of arbitrary degree. He began with the quadratic case, most likely in hopes that he would discover techniques that would help him understand Newton's method for functions of arbitrary degree.

He assumed that the roots of an arbitrary quadratic are distinct, mapped them to ± 1, and focused his attention on $q(z) = z^2 - 1$. Fixing $h = 2$ and choosing $S(z) = \arctan(iz)$, he manipulated a standard trigonometric identity to obtain the relation

$$f(z) = \frac{2z}{1+z^2} = -i\tan(2 \cdot \arctan(iz)), \tag{1.8}$$

which he in turn interpreted as stemming from the Schröder equation

$$S \circ f = 2 \cdot S.$$

[13] We will give another example from Schröder's paper in Section 7.5.3.

The function f is the reciprocal of the Newton's method function N_q for $q(z) = z^2 - 1$, and in a somewhat baroque fashion Schröder used this information to iterate the Newton's method function for $q(z)$, determining that $N_q^n(z) \to \pm 1$ depending on which side of the imaginary axis z lay.[14]

Arthur Cayley also attempted to classify the dynamics of Newton's method for complex polynomials in [**1879a**], [**1879b**], and [**1880**]. Using very different methods, he nonetheless duplicated Schröder's success in the degree two case but was likewise unable to extend his results to higher degree polynomials.[15] Cayley was evidently unaware of Schröder's work, and despite its being oft cited, Cayley's work is probably tangential to the development of complex dynamics. It is unclear what drew Cayley to the problem of Newton's method. It could have been the imperative to explore means to solve equations. Reflective of his interest in Newton's method was his inclusion of a question pertaining to its convergence in the quadratic case on the mathematics exam for 1879 Smith Prize at the University of Cambridge.

1.5. Fundamental concepts of iteration

Although Schröder did not articulate things in this manner, one way to describe the convergence of arbitrary points to the fixed point is via the concept of an *orbit* of a point under iteration:

Definition 1.8 (Forward orbit of a point). Let f be a complex function and $z \in \overline{\mathbb{C}}$. The set $O^+(z)$, called the *forward orbit of* z under f, is defined by $O^+(z) = \{f^n(z)\}$ where n is a nonnegative integer and $f^0(z) = z$.

To extend the definition of an orbit, we need the concept of the *total inverse* of f:

Definition 1.9 (Total inverse of f.). By the *total inverse* of f we mean the following multifunction: $f^{-1}(z) = \{w : f(z) = w\}$.

The backward orbit of a point z under f is thus the orbit of z_0 under the multifunction f^{-1} :

Definition 1.10 (Backward orbit of a point). The *backward* orbit of z is the set $O^-(z) = \{f^{-n}(z)\}$ where, again, n is a nonnegative integer. Thus, the set $O^-(z)$ coincides with $\{w : f^n(w) = z\}$ for some nonnegative integer n. An equivalent way to define the backward orbit of a point z under f is thus to say it consists of the set of *preimages* of z, where a *preimage* of z is a point w satisfying $f^n(w) = z$, for some positive index n. One can of course also consider the orbit of a point under a single branch of the inverse. The *total orbit of* z is the union $O(z) = O^+(z) \cup O^-(z)$.

Another important concept related to the inverse of a function is that of a *critical point* and a *critical value* of f:

Definition 1.11 (Critical value). A *critical point* of a function f is a point z such that $f'(z) = 0$. The point $w = f(z)$ is called a *critical value* of f and is a branch point of the inverse. For this reason, aa critical value of f is sometimes referred to as a *critical point of the inverse of f*.

[14] See Alexander [**1994a**, §1.7].
[15] See Alexander [**1994a**, §1.8].

If f is a rational function of degree d,[16] then

$$w \text{ is not a critical value} \Longleftrightarrow f^{-1}(w) = \{z_1, \ldots, z_d\},$$

namely $f^{-1}(w)$ consists of d distinct points. When w is not a critical value, there are neighborhoods D_i of z_i and D of w such that $f|_{D_i} : D_i \to D$, is one-to-one and onto. Moreover, there is a well-defined branch of the inverse of f at w, f_i^{-1}, mapping D to D_i.

Related to this is the concept of the *post-critical set*.

Definition 1.12 (Post-critical set). Let C denote the set of critical points of a function f. The *post-critical set* is the union of the orbits $O^+(c)$ where $c \in C$.

We can use the terminology of orbits to describe the dynamics of Newton's method for the quadratic: if z is in the right half-plane, then $O^+(z)$ converges to 1; to the left of the imaginary axis, $O^+(z)$ converges to -1. Points on the imaginary axis, which form the boundary of the convergence regions for Newton's method, do not converge to either root under the iteration of N_q and in fact remain on the axis. For this reason, the imaginary axis is said to be *invariant* under iteration by N_q, a fact Schröder understood.

Indeed, Schröder briefly described the behavior of the iteration of N_q on the imaginary axis, which will be discussed in more detail in Section 7.5.3. We will, however, mention two of his observations here. The first requires a definition of a *periodic point*:

Definition 1.13 (Periodic point of order p). If, for $p \in \mathbb{N}$, a point z satisfies $f^p(z_0) = z_0$, with $f^k(z_0) \neq z_0$ for $0 < k < p-1$, then z_0 is said to be a periodic point of f of order p.[17] The orbit $O^+(z_0) = \{z_0, z_1 = f(z), \ldots, z_{p-1} = f^{p-1}(z)\}$ is said to be a *periodic orbit.* If $p = 1$, then z_0 is a fixed point. An orbit is attracting, repelling, or neutral depending on whether

$$\left| \frac{d}{dz} f^p(z)\Big|_{z=z_k} \right|$$

is less than, greater than, or equal to one. The quantity inside the absolute value is called the multiplier of the orbit.[18]

Schröder observed, first, that there were periodic points of N_q of every order on the imaginary axis [**1871**, p. 320]. Although he made no mention of the fact, these periodic orbits are either repelling or neutral. His second observation was that if z is on the imaginary axis but not eventually periodic, then $N_q^k(z)$ takes on infinitely many values [**1871**, p. 320]; in other words, the orbit $O^+(z)$ consists of infinitely many distinct points. The theory developed in the early 20th century tells us that the orbit of a nonperiodic point in the boundary J between two convergence regions is dense in J, so its closure equals J. Schröder's observation begs a question, how

[16]A complex rational function is a function of the form p/q where p and q are polynomials. Its degree is $d = \max\{\deg p, \deg q\}$.

[17]Our definition of a periodic orbit suggests that the concept of "orbit" can be viewed as an infinite sequence of points ordered by the index of iteration, or as the set of distinct values in the sequence, as in the case of a periodic orbit. In general, if we do not use the term "periodic", the reader should assume that the orbit under question consists of infinitely many distinct values ordered by the index of iteration.

[18]An application of the chain rule shows that the multiplier is the same for all points in the orbit.

did he view the character of the infinite set formed by the $N_q^k(z)$? Aside from this second observation, Schröder was silent on this matter. It should be pointed out that at the time of Schröder's investigation, the mathematical language to describe denseness or closure was not in use.

1.6. Newton's method on the Riemann sphere: Two examples

1.6.1. Example I: Newton's method for the cubic. Schröder and Cayley each hoped to understand Newton's method for higher degree polynomials. However, they failed in their attempts to understand Newton's method for the cubic, which following Schröder's lead for the quadratic case, we will express as $c(z) = z^3 - 1$.

One roadblock they encountered was that the techniques they each used for the quadratic case were ad hoc and simply do not generalize. Perhaps the greatest obstacle was that due to the sheer complexity of the behavior of Newton's method for the cubic c, the mathematical tools available to them were not up to the task. Unlike Newton's method applied to the quadratic, which induces a partitioning of the sphere into two convergence regions bounded by a simple closed curve, iteration by the Newton's method function for the cubic, N_c, partitions $\overline{\mathbb{C}}$ into infinitely many convergence regions. It would be over 45 years after Schröder's work before the behavior he confronted with N_c was well understood, as will be discussed in Chapter 7.

The reader wanting a glimpse into the difficulties Schröder and Cayley faced should refer to Color Plate Figure 5.[19] The regions of convergence are color-coded in order to depict the dynamics of N_c from a global perspective. The blue region is the set of points which converge under N_c to $z = 1$, points in the red region converge to $z = e^{i\frac{2\pi}{3}}$, and those in the green to $z = e^{i\frac{4\pi}{3}}$. As one can see, not only does the set of points which converge to a particular root of $c(z) = 0$ consist of infinitely many disjoint connected components, but the boundary between them borders intricate braids of convergence regions that extend from the point at infinity to the origin.

These images also provide insight into the behavior we described in equation (1.1) regarding Newton's method for the function $f(z) = z^3 - 4z^2 - 3z + 5$, where very small variations in the initial conditions of points in the disc of radius 0.005 centered at $\alpha = -0.115$ cause points to converge to different roots: three distinct convergence regions for N_c intersect this disc, analogous to what occurs on a disc in Color Plate 5 intersecting one the braids that form along the edge of the larger convergence regions.

The partitioning of the sphere that occurs with N_c leads to additional definitions: (*or basin*) *of convergence*

Definition 1.14 (Domain of convergence). Given a function f, the set of all z for which $f^n(z)$ converges to a fixed point λ as $n \to \infty$ is called the *total domain* (*or basin*) *of convergence* of λ and is often denoted $A(\lambda)$. It is possible that this set has more than one connected component (if so, as we will see later, it must have infinitely many), in which case the component containing λ is called the *immediate domain* (*or basin*) *of convergence*. If λ is the point at infinity, then the total domain

[19]See page 143 for more information about the images displayed in Color Plate 5.

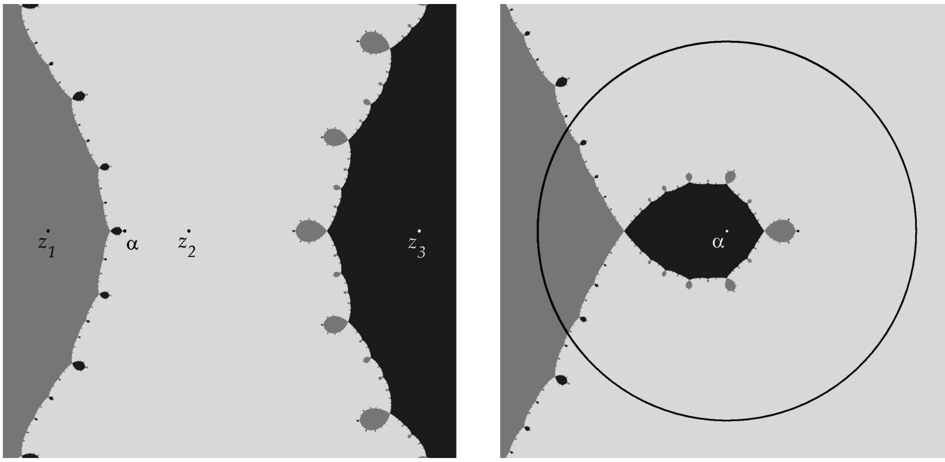


FIGURE 1.1. The dynamics of the Newton's method function N_f for $f(z) = z^3 - 4z^2 - 3z + 5$. The domains of convergence for the three roots of $f(z) = 0$ are pictured, as is the disc D centered at α. Similar to what occurs with Newton's method for the cubic $c(z) = z^3 - 1$, there are innumerable distinct convergence regions within D.

of convergence for λ is called *the basin of infinity*.[20] The domain of convergence of an attracting periodic orbit $\{z_0, \ldots, z_{p-1}\}$ of f is the union of the domains of convergence of f^p to the points z_k.

In Color Plate 5 the union of the blue regions is the total domain of convergence for $z = 1$; the large blue region in the left image represents the immediate domain of convergence of $z = 1$.

The union of the domains of convergence of N_c is an example of what is known as the Fatou set. The complement of the Fatou set—typically the boundary of the union of the domains of convergence—is known as the Julia set. A more precise definition of the Fatou and Julia sets is given the next section, and another in Chapter 7. We will often refer to the Julia and Fatou sets in our book, but the reader should understand that such use is anachronistic since this terminology did not come into common usage until the early 1980s.

1.6.2. The Fatou set, the Julia set, and rotation domains. Repelling and rationally neutral periodic points are in the Julia set. Indeed, for rational and transcendental functions,[21] the Julia and Fatou sets can be defined as follows:[22]

[20]For polynomials, the boundary of the basin of infinity is the what is known as the Julia set, to be discussed in the next section.

[21]Fatou and Julia's initial work focused on the iteration of rational functions of degree two or higher. We will explore the reasons for their choice in subsequent chapters. Fatou later extended his study to the iteration of other kinds of functions, as we will discuss in Chapter 10.

[22]There are historical and mathematical issues associated with this definition. It is the one Julia chose in his monograph on rational functions [**1918a**]. The standard definition, which we will postpone until Chapter 7 when we introduce Montel's theory of normal families, is to define the Fatou set as the domain of normality of the iterative family $\{f^n\}$ and the Julia set as its complement. This is the definition that Fatou used in his monograph on rational functions [**1919-1920**], and it extends naturally to transcendental functions. While Julia's definition also

Definition 1.15 (Julia and Fatou sets). Given a rational or transcendental function f, the Julia set of f, denoted $J(f)$ or J, is the closure of the set of repelling and neutral periodic points of f. The Fatou set, denoted $F(f)$ or F, is the complement of the $J(f)$ in $\overline{\mathbb{C}}$.

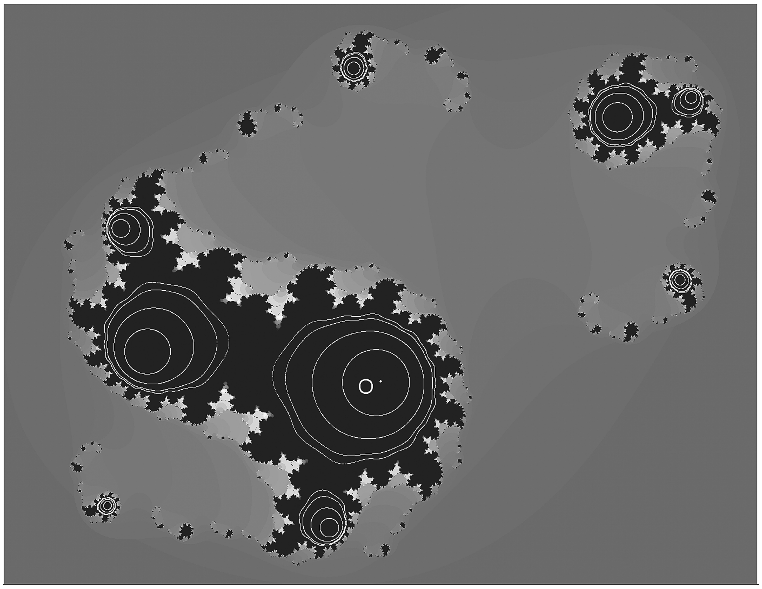


Figure 1.2. The origin O is a center for $f(z) = e^{i2\pi\sqrt{3}}z + z^2 + iz^3$. Four orbits around O are indicated, and the Siegel disc D is the component of the Fatou set containing the center and the four orbits. The Siegel disc has three preimages under f, itself, and the regions with two largest sets of orbits, each of which in turn have three preimages. The large grey region surrounding the preimages is the basin of infinity.

The Fatou set of a rational complex function contains all attracting periodic orbits, as well as their domains of convergence. It can also contain sets called *rotation domains* on which iteration is conformally equivalent via a Schröder equation, to an irrational rotation of a disc or annulus through $2\pi\theta$, where $\theta \in [0, 1)$.[23]

In the event the rotation domain D is conformally equivalent to a disc, it contains in its interior an irrationally neutral fixed point λ, called a *center*. The domain D is also referred to as a *Siegel disc*, and its angle of rotation θ is given by the multiplier $f'(\lambda) = e^{i2\pi\theta}$ of λ. Figure 1.2 illustrates a Siegel disc.

When D is a rotation domain conformally equivalent to an annulus, D is called a *Herman ring* or an *Arnold-Herman ring*. The rotation does not correspond to a

extends, it takes considerably more effort to show this (and indeed took many years), and for this reason it is nowadays generally stated as a consequence of the standard definition of the Fatou set as the domain of normality. We discuss issues involved with the different definitions at the beginning of Section 7.2.2 and on page 223; also see Bergweiler [**1993**, §3.4].

[23] The possible configurations for components of the Fatou set are discussed beginning near the bottom of p. 332 and continuing to the next page.

multiplier of a fixed point, however, and is not as straightforward to compute as it is for a center.[24] We depict a Herman ring in Color Plate 1.[25]

Caption for Color Plate 1: (See p. 63.) This image depicts a Herman ring for the function

$$f(z) = \frac{e^{i2\pi\theta} z^2(z-4)}{1-4z},$$

where θ is chosen so that iteration by f is conjugate to the rotation $z \mapsto ze^{i2\pi\phi}$ where $\phi = (\sqrt{5}-1)/2$.[26] The grayish region containing the origin O, denoted D_0 is the immediate domain of convergence of O, which is a superattracting fixed point. The Herman ring A surrounds D_0, and the paths in A along which the first 1000 points of $O^+(z)$ lie for $z = 0.4, 0.7, 1, 1.3, 1.6$, are indicated. Note that the orbit for $z = 1$ lies on the unit circle, which f fixes.

Also pictured are the myriad preimages of A and D_0. As we will see later, when there is rotary behavior, the degree of connection of a component of the Fatou set is either one or two, hence the Julia set—which is not depicted—snakes through the image, dividing the Fatou set into infinitely many components. Each preimage of A or D_0 is its own distinct component.

While the existence of centers for rational functions of degree greater than one was speculated for quite some time, they were not shown to exist until 1942. It took another 40 years to establish the existence of Herman rings, although there was an anticipation of them in Germany in the early 1930s, which we will discuss in Chapter 11.

1.6.3. Example II: Newton's method for the quadratic. The dynamics of the Newton's method function N_q for $q(z) = z^2 - 1$ are infinitely less complex than those of the cubic, but it is nonetheless worthwhile to explain the local behavior of N_q from a contemporary perspective.

The points $z = \pm 1$ are the fixed points of

$$N_q(z) = \frac{z^2+1}{2z};$$

since $N_q'(\pm 1) = 0$, they are superattracting fixed points. Theorem 1.1 guarantees that both $z = \pm 1$ are surrounded by neighborhoods $D(\pm 1)$ on which all points converge to the corresponding fixed point under iteration by N_q. The question then arises, how far does the convergence of N_q extend beyond $D(\pm 1)$? The local theory developed by Schröder does not answer this question, but, by dint of calculation and good fortune, he and Cayley were independently able to show that the domains of convergence for N_q extend to the left half-plane for $z = -1$ and the right half-plane for $z = 1$.

---

[24]One, in essence, averages the angular distance between consecutive points in an orbit. See Milnor [**2006a**, pp. 164–166] or Steinmetz [**1993**, Chap. 4] for details.

[25]While color images have been separated from the text, we place their captions within our exposition.

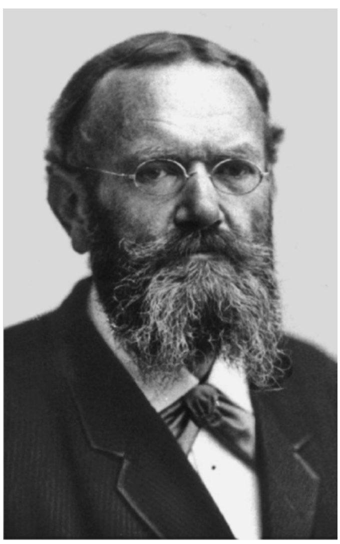
[26]For a rational function, the angle of rotation of a center is simply the multiplier of the fixed point. Finding the angle of rotation for a Herman ring is considerably more complicated. In our example, it cannot be gleaned trivially from the $e^{i2\pi\theta}$ term in f, although the angle of rotation is a continuous function of θ. See Steinmetz [**1993**, pp. 102ff.].

Each half-plane thus comprises both the immediate and complete domain of convergence for the respective fixed points. Such domains are said to be *completely invariant* under iteration since the complete orbit $O(z)$ of a point z remains in the half-plane of its origin. The union of the right and left half-planes forms the Fatou set, and the boundary between the convergence regions, the imaginary axis, is the Julia set.

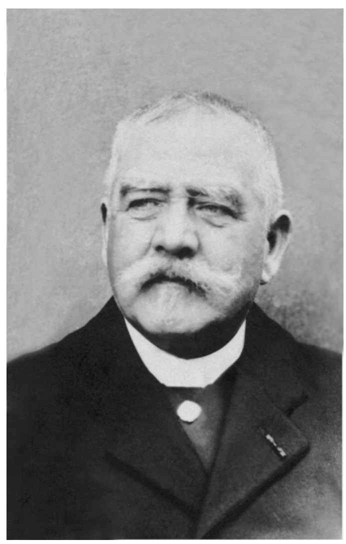
1.7. Towards a deeper understanding of the local theory: The work of Kœnigs and his students

Despite articulating several fundamental ideas concerning the iteration of complex functions, Schröder's understanding of the local theory had several significant gaps. He did not investigate the nature of periodic points or attempt to classify fixed points; nor did he develop a unified approach to the Schröder and Abel equations. These tasks would be left to the French mathematician Gabriel Kœnigs and his students, Auguste Grévy and Léopold Leau.

1.7.1. A rigorous approach to the Schröder functional equation. Kœnigs' three papers from the mid-1880s [**1883, 1884b, 1885b**] represent the most significant exploration of complex dynamics of the 19th century. Although Kœnigs cited Schröder in [**1884b**] (but not in his first paper), it is not clear how much of an influence Schröder was on Kœnigs; the lack of any reference to Schröder in Kœnigs' first paper suggests it may have been minimal.



ERNST SCHRÖDER



GABRIEL KŒNIGS

FIGURE 1.3. Photo of Kœnigs courtesy of Académie des sciences—Institut de France.

Kœnigs developed a rigorous, general approach to the behavior of iteration near attracting fixed points. He produced initial proofs of Theorem 1.1 and the existence of solutions to the Schröder and Abel functional equations when $f(\lambda) = \lambda$, where

$0 < |f'(\lambda)| < 1$ in [**1883**], but later realized they were lacking somewhat in rigor. Prompted by Gaston Darboux,[27] he provided fully rigorous proofs in [**1884b**].

Of most interest to us is his solution to the canonical Schröder equation

$$S \circ f = f'(\lambda) \cdot S \tag{1.9}$$

on a neighborhood D of λ, where λ is an attracting (but not superattracting) fixed point of f. He proved, first in [**1883**] and then more rigorously in [**1884b**], that

$$S(z) = \lim_{n\to\infty} \frac{f^n(z) - \lambda}{z - \lambda} \tag{1.10}$$

is an invertible, analytic solution to equation (1.9) on D. In the proof from [**1883**] he showed only that the limit converges pointwise on D; in his second proof, he used his extensions of Darboux's theorems to prove that the convergence is uniform, which allowed him to conclude that S is analytic. His proof also used the method of majorants, which we describe in a different context in Section 3.2.

Kœnigs derived a solution to the Abel equation by setting

$$A = \frac{\ln(S)}{f'(\lambda)}. \tag{1.11}$$

Routine calculations show that near λ, $A \circ f = A + 1$, but there is a logarithmic singularity of A at λ. This turned out not to be a serious problem. The linearization provided by the Schröder equation is an optimal model for iteration of f near an attracting fixed point λ satisfying $0 < |f'(\lambda)| < 1$, so the translation model suggested by the Abel equation is not particularly germane.

However, as suggested in our discussion in Section 1.4.2, there are times when iteration near a fixed point acts like a translation. While Kœnigs did not use the Abel equation to model this sort of behavior, its utility to do so was recognized by his student Leau in his study of rationally neutral fixed points in the late 1890s, as well as by Fatou over 30 years after Kœnigs' first paper.

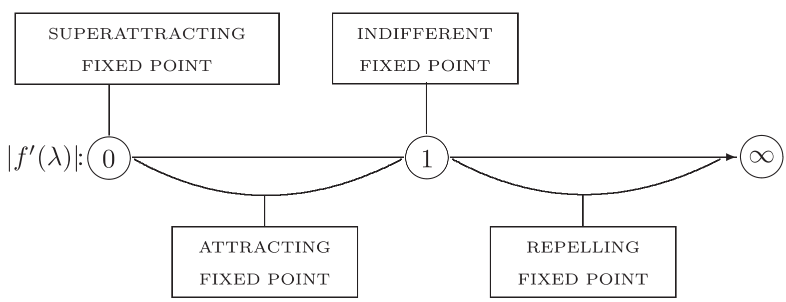


FIGURE 1.4. The local viewpoint, circa 1900: the classification of fixed points after Kœnigs. The horizontal line represents the modulus $|f'(\lambda)|$. Much of the post-Kœnigs 19th century French study of iteration focused on determining the properties of iteration near indifferent and superattracting points.

[27]Kœnigs was a student of Darboux, and at the onset of [**1884b**] he extended to complex functions two theorems of Darboux regarding the uniform convergence of series of real functions. For more information about Darboux and Kœnigs, in particular, the latter's extension of the former's theorems, see Alexander's paper [**1994b**].

1.7.2. Superattracting fixed points: Grévy and Böttcher. Kœnigs did not treat the superattracting case. That was left to his student Auguste Grévy who wrote two papers [**1894, 1896**] that generalized Kœnigs' work in two principal ways. One was to consider functional equations of the form

$$p_0 \cdot G + p_1 \cdot G \circ f + \cdots + p_n \cdot G \circ f^n \equiv 0, \tag{1.12}$$

on a neighborhood D of an attracting fixed point λ of f, where the p_k are given functions analytic on D and G is unknown. Setting $n = 1$, $p_0(z) \equiv f'(\lambda)$, and $p_1 \equiv -1$ yields the canonical Schröder equation. He initially treated the case where $0 < |f'(\lambda)| < 1$ but later extended his treatment to the superattracting case.

His second generalization of Kœnigs' study sought solutions to functional equations analogous to the Schröder equation in the neighborhood of a superattracting fixed point. One such functional equation is the Schröder equation $S \circ f = k \cdot S$, where, for $f(z) = a_k(z - \lambda)^k + \cdots$, with $a_k \neq 0$, he found a solution S with a logarithmic singularity at λ. Grévy's interest in functional equations was by no means restricted to iteration, and he did not seem to intend this equation as a means to model iteration, although if the fixed point were infinity, the mapping $z \mapsto kz$ would to some degree suggest the attracting behavior of a fixed point at infinity.[28]

However, this is not an optimal model. Schröder's heuristic (discussed on page 6), as well as elementary examples, such as $f(z) = z^k$, intimate that iteration near a superattracting fixed point acts like $z \mapsto z^k$. Indeed, the young Polish mathematician Łucyan Böttcher found a solution B to the so-called Böttcher functional equation,

$$B \circ f = B^k,$$

in the neighborhood of a superattracting fixed point λ of a function f satisfying $f(z) = a_k(z - \lambda)^k + \cdots$ in [**1904a, 1904b**], which were written in Russian. One can obtain B from Grévy's function S by setting $B = e^S$, as is easily seen. More will be said about Böttcher's body of work in Section 7.5.3.

1.7.3. Neutral fixed points: Kœnigs and Leau. Kœnigs' treatment of the neutral case was limited to elliptical linear fractional transformations, that is, linear functions L which satisfy

$$\frac{L(z) - a}{L(z) - b} = e^{i\theta} \frac{z - a}{z - b}.$$

He noted that if θ is commensurate with π, the iterative behavior was periodic, and if not, iterates of L would not converge at all. He then remarked, "I have excluded this case from my researches" [**1883**, p. 352]. Since this was the only instance where Kœnigs explicitly examined a neutral fixed point, it is reasonable to interpret this statement as an admission that he was unable to develop a general treatment of neutral fixed points—rational or otherwise.

Such an examination, at least in the rational case, was successfully carried out in 1897 by his student Leau. One of the difficulties in this case is that the Schröder functional equation does not have an analytic solution at a neutral fixed point.

[28] It turns out that a function very much like Grévy's G that also satisfies $S \circ f = k \cdot S$ but in a different context than he envisioned, called the Green's function, is quite useful in the study of iteration of polynomials on the basin of ∞. See for example, Branner [**1989**, p. 88] or Milnor [**2006a**, p. 100].

This can be seen as follows. Let f have a rationally neutral fixed point, which, for convenience we will assume is zero, and let

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

be its Taylor expansion. Assume that an analytic solution

$$S(z) = c_1 z + c_2 z^2 + \cdots$$

to the canonical Schröder equation $S \circ f = a_1 \cdot S$ exists on a neighborhood of the origin.[29]

Solving $S \circ f = a_1 \cdot S$ directly for the c_n yields the recurrence relation

$$c_{n+1} = \frac{c_1[a_1^{n-1}(1-a_1)\cdots(1-a_1^{n-1})a_{n+1} + P_{n+1}(a_1,\ldots,a_n)]}{a_1^n(1-a_1)\cdots(1-a_1^n)},$$

where the P_n are polynomials of the a_k. Since a_1 is a root of unity, infinitely many coefficients of S will have zero in the denominator. Leau used a different method to show that analytic solutions to the Schröder equation do not exist. We choose this method since the above recurrence relation will appear again when we discuss irrationally neutral fixed points in subsequent chapters.

As noted above, a solution to the canonical Abel equation (1.4) exists, albeit one with a logarithmic singularity. However, there is another way to describe iteration near rationally neutral fixed points, the so-called Flower Theorem, articulated by Leau in [**1897**],[30] which yields a helpful visual depiction of iteration in the rationally neutral case, examples of which are given in Figure 7.3 in Section 7.2, where we will discuss the Flower Theorem in more detail.[31]

Iteration near irrationally indifferent fixed points presents much greater difficulties than it does in other cases. Like Kœnigs, Leau made no inroads into the irrational case:

> *...we will discard the case where* $\left|\frac{d\varphi}{dz}\right|_0 = 1$ *and the argument is irrational* [**1897**, p. 65].

1.8. The problem of the division of the plane

Kœnigs concluded his work [**1884b**] with remarks that suggest he was well aware that there was a significant gap in his understanding of the global behavior of iteration:

> *If one envisions the totality of points from the plane which are carried into the interior of* C_x[32] *and, as a result, to the point* x*, one can extend to this region our general theorems. But nothing is known of the general manner in which the region is bounded, and*

[29]As we have previously noted, routine calculations show that any solution S to a canonical Schröder equation maps the fixed point to the origin; thus, in this case, it fixes the origin.

[30]Another French mathematician, Ernest Lémeray, gave a less polished version of the Flower Theorem in [**1897a**], the same year as Leau. Lémeray's paper also contains an early example of the technique of graphical iteration which is used to depict the iteration of a real function near an attracting fixed point. See Figure 8.2 in Section 8.4 for a diagram illustrating this technique. Lémeray's work is discussed in Alexander [**1994a**, §5.2].

[31]The reader curious to know more of the mathematics of the Flower Theorem is invited to consult one of the standard complex dynamics references listed above in footnote 1. See Alexander [**1994a**, Chap. 5, 6] and Audin [**2009**, p. 93] for discussions of the history of the Flower Theorem.

[32]That is, a neighborhood of an attracting fixed point x.

> *one cannot affirm a priori that its mode of delimitation is not of a nature which restricts such an extension.*
>
> *The importance of the division of the plane into regions according to the limit points to which they are carried is thus once again underscored. But in realizing that in general there are infinitely many circular limit groups,*[33] *since the index* [of iteration] *can be arbitrarily large, one grasps the difficulty attached to this problem.* [**1884b**, p. 40]

The common interpretation of Kœnigs' remarks is that he was speculating that there could be infinitely many periodic attracting orbits. That may be so, but he could also have been suggesting that periodic points are not in general attracting, as Schröder showed in the case with Newton's method for the quadratic. If so, Kœnigs turned out to be correct. In any case, the fundamental theorem of algebra guarantees that for rational f, $f^n(z) = z$ has solutions for all n, so there are infinitely many periodic groups, and the task of classifying them would have been daunting, especially given the lack of set theoretic or topological tools available to him.

A little over a decade later, Leau echoed Kœnigs' pessimism at the conclusion of [**1897**]. His study of the domain of solutions of functional equations led him to also confront the intractability of the problem of "the division of the plane":

> *The solutions of functional equations are defined in a small domain* [referred to below as the primitive domain]. *One can prolong the functions thus obtained to the regions from whence the points came if, after a finite number of iterations, they fall in the primitive domain. Naturally, we suppose that the solution of the equation is still defined in these new regions. ... The problem of the extension of these solutions reduces to the division of the plane into regions which are transformed into one another* [under iteration], *a problem which is without a doubt in general impractical, if one supposes that the function to be iterated is given in a series development.* [**1897**, p. 55]

What is particularly interesting is that he evidently did not find an analytic continuation, one of the most powerful tools for extending domains of solution, up to the task of finding maximal domains of solutions to functional equations. We will see Fatou's response to the issue of continuation in our discussion of Chapter VII of his monograph on rational functions, [**1919-1920**] in Section 7.3.

[33] Period p points.

CHAPTER 2

Introduction: Dynamics of a Complex History

2.1. Introduction

The development of complex dynamics is characterized by an ebb and flow that induces a partition of the study into four distinct phases. At the end of each of the first three phases lies a fallow period, in which interest in the subject seems to wane. But then a spark arises, and new ideas burst forth.

The first phase, discussed in the opening chapter, began in 1870 with Ernst Schröder's study and ended with a paper [**1897**] by Léopold Leau. Pierre Fatou's *Comptes Rendus* notice [**1906b**] initiated the next phase, which included works prepared in anticipation of the 1918 *Grand Prix des Sciences Mathématiques* and concluded with Carl Ludwig Siegel's paper [**1942**]. While characterized by the development of a robust global theory of rational dynamics,[1] it also witnessed the first attempts to understand the iteration of algebraic and transcendental functions.

The aftermath of World War II through the early 1970s constitutes the third phase. It is commonly thought that not much happened between Siegel's paper and the onset of the computer, but an examination of the record, about which we say more shortly, belies this. The advent of the personal computer helped trigger the fourth phase, in which the pioneering work of Adrien Douady, John H. Hubbard, and Benoit Mandelbrot in the late 1970s and early 1980s launched an energetic and far-reaching body of work that, as of the completion of our manuscript, shows no signs of abating.

Our book concerns the second phase and details its progress in Europe and the United States from 1906 to 1942. Alexander's book [**1994a**] treats the initial phase, as well as some of the details of Fatou and Julia's works. Audin's book [**2009**] discusses the events, people and mathematics surrounding the 1918 *Grand Prix* and follows certain reverberations through the 1960s. Histories of the other phases have yet to be written, although texts by Burckel [**1979**], Milnor [**2004**], and Steinmetz [**1993**], as well as Bergweiler's paper [**1993**], contain rich historical details.

[1] A note on terminology: The phrase *rational* (or *holomorphic*) *dynamics* refers to the iteration of functions which are analytic on the Riemann sphere $\overline{\mathbb{C}}$. Likewise, *transcendental dynamics* refers to the study of the iteration of a function defined throughout the plane with the point at infinity an essential singularity. *Complex dynamics* refers to the iteration of complex functions in general which includes rational dynamics, transcendental dynamics, meromorphic dynamics, as well as the iteration of algebraic functions. By *meromorphic dynamics* we mean transcendental functions with multiple poles (see Bergweiler [**1993**] for more on this subject). Unless otherwise specified, the functions we refer to are of a single complex variable.

PIERRE FATOU

FIGURE 2.1. Photo courtesy of Pierre Fatou's family members

2.2. Our story begins: Fatou's 1906 notice

As we observed in the last chapter, Schröder and Gabriel Kœnigs (1842–1917) independently mapped out a robust local theory of iteration in the last third of the 19th century. Beset by a seeming inability to extend the study beyond the neighborhood of a fixed point, interest in the study of iteration waned in the late 1890s and the early 1900s, and the study stalled.[2]

Progress came early in the new century with Pierre Fatou's *Comptes Rendus* notice [**1906b**] where he examined the iteration of two kinds of rational functions.[3]

[2]While there were no French studies of substance focusing on dynamical properties of iteration for several years following Leau's work, there was an interest in the functional equations related to iteration. See, for example, Carlo Bourlet's papers [**1897a, 1897b, 1899**], which examine iteration as well the related functional equations from the point of view functional operators. The Italian mathematician Salvatore Pincherle explored ideas similar to Bourlet's around the same time in [**1894-95, 1896a, 1896b, 1896c, 1897a**], and Pincherle's survey [**1912**, §7,§16] of functional equations discusses Bourlet's papers in the context of functional analysis. The lack of discussion of Bourlet's line of investigation in the work of Fatou and Julia suggests that his work did not have significant influence on the development of their ideas.

[3]There was one another mathematician who published an important paper on iteration in the first years of the 20th century, Lucyan Böttcher, whose work we touched upon in Section 1.7.2. His papers [**1904a, 1904b**] appeared in a Russian journal in 1904, but, perhaps due to the language barrier, there was virtually no mention of them in France until Fatou's monograph [**1919-1920**].

The first, which as Fatou observed is typified by the family $f_k(z) = z^k/(z^k+2)$ where k is a natural number greater than one, has a single attracting orbit consisting of a fixed point at the origin, and the second has two attracting orbits, one, again, a fixed point at the origin, and the other the point at infinity. The only example Fatou provided for the second kind of function was $g(z) = (z^2+z)/2$.

Fatou successfully solved the problem of the division of the plane referred to in Section 1.8 for both kinds of functions. The nature of the respective divisions is quite different. For functions f such as the family f_k, Fatou argued that all points in $\overline{\mathbb{C}}$ aside from those in a *totally disconnected perfect set*[4] J converge under iteration to the origin. Points from J remain in J under iteration by f; moreover, it is clear that solutions to $f^n(z) = z$ other than zero are in J as well—something Fatou pointed out in a footnote.

For g, both the origin and infinity are attracting fixed points, hence there are two domains of convergence. Fatou observed that a simple closed curve Γ divides these domains but is not analytic, by which he meant that points on Γ generally do not posses a well-defined tangent line. Both J and Γ are examples of the Julia set discussed in Section 1.6.2.

Comptes Rendus notices typically do not contain rigorous details of proofs, and Fatou's was no exception. Nonetheless, he indicated why the behavior he observed occurred and emphasized that the behavior observed in both g and f_k are not specific to his examples. Color Plates 2 and 3, respectively, depict the dynamics of f_2 and g.

Caption for Color Plate 2: (See p. 63.) We use a representative of the family $\{f_k\}$ referred to in Fatou's 1906 *Comptes Rendus* notice [**1906b**],

$$f_2(z) = \frac{z^2}{z^2+2},$$

to illustrate the iteration a function whose Julia set J is a Cantor set. The function f_2 admits fixed points $O = 0$, $A = \frac{1}{2} + i\frac{\sqrt{7}}{2}$, and $B = \frac{1}{2} - i\frac{\sqrt{7}}{2}$ (not indicated). The origin is superattracting, while A and B are repelling and belong to the Julia set J. All points on the complex plane except those in J give rise to orbits that converge to the origin.

These images are obtained via the escape-time method (see Appendix P) which assigns each point a particular color depending upon how many iterations are needed for its orbit to enter a predetermined region R surrounding an attracting point. In our case the region R is the dark blue ellipse centered at the origin. In each image, points whose color is closer to blue take fewer iterations to reach R, while those closer to red require more.

Top right: A blow-up of the region near A in the left image. Bottom right: A magnification of the region indicated by the box in the upper picture.

[4]Totally disconnected perfect sets are closed sets with no isolated points (every point of the set is a limit point of the set) whose maximal connected components are single points. The Cantor middle third set is perhaps the most well known totally disconnected perfect set, and for that reason such sets are often referred to as Cantor sets, a terminology we will adopt.

Fatou began his construction of the set J for a rational function f of the first kind by defining a circle C of radius r surrounding the fixed point where r satisfies two conditions. On the one hand, it has to be sufficiently large so that all the critical values of f are interior to C.[5] This ensures that all points on C and its exterior will have d preimages, where d is the degree of f. On the other hand, r must be small enough so that all points interior to C remain in its interior under iteration by f before converging to the origin. However, for the example of functions of the first kind that Fatou provided—the family $\{f_k\}$ given above—such a C does not exist because each f_k has the critical value $z = 1$, and any circle large enough to contain $z = 1$ on its interior no longer has the property that its interior is invariant under iteration by f_k.[6]

This does not present a serious problem since one can instead use a closed curve containing the critical values. We suspect that Fatou took it for granted that the reader would understand this. For example, the boundary of the dark blue region in Color Plate 2 is an ellipse with $z = 1$ in its interior for which the properties hold that Fatou claimed for C.

Fatou's analysis of the iteration of f relied on a family of sets E_n. His definition of these sets is a bit muddled since he defined E_n as the set of points $\{z\}$ such that $|z| > r, |f(z)| > r, \ldots, |f^n(z)| \geq r$ while $\left|f^{n+1}(z)\right| < r$, which is the set of points that remain outside of C for exactly n iterations but are interior to C starting with the $(n + 1)$st. This creates a family of disjoint sets since $z \in E_n$ satisfies $\left|f^{n+1}(z)\right| < r$ while all $z \in E_{n+1}$ satisfy the contrary condition $\left|f^{n+1}(z)\right| \geq r$. However, not only did he remark that the sets are contained within one another, but his argument that J is a Cantor set relies upon their nesting, [**1906b**, pp. 546–547].

In order to make the definition of E_n consistent with the details of Fatou's argument, we take the liberty of removing the condition that $\left|f^{n+1}(z)\right| < r$; hence, the E_n can be viewed simply as the set of points that remain exterior to C for at least n iterations. Since the points of E_{n+1} remain exterior to C for $n+1$ iterations, E_{n+1} is contained in E_n, as Fatou stipulated.

The set E_n has d^n components, where d is the degree of f, each of which is bounded by a simple closed curve denoted collectively as Γ_n that, because of the definition of E_n, satisfies $f(\Gamma_n) = \Gamma_{n-1}$. Fatou observed that if, beyond a certain index, $|f'(z)| > A > 1$ on Γ_n, then the circumference of the Γ_n go to zero as $n \to \infty$; hence, the intersection of the closure of the E_n (the set of points $\{z\}$ for which $f^n(z)$ is exterior to C for all n) gives rise to infinitely many nestings, each of which intersects down to a single point. For example, referring to the upper right image in Color Plate 2, we obtain one nesting if we take a route beginning with the upper light green component, followed by the leftmost yellow, orange, and red-components, and so forth, and another if we instead follow rightmost components.

[5]Recall from Chapter 1 that the critical values of a function are the branch points of the various branches of its inverse; hence for f_2, for example, two distinct inverses do not exist at the critical points. See Definition 1.11. For the functions f_k the critical values are $z = 0$ and $z = 1$.

[6]It is possible to choose a circle for C and still obtain a Julia set in the manner Fatou describes, but one gives up either the invariance of iteration on the interior of C or the property that the preimage of C consists of discrete curves. For example, if C is the unit circle, the preimage of C under f_2 consists of two branches of a hyperbola that meet at the point of infinity, forming a figure eight, which is replicated in subsequent preimages. For a depiction of the dynamics of f_2 when C is the unit circle, see Alexander [**1994a**, §7.5].

Since each component of a given E_n contains infinitely many other components of higher order whose diameters go to zero, within a given ϵ of each nesting will be infinitely many others, so each point in J will also be a limit of J. The result is a totally disconnected perfect set.

The analogy between J and the Cantor middle-thirds set is obvious. One way to view the formation of J is to first remove the points inside C, then the points which arrive in the interior of C after one iteration, and so forth.

Consistent with the nature of a *Comptes Rendus* notice, Fatou provided very little justification for his assertion that the lengths of Γ_n go to zero, which is the cornerstone of his argument that J is a Cantor set. A little over a decade later he gave a similar construction in his monograph [**1919-1920**, Chap. IV, p. 245 ff.],[7] where he showed that the Julia set for a particular class of rational functions is a totally disconnected perfect set.[8]

As was the case in his 1906 notice, Γ_n comprised d^n closed curves. Fatou showed that for a representative function f from the class he was considering, there was an N such that when $n \geq N$, $|f'(z)| > A > 1$ on Γ_n, which implies that the length of Γ_n (and with it, those of the curves that compose it) go to zero. His justification for this last statement was based on differentials and is essentially as follows.

Since $f(\Gamma_n) = \Gamma_{n-1}$, we can write $\gamma_{n-1} = f(\gamma_n)$ where γ_n is one of the curves that make up Γ_n and γ_{n-1} is its image. An application of the chain rule yields

$$d\gamma_n(t)\,|f'(\gamma_n(t))| = d\gamma_{n-1}(t),$$

where $d\gamma_n$ is an arc length element of γ_n. For $n \geq N$, $|f'(z)| > A > 1$ on Γ_n, hence

$$d\gamma_n < \frac{1}{A} d\gamma_{n-1},$$

which, because the union of the γ_n is Γ_n, implies

$$\ell(\Gamma_n) < \frac{1}{A}\ell(\Gamma_{n-1}).$$

It readily follows that

$$\lim_{n\to\infty} \ell(\Gamma_n) = 0.$$

Was this the argument he had in mind in 1906? We of course cannot be certain. But the fact that his justification from [**1919-1920**] followed the outline he had sketched in 1906 suggests that it may have been.

The dynamics involved in the iteration of the second kind of function, typified by the function $g(z) = (z^2 + z)/2$, are quite different from those of the f_k. As noted above, Fatou observed that for g, the boundary between the domains of convergence of $z = 0$ and the point at infinity is a simple closed curve Γ whose points do not in general posses a well-defined tangent line. The curve Γ and the domains of convergence are depicted in Color Plate 3.

Caption for Color Plate 3: (See p. 63.) At left: The Fatou and Julia sets for the function

$$g(z) = \frac{z^2 + z}{2},$$

[7]We will discuss this monograph briefly in the next section and in much greater detail in subsequent chapters.

[8]See Theorem 7.13 in Chapter 7.

also from Fatou's 1906 notice, which admits the fixed points $O = 0$ (superattracting), $A = 1$ (repelling), and the point at infinity (attracting). Points in the red region converge to zero under iteration; points on its boundary are in the Julia set J. Points shaded blue are in the basin of attraction for the point at infinity and are colored according to the number of iterations it takes for their moduli to reach a particular magnitude M.

At right: Two magnifications near $z = 1$ reflecting the fact that J retains the same general pattern of irregularity regardless of how much it is magnified. This property is often referred to as self-similarity or homogeneity. The upper image at right is a magnification of the region near A indicated in the larger image. The lower image is a magnification of the region indicated by the square in the top right image. The cusps indicate why tangents do not exist: they pervade Γ and become more evident upon subsequent magnifications.

Fatou's justification that Γ is not analytic is a little obscure. He noted first that Γ is symmetric with respect to the real axis, hence, if Γ is analytic, then there would be an analytic function φ mapping a point z on Γ to its complex conjugate, $\overline{z}$, which, due to Γ's symmetry, is also on Γ. Although he did not say why this is so, it follows from the Schwarz reflection principal and φ—whose domain can be extended to an annulus surrounding Γ—is sometimes referred to as a Schwarz function. Since g is a polynomial with real coefficients, one can easily show that $\overline{g(z)} = g(\overline{z})$, from which $\varphi \circ g = g \circ \varphi$ readily follows. Without going into details, Fatou noted the fact that g and φ satisfy the system of functional equations $\varphi \circ g = g \circ \varphi$ and $\varphi^2 = I$ leads to a contradiction;[9] hence he concluded that Γ is not analytic.

It is unclear what argument he had in mind, but an argument for a similar proposition from his *Comptes Rendus* notice [**1918a**] may offer some clues. In that article he showed, under similar conditions, that Γ is analytic only if it is a circle. His argument relied on the Schwarz reflection principle and asserted that if Γ is analytic, it is the image under a linear mapping of a circle surrounding the origin. Perhaps he had an inchoate version of this argument in mind. Indeed, if φ is linear, one can use the conditions $\varphi(-2) = -2$, $\varphi(1) = 1$ (which follow from the fact that $z = 1, -2$ are points on Γ) and $\varphi^2 = I$ to deduce an explicit equation for φ,

$$\varphi = \frac{4 - z}{1 + 2z}.$$

Since g and φ are defined on a set with a limit point, the relation $\varphi \circ g = g \circ \varphi$ holds beyond Γ. But one can easily choose a point not on Γ for which $\varphi \circ g = g \circ \varphi$ does not hold.

An example where Γ is analytic occurs with $g(z) = z^2$, where $\varphi(z) = 1/z$, Γ is the unit circle, and the system of functional equations $\varphi \circ g = g \circ \varphi$ and $\varphi^2 = I$ is not only satisfied on Γ, but throughout $\mathbb{C}$. The inference from Fatou's argument from 1918 is that if the dynamical system given by $z \mapsto z^2$ is perturbed in such a way that Γ is not a circle, its Julia set is no longer an analytic curve.

Because it appeared during one of those dips of interest that occur periodically within the study of iteration, Fatou's groundbreaking notice garnered no response

[9] I represents the identity function.

for quite some time. Even though it would be ten years before Fatou would again take up the study of the iteration of rational functions in print, Audin [**2009**, p. 17], citing a 1907 letter from Fatou to Maurice Fréchet, points out that Fatou continued to explore the subject subsequent to the publication of his 1906 notice.

2.3. Where we are headed

In 1917 Fatou, Gaston Julia, Samuel Lattès, and Joseph Fels Ritt published a flurry of papers on the global behavior of iterates of complex rational functions. The motivation for Lattès and Julia—and most likely Fatou—was the 1915 announcement by the French Académie des Sciences that its 1918 *Grand Prix des Sciences Mathématiques* competition would be devoted to the study of the iteration of complex functions. This flurry became a blizzard, and two lengthy monographs from Fatou and Julia ([**1919-1920**] and [**1918a**], respectively) brought a substantial global theory of rational dynamics into being.

There was still work to do, however. Despite their best efforts, neither Fatou nor Julia were able to determine the iterative behavior of rational functions near an irrationally neutral fixed point, and they left essential questions unanswered. An understanding of iteration near these points began to emerge in the late 1920s with the work of the young German mathematician Hubert Cremer [**1927**], and Carl Ludwig Siegel gave a conclusive answer in [**1942**] to one of the most significant questions left open by Fatou and Julia.

Moreover, the studies Fatou and Julia prepared in anticipation of the *Grand Prix* focused almost exclusively on the iteration of rational functions. Consequently, very little was known about the global behavior under iteration of other kinds of complex functions. In an attempt to remedy this, Fatou published two introductory papers, one on the iteration of algebraic functions [**1922a**] and another [**1926**] on transcendental dynamics. These papers found little response for quite some time. Due in part to the disruptions caused by World War II, Siegel's 1942 paper signaled an end to the phase of complex dynamics that began with Fatou's 1906 notice.

It would be another 40 years or so before the study of complex dynamics would again receive the sort of attention that it had at the time of 1918 *Grand Prix*. Beguiling images such as the Mandelbrot set and Newton's method for the cubic circulated widely in the early 1980s, even gaining the attention of audiences beyond the community of mathematicians.[10] Images which Fatou and Julia could only describe with metaphor could now be seen. Extraordinarily bountiful lines of inquiry opened, and advances in topology and analysis provided the tools mathematicians needed to analyze what they saw. The field had entered its mature phase.

Before we continue, we would like to dispel the commonly held notion that very little happened between 1942 and the studies of Adrien Douady, John H. Hubbard, and Benoit Mandelbrot of the late 1970s and early 1980s. Not only did Irvine Noel Baker develop a substantial research program on the iteration of entire, transcendental maps beginning in 1955 (Rippon and Stallard's tribute to Baker lists 79 papers by Baker related to the iteration of entire maps [**2008**, pp. xv–xix]

[10] The August 1985 popular science magazine *Scientific American* featured the Mandelbrot set on its cover, and complex dynamics was often portrayed as part of the emerging science of chaos, which caught the popular imagination in the 1980s. The best-selling general-audience book *Chaos* by James Gleick, published in 1987, devoted a chapter to complex dynamics.

over half of which appeared before 1980) but we found well over 20 other works on the iteration of complex functions written between 1942 and 1980.

Some of these works are a residue of genealogy: Cremer's student Hans Töpfer wrote two papers on iteration [**1939, 1949**], as did his own student Bruno Klingen [**1962, 1968**]. Likewise, Baker's students Prodipeswar Bhattacharya [**1969**] and Lennox Liverpool[11] [**1971**] devoted their doctoral dissertations to iteration. Other works written during this period include Hans Brolin [**1965**], Thomas Cherry [**1964**], Maurice Heins [**1941**], Pekka Myrberg [**1960**], Hans Rådström [**1953**], and Paul Charles Rosenbloom [**1948, 1952**].

Perhaps one reason for the perception that interest in iteration withered is that its principals did not continue to research it: Ritt, Julia, and Cremer moved on to study other things; Fatou died early, at age 51, in 1929; Siegel's interest in complex dynamics was passing, his 1942 paper being a bit of a one-off. Nor, with the exception of Cremer, did they develop students who carried on the study. But interest did not fade after 1942 or even go into hiatus. Rather it persisted, but just not as prominently as it once had and would be again.

In the next two chapters we devote considerable space to the study of differential equations. While the studies of iteration as practiced by Fatou and Julia are divergent from the use of iteration in celestial mechanics, as exemplified by Henri Poincaré, the two studies are not unrelated. First of all, in its *Grand Prix* announcement the Académie referred to Poincaré's use of iteration. We wish to explore that reference with a discussion that includes Poincaré's return map. To put the return map in its proper context requires an excursion into Hamiltonian systems and the concept of stability. Moreover, Lattès extended Poincaré's method of investigating differential equations and adapted them to his own study of the functional equations that naturally arise in iteration, and our examination of Poincaré's return map will help the reader to fully appreciate what Lattès did.

We ask those readers whose interest is not differential equations to have patience: we will return to the study of complex dynamics, but not before we embark on an examination of the small divisors problem in the context of celestial mechanics and perturbation theory in Chapter 4. We do this for two reasons: The first is that small divisors surface at several turns in our text. They are the principal impediment to understanding the dynamics of iteration near an irrationally neutral fixed point, and they vexed many of the mathematicians we discuss, from the American mathematician George Adam Pfeiffer, who was the first to encounter them in the context of complex dynamics around 1915, to Fatou, Julia, and Cremer. We hope that our portrayal of the difficulties that small divisors problems posed for these mathematicians will foster a deeper appreciation of their struggles and accomplishments. The second reason is that Siegel's demonstration of the existence of centers[12] in complex dynamics represents a rare—at least for the time—success with a small divisors problem. This event influenced subsequent developments in KAM theory, an influence we explore first in Chapter 4 and then, again, in more detail in Chapter 12.

We return, finally, to complex dynamics in Chapter 5 with an account of the early 20th century developments in the U.S., which largely took place prior to the

[11] Liverpool's own student, Kenneth Yugunda, wrote his thesis on iteration as well, in 1998, at the University of Jos in Nigeria.

[12] Centers are discussed in Section 1.6.2.

U.S. entry into World War I, and before the appearance of the works of Fatou and Julia prepared in anticipation of the 1918 *Grand Prix*. While informed by the work of Kœnigs and his students, particularly their work with functional equations, Edward Kasner's study of invariants in conformal geometry [**1913**] was also essential to the beginnings of the American study. The work of two of the Americans discussed in this chapter, Ritt and Pfeiffer, factors into subsequent European developments.

The 1918 *Grand Prix des Sciences Mathématiques* is obviously at the center of our story, and we devote the next several chapters to events revolving around the competition for the prize. We examine the responses of Fatou, Julia, Lattès, and others, as well as questions left unanswered. Paramount among the open questions is what is often referred to as the *center problem*, which seeks to determine whether the iteration of a nonlinear rational function near an irrationally neutral fixed point is ever conformally equivalent to a rotation of the unit disc.

A reason to suspect that it might be stems from an examination of the canonical Schröder functional equation discussed in Section 1.4.1,

$$S \circ f = f'(\lambda) \cdot S, \tag{2.1}$$

where λ is a fixed point of f. This equation is fundamental to the study of iteration, and almost everyone who studied iteration in our story considered it. As we noted, when λ is an attracting but not superattracting fixed point (i.e., the modulus $|f'(\lambda)|$ is strictly between zero and one), this equation can always be solved, and its effect is to linearize iteration via the mapping $z \mapsto f'(\lambda)z$. It is natural to ask, then, if solutions exist in the irrationally indifferent case. If so, iteration would be conformally equivalent to the rotation $z \mapsto e^{i2\pi\theta}z$ where $f'(\lambda) = e^{i2\pi\theta}$.

However, due in large measure to issues involving small divisors, proving the existence of such behavior is quite difficult, and though they tried, Fatou and Julia made little progress toward the solution of the center problem. Nonetheless they explored what might happen if centers did exist and gained a not inconsiderable insight into the dynamics of the problem. We describe their examination in Chapter 9. In Chapter 10, the last to treat the *Grand Prix* and its immediate aftermath, we survey the study of iteration in the 1920s.

In Chapters 11 and 12 we continue with the history of the center problem. The former chapter concerns the mathematics of Cremer who, in the late 1920s and 1930s, found conditions that imply solutions to equation (2.1) do not exist; the latter chapter discusses Siegel, who found conditions under which they do. The final chapter surveys mathematicians around the world who studied iteration of complex functions in the 1920s and 1930s.

Our book closes with a series of appendices by present-day researchers that treat technical details involved in the topics we investigate and provide additional historical perspectives by describing subsequent developments. We thus hope to link old discoveries to current investigations, thereby providing a fuller context to our discussion. Included in the appendices are brief accounts of the lives of Fatou and Julia, capsule biographies of other mathematicians involved in our story, and a discussion of the technical details of the computer graphics we include.