

Newton's method applied to two quadratic equations in \mathbb{C}^2 viewed as a global dynamical system

John H. Hubbard and Peter Papadopol

In this paper, we will study Newton's method for solving two simultaneous quadratic equations in two variables.

In one dimension, if F is a polynomial, the Newton mapping is a rational function and we can apply the now rather well developed theory of one-dimensional complex analytic dynamics. The subject is far from completely understood, but much progress has been made, particularly by J. Head, Tan Lei, and M. Shishikura. More recently, [HSS] give precise results on how to find all the roots of a polynomial, based on the topology and complex analysis of the basins.

Presumably, there is no need to motivate a study of Newton's method, in one or several variables. The algorithm is of immense importance, and understanding its behavior is of obvious interest. It is perhaps harder to motivate the case of two simultaneous quadratic equations in two variables, but this is the simplest non-degenerate case.

The classical theory, due to Ostrowski and Kantorovitch, asserts that a non-degenerate zero of a C^2 function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a superattractive fixed point of the associated Newton map

$$N_f(x) = x - [Df(x)]^{-1}f(x).$$

For an elementary presentation of this theory, see [HBH], Section 2.7; we will follow the notation used there. Shub and Smale [ShSm] have also made an extensive study of Newton's method from the point of view of complexity theory.

When $f : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ is analytic, a much more detailed description of the dynamics near a fixed point is given in [HP], which gives a partial description of the basins of the roots. In that paper, we show that associated to each root x_0 there is a rational function $g_{x_0} : \mathbb{P}^1 \rightarrow \mathbb{P}^1$ constructed from the quadratic terms of the Newton map at the root. This rational function (its Julia set, critical orbits, . . .) is the most important element in the description of the basin, and in the present paper it will also play an important role.

What makes N_f such an intractable dynamical system is the simultaneous presence of critical points and points of indeterminacy. There is an extensive theory for the dynamics of birational maps $\mathbb{P}^2 \rightsquigarrow \mathbb{P}^2$ [BS1-8],[FS1-3], [HO1-2],[Dil], [Br], [Jo], though in no sense do they give the sort of information available for functions of one complex variable. Of particular relevance to this paper is [HPV], where the technique of real oriented blow-ups is introduced. These play an important role here too; in [HPV], we showed something of what complications can lurk in points of indeterminacy (solenoids, etc.), but here the same sort of construction yields for more elaborate structures.

There is also a reasonable theory of endomorphisms of \mathbb{P}^n [HP],[FS], [Ueda], though again the descriptions provided are not nearly so precise as one might like, and there is a drastic shortage of examples that are actually understood.

But next to nothing is known about mappings that have topological degree > 1 (and in particular, critical curves), and points of indeterminacy. One important result is due to Russakovskii

and Shiffman [RS]. There they prove that if $f : \mathbb{P}^n \rightsquigarrow \mathbb{P}^n$ is a rational mapping, and various inequalities on the algebraic degree and topological degree are satisfied, then if you take a “generic” point, take all its inverse images by f^n , and give them all equal weight so as to form a probability measure μ_n , then these measures converge as $n \rightarrow \infty$ to a measure μ independent of the original point. Here *generic* means “chosen off a pluripolar set.”

Newton’s method for solving two simultaneous quadratics does satisfy their conditions; in fact, Russakovskii and Shiffman consider specifically the Newton map associated to

$$f(x, y) = (y - x^2, x - y^2).$$

This map, like all the Newton maps N_f associated to pairs of quadratics, has two points of indeterminacy at infinity. Russakovskii and Shiffman could not determine whether their measure charges the points of indeterminacy at infinity; in this paper we prove that it does not (Corollary 3.3.3).

Outline of the paper

This paper is organized into six chapters.

Chapter 1 presents various facts about Newton’s method in general, and some specifics about the Newton method as applied to two quadratic equations in two variables. All the results follow from straightforward computations, but some are rather surprising:

- The Newton mapping N_f associated to the equations $f(x, y) = 0$ to find the intersection of two conics depends only on the points of intersection, and not on the equations of the conics passing through these points. In particular, Newton’s method to intersect two conics is conjugate to a normalized one where the conics intersect at $(0, 0)$, $(1, 0)$, $(0, 1)$, and (α, β) , and (α, β) is unique once the intersection points are marked.
- The lines joining pairs of roots are invariant under N_f , and the restriction of N_f to the line is simply Newton’s method for solving the quadratic equation for the two roots in the line.
- Newton’s method has topological degree 4; since it also has local degree 4 at the roots, it follows that the basins of the roots are connected.
- If the Newton map is normalized as above, then the rational function of degree 2 at each root runs through all rational functions of degree 2 as (α, β) vary. In particular, these Newton maps have at least the complexity of all rational functions of degree 2.

We also analyze the points of indeterminacy (there are five of them), the critical locus, which turns out to be the union of two non-singular cubics, and the critical value locus, which consists of the union of two parabolas.

Chapter 2 is concerned with stable and unstable manifolds of invariant subsets. One important result is Theorem 2.4.1. As stated above, each line joining a pair of roots is invariant, and the restriction of Newton’s method to such a line has a Julia set, which is the real straight line bisecting the two roots in the complex invariant line. The question whether this Julia set has a stable manifold is quite subtle. It never has a genuine stable manifold, but sometimes there is a 3-dimensional manifold associated to it which has some features of a stable manifold and some features of a center manifold; this is the content of Theorem 2.4.1. Under more general conditions, it is an *ergodic attractor* [PS], and has a measure-theoretic stable manifold, the topology of which is still pretty mysterious.

In Chapter 3, we examine a case where the algebra simplifies: the intersection of the parabolas $y = x^2$, $x = y^2$. For this map, we will analyze the behavior at infinity, using infinitely many blow-ups. One pay-off is that we are able to untangle the regions at infinity where Newton’s method

is attracting and repelling, and show that the line at infinity is not charged by the Shiffman-Russakovskii measure.

In Chapter 4, we analyze in the abstract the process of infinitely many blow-ups used in Chapter 3. This construction, which we call the *Farey blow-up*, is fundamental if one wants to understand the behavior of mappings with critical points and points of indeterminacy. In some sense, the idea of constructing this blow-up goes back to Max Noether, whose “monster” has intrigued many mathematicians [Ma], [De], without ever really becoming an accepted citizen of algebraic geometry, because the topology of the monster is so wild.

In contrast, we will show that the topology of Farey blow-ups is fairly tame, and quite easy to describe in the language of *real oriented blow-ups*, as pioneered in [HPV]. In a rather surprising coincidence, the topology of these real oriented blow-ups is reminiscent of the topology of non-integrable Hamiltonian systems with two degrees of freedom, with zones corresponding to rational “frequencies” separated by tori corresponding to irrational frequencies.

We also compute the second homology groups of Farey blow-ups. This is a space with a natural quadratic form coming from the intersection product. This quadratic form is negative definite, and the completion of the homology with respect to the quadratic form turns out to be a very classical Sobolev space of functions with one derivative in L^2 . We also compute the operation of appropriate mappings on the homology.

In Chapter 5, we perform infinitely many Farey blow-ups to construct a compact space X_∞ on which the Newton map operates, and analyze in detail the structure at infinity.

The real oriented blow-up of X_∞ has a 3-dimensional boundary, and now the resemblance with the structures appearing in the KAM theorem [Ar] is uncanny. “First-order tori” corresponding to irrational numbers, and carrying irrational foliations, separate zones corresponding to rational numbers. Within these zones, further “second-order tori” again with irrational foliations, separate zones corresponding to rationals, etc., the whole thing accumulating on some Cantor set of solenoids.

Another way of thinking of this structure is in terms of Puiseux expansions, with the first-order tori corresponding to certain one-term Puiseux expansions of transcendental curves, separating zones corresponding to Puiseux exponents of algebraic curves. Within these zones, there are further structures corresponding to two-term Puiseux expansions, some transcendental corresponding to tori, and some algebraic, which can then be further enriched, etc. We do this in the case of the same special Newton map as in Chapter 3, to simplify the computations.

In Chapter 6, we show that essentially all the results of Chapters 3 and 5 generalize to any Newton map associated to a pair of quadratic equations in two variables (except some degenerate cases).

Acknowledgments

We thank Jeff Diller, who helped with the computation of Section 2.1. Robert Strichartz saw the connection of the homology with the Sobolev space \mathcal{H}^1 . Eric Bedford made many helpful comments about an early version of the paper. Karl Papadantonakis contributed the computer program which made the color pictures. Ramin Farzaneh wrote an earlier program, which raised many of the problems we address in this paper. A big thanks to Barbara Burke Hubbard, for her help with editing, and finding several mathematical mistakes.

Peter Papadopol thanks the Department of Mathematics of Cornell University, and the chairmen Keith Dennis, Peter Kahn, Robert Connelly and especially John Smillie, for many years of hospitality. From Grand Canyon University, he thanks former President Bill Williams for his support for many years, President Gil Stafford and Chairperson Beth Dawkins for support and understanding.

A computer tour of Newton's method

Before embarking on the rigorous part of this paper, we want to give an overview of what computer exploration shows, emphasizing those aspects that we will be able to explain.

The example we will examine is the Newton's method for finding the intersections of two conics when the intersections are at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, and $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$; as we will see in Corollary 1.2.2, this specifies a unique Newton mapping. As far as we know, there isn't anything very special about these values of the parameters, except that they are real, and we can draw pictures of the basins in \mathbb{R}^2 , shown in Figure 1.

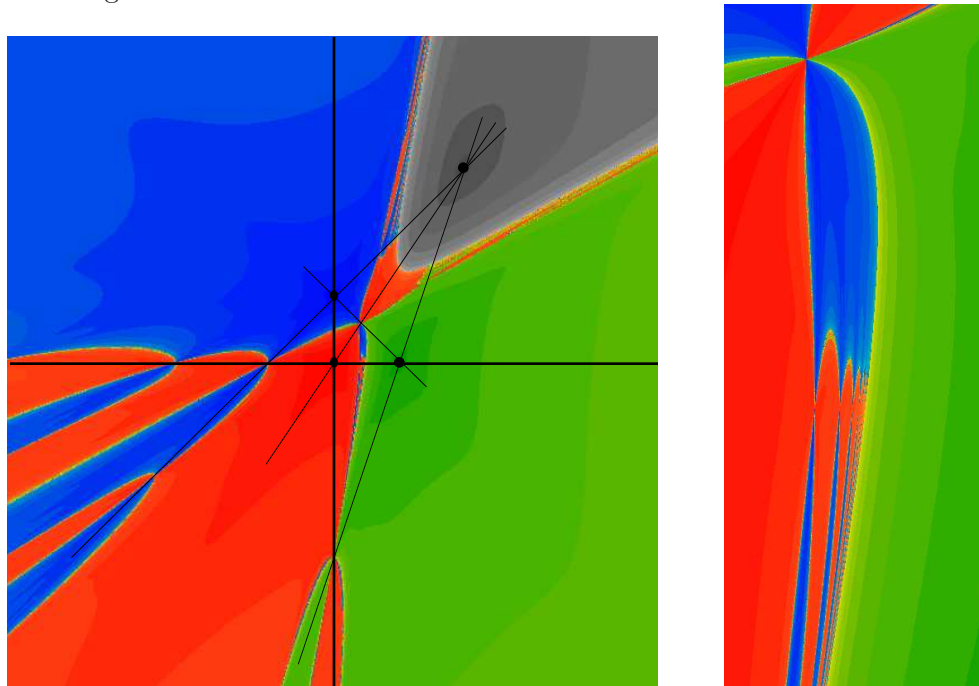


FIGURE 1. At left, we see the basins of the roots, marked with black dots: the basin of $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ in red, the basin of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in green, the basin of $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in blue, the basin of $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ in grey. The region displayed is $-4 \leq x, y \leq 4$. Notice that the x -axis is entirely red and green, and the y -axis is entirely red and blue, in fact all the lines joining pairs of roots contain only two colors. At right is a blow-up of the figure at left; notice the blue fingers coming from above and below to touch the x -axis.

The main thing we see in these pictures is the importance of the *polar locus*, i.e., the preimage of the line at infinity. In Figure 1, it is the hyperbola, one branch of which bounds the basin of $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$, drawn in grey in the picture. The other branch is a bit harder to see; one part bounds the

basin of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The inverse image of this polar hyperbola is a curve of degree 6, which has three branches in this picture. It is a bit difficult to see, since two of its branches are almost superposed to the hyperbola.

Figure 2 represents Newton’s method applied to two quadratic equations, in the case when the roots are at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, and $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$. The polar curve is the ellipse one sees emphasized in the middle. More generally, if two real conics intersect in four real points that form a convex quadrilateral, then the polar locus is a hyperbola, but if one intersection is in the convex hull of the others, the polar locus is an ellipse. This time the real part of its inverse image is the 3-leafed clover, a bit more lightly emphasized. Actually, this is only one branch of the inverse image; there is another branch in \mathbb{C}^2 .



FIGURE 2. Left: the basins of the roots $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ (red basin), $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (green), $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ (blue), and $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$ (grey). The polar curve is an ellipse in this case, and it together with its first and second inverse images are drawn on the right. They appear to form the boundary of the basins of the roots, or at least it seems that the closure of the union of the iterated inverse images of the polar curve does.

It is not really surprising that the inverse images of the line at infinity appear to make up the boundaries of the basins of the roots. After all, the line at infinity itself is never in the basin of any root, so neither are its inverse images.

Since our objective is to understand the dynamics of Newton’s method in \mathbb{C}^2 , we need to find a way of making complex pictures. We will also need to find the complex analog of the line at infinity, and its successive inverse images. Of course, the line at infinity is still there, and so are its inverse images, but they are now of codimension 2, and cannot form the boundary of anything.

The most obvious picture to make is a 1-dimensional complex slice through \mathbb{C}^2 , in which we draw the basins. If we use the same Newton map with roots at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, and $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$, and slice by the line $y = (1 + i)x$, then the region $0 \leq \text{Re } x \leq 2$, $|\text{Im } x| < 1$ is represented on the left of Figure 3.

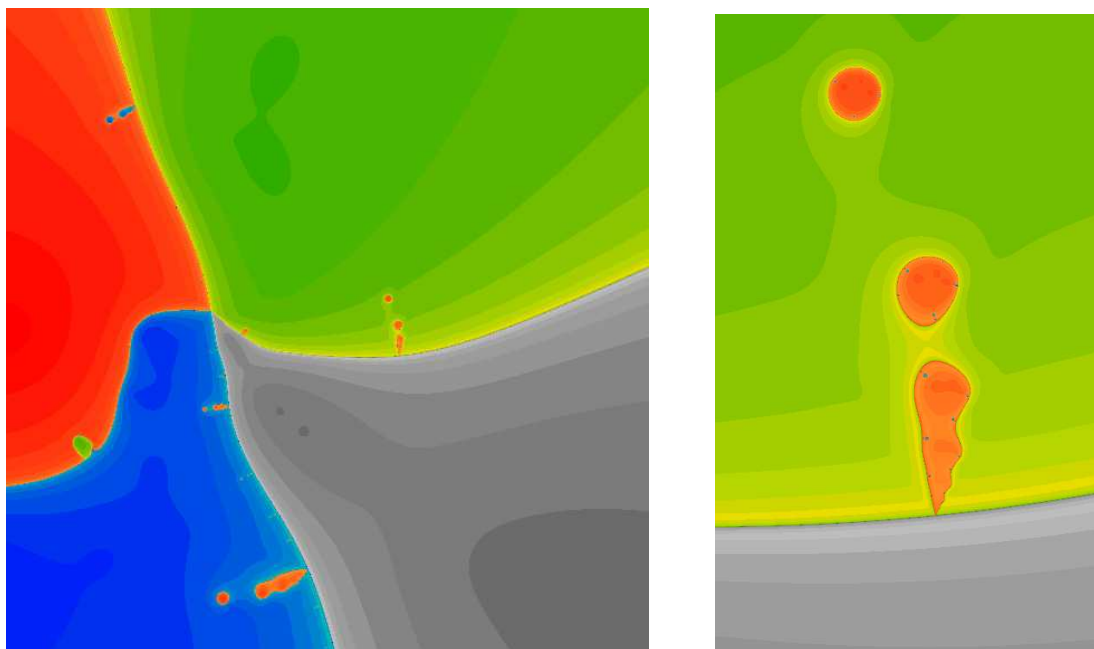


FIGURE 3. The same Newton's map shown in Figure 2, sliced by the complex line $y = (1 + i)x$. The different shades represent the basins of the four roots. On the right, a blow-up of the figure on the left.

What do we see? The basins of the roots are bounded by what seem to be smooth curves. Inside some of the basins there seem to be archipelagoes of islands of a different basin, which appear to converge to a point of a boundary with a third basin, where however the boundary appears perfectly smooth (see Figure 4).

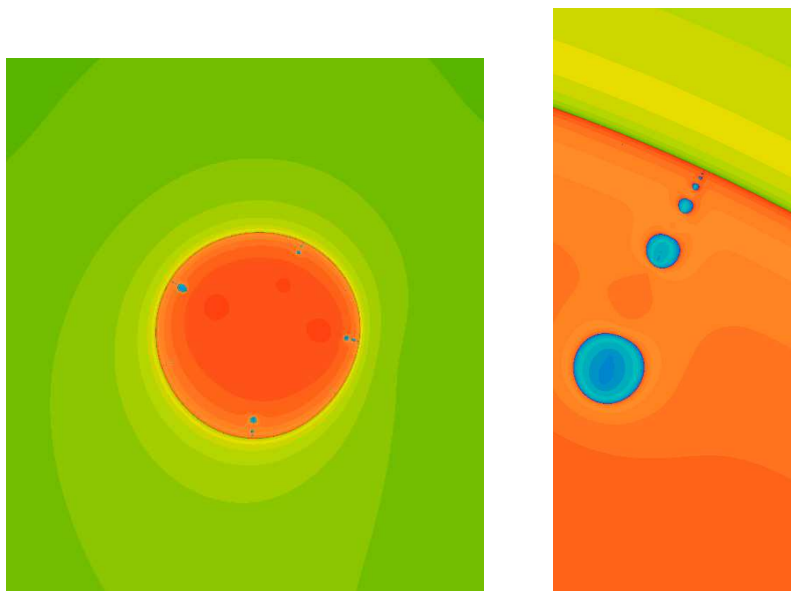


FIGURE 4. A sequence of “bubbles” of one domain inside another basin, converging to a boundary point of a third basin. The left hand figure is the “top bubble” of the right-hand picture in Figure 3, and the right-hand picture is a further blow-up of that. The reader might notice some similarity between the blue “fingers” of Figure 1 and the blue bubbles in this figure; we will explore this similarity in Section 2.5.

In other places, the picture appears to be self similar, but the curves forming the boundaries undulate, and these undulations do not appear to become smaller as they get closer to the center of self-similarity. These curves form a definite angle at the point of self-similarity.

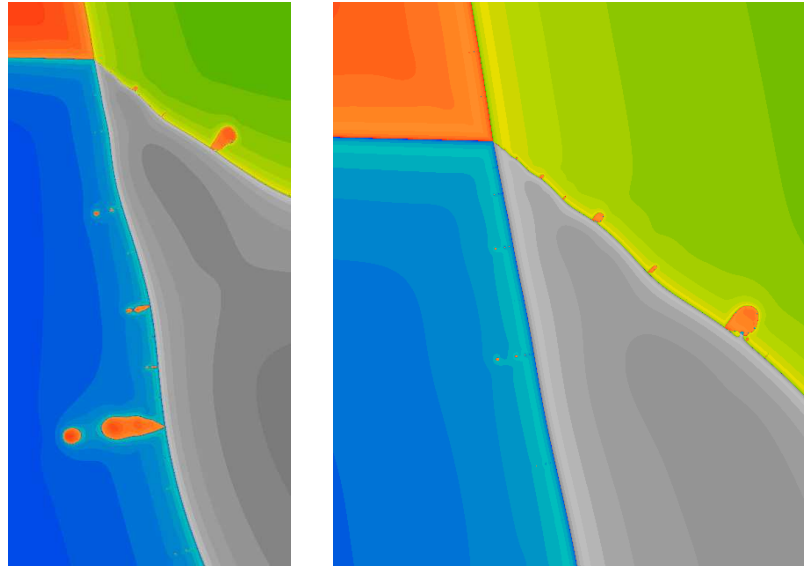


FIGURE 5. A blow-up near the center of the left-hand side of Figure 3, and a further blow-up of it.

The pictures seen so far are misleadingly simple. As we will see in Chapter 2, things tend to be a bit simpler when the coefficients of a Newton map are real. Figure 6 represents the line of equation $y = sx$. We see the basins of the roots for Newton's method, where the roots are at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, where $\alpha = 2 + 2i$ and $\beta = -.3$. Again, we understand roughly why the drawing presents the sort of complication present, without understanding in detail just why the structure is what one sees.

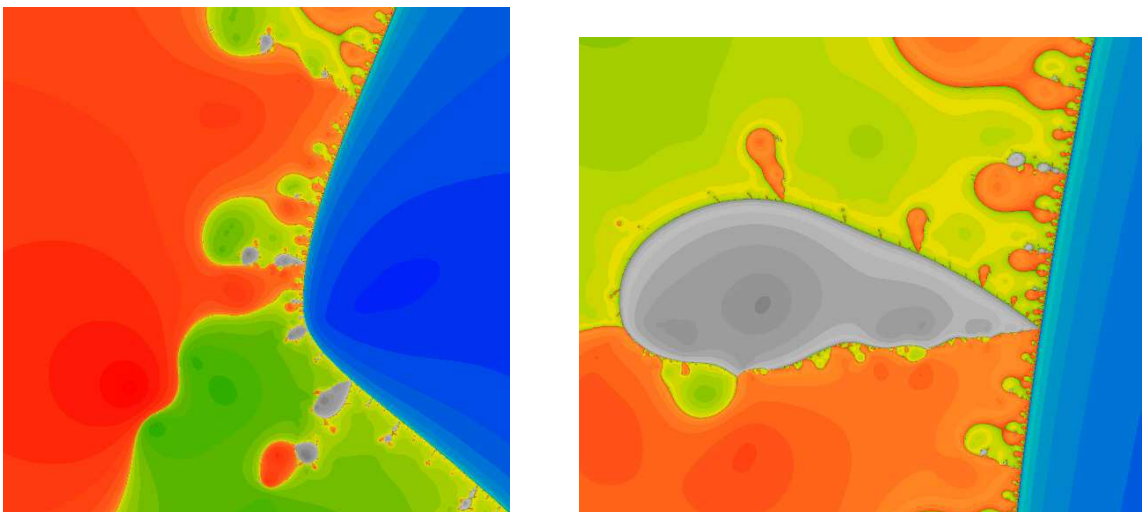


FIGURE 6. At right, a blow-up of the figure on the left.

Some of the phenomena above will be explained in this paper: the apparent smoothness of basin-boundaries, the self-similarity, the undulation of some boundary curves. There are many things about these pictures which we don't understand yet, more particularly the combinatorial properties of all these pictures.

A question we have not been able to solve is: are the basins dense? This hides two quite different questions:

- Are there any attracting cycles other than the basins?
- Are there wandering domains?

We conjecture that the answer to both these questions is no. But the questions are not of equal importance: for the Newton map as applied to polynomials of degree greater than 2, there certainly are other attracting cycles. But we conjecture that there are never any wandering domains for Newton's method as applied to pairs of polynomials in two variables. This stands out as a central problem of dynamics in several variables: the tools used by Sullivan in the proof in one dimension are simply not available.

Remark. The figures shown here are chosen essentially at random; practically any other parameters cut by any line are equally as interesting and complicated. A much larger gallery of pictures can be seen at web site

<http://www.math.cornell.edu/dynamics/>

The site also includes Macintosh programs that the reader can download. \triangle

1

Generalities about Newton's method

Given two vector spaces V and W of the same dimension, and a mapping $F : V \rightarrow W$, the associated Newton map $N_F : V \rightarrow V$ is given by the formula

$$N_F(\mathbf{x}) = \mathbf{x} - [Df(\mathbf{x})]^{-1}(F(\mathbf{x})). \quad (1.1)$$

A well-known theory asserts that if $\mathbf{x}_0 \in V$ satisfies $F(\mathbf{x}_0) = 0$ and if the derivative $[DF(\mathbf{x}_0)]$ is invertible, then \mathbf{x}_0 is an attracting fixed point of the Newton map, and the rate of attraction is at least quadratic.

1.1. Generalities about Newton's method

The Newton map behaves very pleasantly under linear, or even affine, changes of variables.

Lemma 1.1.1. *If $A : V \rightarrow V$ is affine and invertible, and $L : W \rightarrow W$ is linear and invertible, then*

$$N_{L \circ F \circ A} = A^{-1} \circ N_F \circ A. \quad (1.2)$$

Proof. This is a straightforward application of the chain rule:

$$\begin{aligned} N_{L \circ F \circ A}(\mathbf{x}) &= \mathbf{x} - [D(L \circ F \circ A)(\mathbf{x})]^{-1}(L \circ F \circ A)(\mathbf{x}) \\ &= \mathbf{x} - \left([DA(\mathbf{x})]^{-1} \circ [DF(A(\mathbf{x}))]^{-1} \circ L^{-1} \circ L \circ F \circ A \right)(\mathbf{x}) \\ &= \mathbf{x} - [DA(\mathbf{x})]^{-1} \circ [DF(A(\mathbf{x}))]^{-1}(F(A(\mathbf{x}))). \end{aligned} \quad (1.3)$$

On the other hand, we have

$$(A^{-1} \circ N_F \circ A)(\mathbf{x}) = A^{-1} \left(A(\mathbf{x}) - [DF(A(\mathbf{x}))]^{-1}(F(A(\mathbf{x}))) \right). \quad (1.4)$$

Since A^{-1} is affine, we have

$$A^{-1}(\mathbf{y} + \mathbf{z}) = A^{-1}(\mathbf{y}) + [DA]^{-1}(\mathbf{z}) \quad (1.5)$$

for any \mathbf{y} and \mathbf{z} , and we have not indicated where the derivative is evaluated since it is constant. This gives

$$(A^{-1} \circ N_F \circ A)(\mathbf{x}) = A^{-1}(A(\mathbf{x})) - [DA(\mathbf{x})]^{-1} \circ [DF(A(\mathbf{x}))]^{-1}(F(A(\mathbf{x}))), \quad (1.6)$$

which agrees with equation (1.3). \square

It is clear from the definition that even if F is defined on all of V , the associated Newton map will not be defined where the derivative of F is not invertible; we will call this locus the *polar locus* P_F . It is generically of codimension 1, and is given by the single equation

$$P_F = \{\mathbf{x} \in V \mid \det[DF(\mathbf{x})] = 0\}. \quad (1.7)$$

Off the polar locus, the Newton map is differentiable; the derivative is a bit delicate to compute. In one dimension, the derivative of the Newton map is given by

$$(N_f)'(x) = \frac{f(x)f''(x)}{f'(x)^2}. \quad (1.8)$$

We want to extend this computation to several dimensions. This requires a bit of notation to decide what the second derivative is. We will write the Taylor expansion of F in a neighborhood of \mathbf{x} as

$$F(\mathbf{x} + \mathbf{u}) = F(\mathbf{x}) + [DF(\mathbf{x})](\mathbf{u}) + \frac{1}{2}[D^2F(\mathbf{x})](\mathbf{u}, \mathbf{u}) + \dots, \quad (1.9)$$

so that

$$\frac{1}{k!}[D^kF(\mathbf{x})]F : V \times \dots \times V \rightarrow W \quad (1.10)$$

denotes the symmetric k -linear mapping whose restriction to the diagonal is the terms of degree k of the expansion.

Lemma 1.1.2. *The derivative of the Newton map is given by the formula*

$$[DN_F(\mathbf{x})](\mathbf{v}) = [DF(\mathbf{x})]^{-1}[D^2F(\mathbf{x})]\left([DF(\mathbf{x})]^{-1}F(\mathbf{x}), \mathbf{v}\right). \quad (1.11)$$

This formula reduces to the simple one above in dimension 1 (Equation (1.8)).

Proof. If $M : X \rightarrow \text{Hom}(V, W)$ is a differentiable matrix-valued function, then

$$[DM^{-1}(\mathbf{x})](\mathbf{y}) = (M(\mathbf{x})^{-1})\left([DM(\mathbf{x})](\mathbf{y})\right)(M(\mathbf{x})^{-1}), \quad (1.12)$$

where the middle term is naturally a linear transformation in $\text{Hom}(V, W)$, so that the composition makes sense. Using this formula, we compute

$$[DN_F(\mathbf{x})](\mathbf{v}) = \mathbf{v} - [DF(\mathbf{x})]^{-1}[DF(\mathbf{x})](\mathbf{v}) - [DF(\mathbf{x})]^{-1}[D^2F(\mathbf{x})]\left([DF(\mathbf{x})]^{-1}(F(\mathbf{x})), \mathbf{v}\right), \quad (1.13)$$

and the first two terms on the left-hand side cancel, to give the desired result. \square

Of course, since roots of F are superattractive for Newton's method, the derivative vanishes there, as is clear from Lemma 1.1.2. It is often useful to know the quadratic terms also.

Lemma 1.1.3. *If $F(\mathbf{x}_0) = \mathbf{0}$ and $[DF(\mathbf{x}_0)]$ is invertible, then we have*

$$N_F(\mathbf{x}_0 + \mathbf{u}) = \mathbf{x}_0 + \left(\frac{1}{2}[DF(\mathbf{x}_0)]^{-1}[D^2F(\mathbf{x}_0)](\mathbf{u}, \mathbf{u})\right) + o(\|\mathbf{u}\|^2). \quad (1.14)$$

Proof. We just expand $N_F(\mathbf{x}_0 + \mathbf{u})$ into a Taylor polynomial of degree 2:

$$\begin{aligned} N_F(\mathbf{x}_0 + \mathbf{u}) &= \mathbf{x}_0 + \mathbf{u} - [DF(\mathbf{x}_0 + \mathbf{u})]^{-1}(F(\mathbf{x}_0 + \mathbf{u})) \\ &= \mathbf{x}_0 + \mathbf{u} - [DF(\mathbf{x}_0 + \mathbf{u})]^{-1}\left([DF(\mathbf{x}_0)]\mathbf{u} + \frac{1}{2}[D^2F(\mathbf{x}_0)](\mathbf{u}, \mathbf{u})\right) + o(\|\mathbf{u}\|^2) \\ &= (\mathbf{x}_0 + \mathbf{u}) - [Df(\mathbf{x}_0)]^{-1}[DF(\mathbf{x}_0)](\mathbf{u}) - \frac{1}{2}[DF(\mathbf{x}_0)]^{-1}[D^2F(\mathbf{x}_0)](\mathbf{u}, \mathbf{u}) \\ &\quad + [D^2F(\mathbf{x}_0)]\left(\mathbf{u}, [Df(\mathbf{x}_0)]^{-1}[DF(\mathbf{x}_0)](\mathbf{u})\right) + o(\|\mathbf{u}\|^2). \end{aligned} \quad (1.15)$$

In Equation (1.15), the constant term is \mathbf{x}_0 , the linear terms cancel, and the quadratic terms give

$$\left(-1 + \frac{1}{2}\right) \left([DF(\mathbf{x}_0)]^{-1}[D^2F(\mathbf{x}_0)](\mathbf{u}, \mathbf{u})\right), \quad (1.16)$$

which gives the desired result. \square

The critical locus

The critical points of Newton's map are of course those \mathbf{x} for which the derivative above is not invertible, i.e, those for which the mapping

$$\mathbf{v} \mapsto [D^2F(\mathbf{x})]([DF(\mathbf{x})]^{-1}F(\mathbf{x}), \mathbf{v}) \quad (1.17)$$

is not invertible. We can analyze this condition further.

Call a quadratic map $Q : V \rightarrow W$ *non-degenerate* if $Q^{-1}(0) = 0$. In that case Q induces a map

$$q : \mathbb{P}(V) \rightarrow \mathbb{P}(W) \quad (1.18)$$

whose critical locus C_q is a hypersurface in $\mathbb{P}(V)$ of degree $\dim W$. Its inverse image C_Q in V is the critical locus of Q . In the case $\dim V = \dim W$, which is the main one of interest here, C_q consists of two points, and C_Q consists of two lines.

The vectors in V in the directions of C_Q will be called the *critical directions of Q* .

Lemma 1.1.4. *If the point \mathbf{x} is not a critical point of F , then it is a critical point of the Newton map N_F if and only if the Newton vector $[DF(\mathbf{x})]^{-1}F(\mathbf{x})$ points in a critical direction of the quadratic function*

$$\mathbf{u} \mapsto [D^2F(\mathbf{x})](\mathbf{u}, \mathbf{u}). \quad (1.19)$$

Proof. This follows from Equation (1.14); the composition by $[DF(\mathbf{x})]^{-1}$ on the left does not affect which vectors are mapped to zero. \square

1.2. The intersection of conics

One particularly simple example of Newton's map is the one used to find the intersection of two quadratic curves; in fact this might be the simplest non-degenerate case, since the intersection of a line and a curve of any degree gives a singular Newton map.

Two such curves usually intersect in four points, and we will need to understand the space of quadratic functions vanishing at four distinct points.

Lemma 1.2.1. (a) *The space of quadratic functions vanishing at four distinct points is 2-dimensional, and GL_2 acts transitively (by composition on the left) on pairs of such functions that are linearly independent.*

(b) *Among these functions exactly two define parabolas (possibly degenerate), and exactly three define pairs of lines.*

From this and Lemma 1.1.1, we immediately see:

Corollary 1.2.2. *Newton's method to find the intersection of two conics depends only on the intersection points and not on the choice of curves.*

The family of curves given by a 2-dimensional vector space of functions is a 1-dimensional family of curves, called a *pencil* of curves. Lemma 1.2.1 says that the conics passing through four distinct points form a pencil. More generally, the curves of degree d passing through $(d^2 + 3d - 2)/2$ points form a pencil. When $d = 3$, we have $(d^2 + 3d - 2)/2 = 8$, and given any eight points in \mathbb{C}^2 in general position, we can consider the Newton map defined by any two cubics passing by these points. This map is independent of the choice of cubics in the pencil. A well-known theorem asserts that if three cubics are concurrent in eight points, then they are also concurrent in a ninth; the argument above gives another proof of this fact. We can also see that the Newton maps applied to find the intersections of two curves of degree d form a family of dimension $d^2 + 3d - 8$.

Proof of Lemma 1.2.1. By an affine change of variables in the domain, the four points can be moved, for instance to $\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$; so up to conjugation by an affine mapping the corresponding Newton maps depend on the two parameters α, β . We will find it convenient to set

$$A = \frac{1 - \alpha}{\beta} \quad \text{and} \quad B = \frac{1 - \beta}{\alpha} \quad (1.20)$$

as these quantities arise frequently in computation.

The general quadratic function vanishing at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ is

$$P(x^2 - x) + Qxy + R(y^2 - y), \quad P, Q, R \in \mathbb{C} \quad (1.21)$$

where $Q = PA + RB$. We can take as our equations any two linearly independent functions in this family.

One possibility is to require that $Px^2 + Qxy + Ry^2$ be a perfect square: this will happen when

$$Q^2 = (PA + RB)^2 = 4PR. \quad (1.22)$$

Up to multiples, there are two such equations in general, defining parabolas.

We could also choose $P = 1, R = 0$, and $P = 0, R = 1$, to find

$$F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 + Axy - x \\ y^2 + Bxy - y \end{pmatrix} \quad (1.23)$$

and finally the Newton map

$$\begin{aligned} N_F \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} x \\ y \end{pmatrix} - \begin{bmatrix} 2x + Ay - 1 & Ax \\ By & 2y + Bx - 1 \end{bmatrix}^{-1} \begin{pmatrix} x^2 + Axy - x \\ y^2 + Bxy - y \end{pmatrix} \\ &= \frac{1}{\Delta} \begin{pmatrix} x(Bx^2 + 2xy + Ay^2 - x - Ay) \\ y(Bx^2 + 2xy + Ay^2 - Bx - y) \end{pmatrix} \end{aligned} \quad (1.24)$$

where

$$\Delta = 2Bx^2 + 4xy + 2Ay^2 - (2 + B)x - (2 + A)y + 1. \quad \square \quad (1.25)$$

Most of this paper is devoted to understanding the dynamics of the Newton mapping given by Equation (1.24) in terms of the parameters α and β . Sometimes, instead of studying that mapping, we will switch to its conjugate (given by Equation (1.34)), which is sometimes simpler. This conjugate is obtained by intersecting parabolas.

The following result is essential to many of our constructions.

Lemma 1.2.3. *The lines joining the roots are invariant under Newton's method, and in these lines Newton's method induces the 1-dimensional Newton's method to find the roots of a quadratic polynomial.*

Proof. By Lemma 1.1.1, it is enough to prove the invariance of the x -axis in the above normalization. The Newton mapping becomes

$$N_F \begin{pmatrix} x \\ 0 \end{pmatrix} = \frac{1}{2Bx^2 - (2+B)x + 1} \begin{pmatrix} x(Bx^2 - x) \\ 0 \end{pmatrix} = \begin{pmatrix} x^2/(2x - 1) \\ 0 \end{pmatrix}, \quad (1.26)$$

The first coordinate is Newton's method for the polynomial $x^2 - x$ with roots at 0 and 1. \square

Lemma 1.2.3 explains why the lines joining pairs of roots in Figure 1 are colored in only two colors.

Remark. Lemma 1.2.3 is a special case of the following more general result: if two curves of degree d in the plane have d of their intersection points on a line, then this line is invariant under the Newton map, and the restriction to the line is the Newton map for solving the one-variable polynomial with its roots at those points.

Of course, if $d > 2$, it is exceptional for two curves of degree d to have d of their intersection points aligned. \triangle

We will refer to the lines joining pairs of roots as the *invariant lines* of the Newton map N_F . Their intersections also play an important role: they are points of indeterminacy of N_F . This should be clear: if the Newton map were defined there, the orbit would need to be in both lines, so the point of intersection would have to be fixed, and it isn't.

We will label these points

$$\mathbf{q}_1 = \begin{pmatrix} 1/B \\ 0 \end{pmatrix}, \quad \mathbf{q}_2 = \begin{pmatrix} 0 \\ 1/A \end{pmatrix}, \quad \mathbf{q}_3 = \begin{pmatrix} \alpha/(\alpha + \beta) \\ \beta/(\alpha + \beta) \end{pmatrix}. \quad (1.27)$$

In Figure 1, these points are the intersections of pairs of lines through disjoint pairs of roots. For instance, the point \mathbf{q}_1 is the point on the x -axis at $x = -1$.

As points of indeterminacy, these points are fairly innocent. A standard result from algebraic geometry says that N_F will be well defined after an appropriate sequence of blow-ups in the domain; in this case, it is enough to blow up once.

Lemma 1.2.4. *The Newton map extends algebraically to the exceptional divisors of \mathbb{C}^2 blown up at the points $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$, and the images of the exceptional divisors are straight lines.*

We will see later just which straight lines the images of the exceptional lines are.

Proof. It is enough to prove this at \mathbf{q}_1 . We need to show that the limit

$$m \mapsto \lim_{t \rightarrow 0} N_F \begin{pmatrix} 1/B + t \\ mt \end{pmatrix} \quad (1.28)$$

exists. This is a straightforward computation using Equation (1.24); the constant terms in t vanish, and after canceling a factor of t in the numerator and denominator of both coordinates, we find

$$\lim_{t \rightarrow 0} N_F \begin{pmatrix} 1/B + t \\ mt \end{pmatrix} = \begin{pmatrix} \frac{\frac{1}{B} + \frac{2m}{B^2} - \frac{Am}{B}}{2 - B + m(\frac{4}{B} - 2 - A)} \\ \frac{\frac{m}{B} - m}{2 - B + m(\frac{4}{B} - 2 - A)} \end{pmatrix}. \quad (1.29)$$

This is a parametrization of a line. \square

A degenerate case

The Newton map is special when two roots are on a line parallel to the line passing through the other two roots. In the form above, this occurs when $\alpha = 1$, or $\beta = 1$, or $\alpha = -\beta$. These are

equivalent under an affine change of variables, so we can restrict ourselves to the case $\alpha = 1$, i.e., $A = 0$. A bit of computation shows that in this case, the first coordinate of N_F is simply $x^2/(2x - 1)$, which is the one-variable Newton map for solving $x^2 - x = 0$. Thus all points with $\operatorname{Re} x < 1/2$ are attracted to the line $x = 0$, and all points with $\operatorname{Re} x > 1/2$ are attracted to the line $x = 1$. Within each of these lines we are again faced with a one-variable Newton map, so the dynamics are simple. Thus all the really interesting dynamics occurs in the real 3-dimensional manifold $\operatorname{Re} x = 1/2$.

In this 3-manifold, we have what might be called a *fibred family of rational functions*, i.e., a map of the form

$$G : \mathbb{R}/\mathbb{Z} \times \mathbb{P}^1 \rightarrow \mathbb{R}/\mathbb{Z} \times \mathbb{P}^1, \quad \begin{pmatrix} \theta \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2\theta \\ g_\theta(y) \end{pmatrix} \quad (1.30)$$

where g_θ is the rational function

$$y \mapsto \frac{y(yz^2 + Bz - y)}{(z + 1)((2y - 1)(z - 1) - Bz)}, \quad z = e^{2\pi i \theta}. \quad (1.31)$$

There has been some work on dynamical systems of this sort, mainly when g_θ is a polynomial [Se]. Computer investigation indicates that the system is quite interesting: we intend to write another paper on the subject.

In the case where three roots are aligned, there does not appear to be any way of making sense of the Newton map.

The parabolic normalization

We will often find a different normalization more convenient: by an affine change of coordinates in the domain we can impose that the axes of the two parabolas in a given pencil should be the coordinate axes, and the remaining freedom to scale the coordinates allows us to write these parabolas as

$$y = x^2 + a \quad \text{and} \quad x = y^2 + b. \quad (1.32)$$

This corresponds to setting

$$F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 - y + a \\ y^2 - x + b \end{pmatrix}. \quad (1.33)$$

In this form, the Newton mapping becomes

$$\begin{aligned} N_F \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} x \\ y \end{pmatrix} - \frac{1}{4xy - 1} \begin{bmatrix} 2y & 1 \\ 1 & 2x \end{bmatrix} \begin{pmatrix} x^2 - y + a \\ y^2 - x + b \end{pmatrix} \\ &= \frac{1}{4xy - 1} \begin{pmatrix} 2x^2y + y^2 - 2ay - b \\ 2xy^2 + x^2 - 2xb - a \end{pmatrix}. \end{aligned} \quad (1.34)$$

Actually, this normalization excludes the degenerate cases where one of the parabolas is a pair of parallel lines, which has features of its own. The locus in the (a, b) -plane where this occurs is easy to parametrize by the slope m at the point of tangency $\begin{pmatrix} x \\ y \end{pmatrix}$: the relations

$$2x = m, \quad 2my = 1 \quad \text{lead to} \quad m \mapsto \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y - x^2 \\ x - y^2 \end{pmatrix} = \begin{pmatrix} 1/(2m) - m^2/4 \\ m/2 - 1/(4m^2) \end{pmatrix}. \quad (1.35)$$

The equation of this curve is

$$2^8(a^2b^2 + a^3 + b^3) + 2^53^2ab = 3^3; \tag{1.36}$$

it is a quartic of genus 0, with three cusps; the real locus of this quartic is shown in Figure 7.

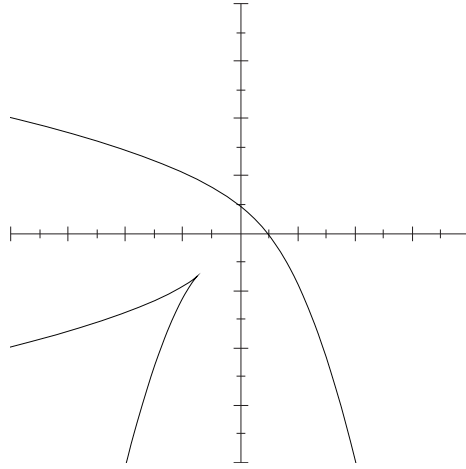


FIGURE 7. The locus of pairs (a, b) , $-4 \leq a, b \leq 4$ for which the parabolas $y = x^2 + a$ and $x = y^2 + b$ are tangent.

In one dimension, a multiple root is an attracting fixed point of the Newton map, but not superattractive. In two dimensions, the situation is a great deal more complicated. It is easier to study if we go back to a variant of the form (1.24), and consider the pencil of conics that pass through $\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and have slope m at the origin. The space of quadratic functions defining such conics is 2-dimensional, and we may take as basis the two functions

$$xy, \quad mx^2 - y^2 - mx + y. \tag{1.37}$$

The Newton map associated to the equations $xy = 0, \quad mx^2 - y^2 - mx + y = 0$ is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{2mx^2 - mx + 2y^2 - y} \begin{pmatrix} x(mx^2 + y^2 - y) \\ y(mx^2 + y^2 - mx) \end{pmatrix}. \tag{1.38}$$

In this form, it is fairly easy to see that the origin is a point of indeterminacy, in particular, it is not an attractive fixed point. But it almost is: in the x -axis and in the y -axis it is superattracting. In the line of tangency $y = mx$ it is linearly attracting: the Newton map restricts to $x \mapsto x/2$.

As a point of indeterminacy, the origin is a bit more complicated than the points examined so far: after one blow-up, the exceptional divisor is collapsed to the point corresponding to slope m , except for the point corresponding to slope $-m$, which must be blown up again to make the Newton map well defined.

Computer investigation seems to show that this case is not much simpler than the general case, as Figure 8 indicates.

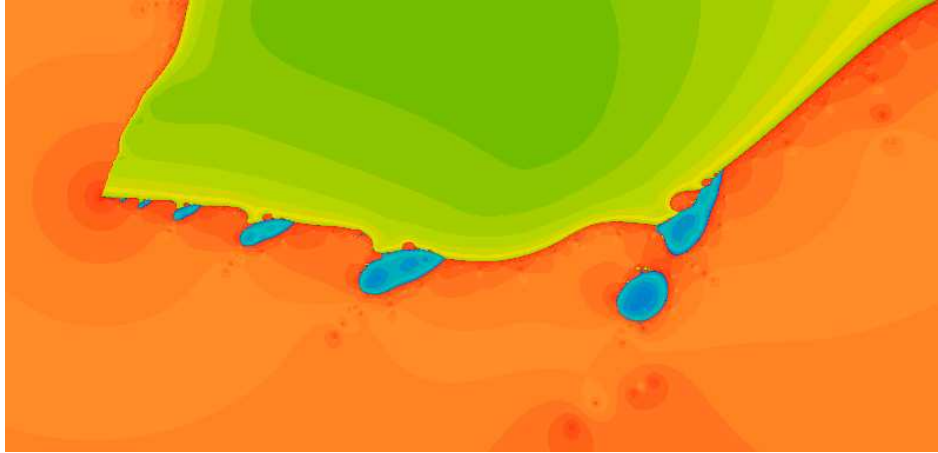


FIGURE 8. The picture represents a slice of \mathbb{C}^2 , for the dynamical system (1.38). In red is the basin of the double root, in green the basin of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, in blue the basin of $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Comparison of the two ways of writing the Newton map

Except for the degenerate cases, every map of the form (1.24) is conjugate to a map of the form (1.34), and vice-versa. Let us spell this out. Consider the space \mathcal{N} of conjugacy classes of Newton maps used to find the points of intersection of two conics that intersect in four distinct points, not on parallel lines. Then the (α, β) -plane, with the six lines $\alpha = 0, 1$, $\beta = 0, 1$, $\alpha + \beta = 0, 1$ removed, is a 24-fold cover of \mathcal{N} , with the fiber above a point corresponding to the 24 ways of labeling the roots.

The (a, b) -plane, with the quartic curve corresponding to tangency of the parabolas $y = x^2 + a$ and $x = y^2 + b$ removed, is a double cover of \mathcal{N} , with the fiber above a point corresponding to labeling the parabolas passing through the roots.

There is no covering mapping between these two covering spaces, and the fibered product over \mathcal{N} of the two covering spaces is a 48-fold cover, corresponding to labeling both the roots and the parabolas. Despite this, we will treat both forms as the “same family”, and label N_F the Newton maps associated to both.

The relation between the families does not respect reality: in the case where two real conics intersect in a real non-convex quadrilateral, the Newton map in the first form has real coefficients, but not in the second form: the change of variables bringing this quadrilateral to the second form is not real, since the parabolas in the pencil of conics passing by those points are not real.

1.3. The polar locus and the extension to infinity

Using the second form (1.34) for our conics, we have

$$\left[DF \begin{pmatrix} x \\ y \end{pmatrix} \right] = \begin{bmatrix} 2x & -1 \\ -1 & 2y \end{bmatrix}, \quad (1.39)$$

and we see immediately the following result.

Lemma 1.3.1. *For the mapping F of (1.33), the polar locus P_F is the hyperbola of equation $4xy = 1$.*

We simply cannot exclude P_F from our dynamical system. We would then have to exclude its inverse image, and the inverse image of that, etc., and we would lose all control over the domain of our dynamical system. In addition, we saw in the introduction to this paper, in the computer tour of Newton's method, that at least in the real plane, the polar locus and its inverse images form the most interesting part of all the dynamics.

There is an obvious cure: consider N_F as a mapping in the projective plane.

Lemma 1.3.2. *The map N_F extends to the line at infinity, except at the points at infinity on the axes of the parabolas (1.32). Moreover, the extension is the identity on the line at infinity (except at these points), and the eigenvalues of the derivative are 1 and 2. The eigenvalue 1 corresponds to the direction of the line at infinity.*

Proof. In homogeneous coordinates, the Newton map becomes

$$N_F : [x : y : z] \mapsto [2x^2y + y^2z - 2ayz - bz^3 : 2xy^2 + x^2z - 2bxz^2 - az^3 : 4xyz - z^3]. \quad (1.40)$$

In particular, if $x \neq 0, y \neq 0$, then

$$N_F : [x : y : 0] \mapsto [2x^2y : 2xy^2 : 0] = [x : y : 0]. \quad (1.41)$$

This shows that the map extends, and that the extension is the identity. We leave the computation of the eigenvalues to the reader. \square

The computation above shows that the two points at infinity on the axes of the parabolas (which are also the points where the parabolas are tangent to the line at infinity) are points of indeterminacy. We will label them $\mathbf{p}_1 = [1 : 0 : 0]$ and $\mathbf{p}_2 = [0 : 1 : 0]$, at least if we are writing Newton's method in the form (1.34); in the form (1.24), these points form a non-trivial double cover of the parameter plane and cannot be unambiguously labeled.

As was the case for the points \mathbf{q}_i , the points \mathbf{p}_i are apparently rather innocent points of indeterminacy, and N_F extends to the projective plane blown up at these points.

Lemma 1.3.3. *Let $E_{\mathbf{p}_i}$ be the exceptional divisor of \mathbb{P}^2 blown up at \mathbf{p}_i . Then N_F extends algebraically to $E_{\mathbf{p}_i}$, and maps $E_{\mathbf{p}_i}$ isomorphically to the line at infinity.*

Proof. We will work at \mathbf{p}_1 , and use the coordinates $(u = 1/x, y)$ in the domain, and $(u, v = y/x)$ in the range. Note that (u, y) are coordinates in the blown-up plane, whereas (u, v) are coordinates in the projective plane itself. Then $u = 0$ defines the exceptional divisor of the blow-up at \mathbf{p}_1 , and y parametrizes this divisor. In these coordinates, the Newton map is written

$$\begin{pmatrix} u \\ y \end{pmatrix} \mapsto \begin{pmatrix} \frac{u(4y - u)}{2y + y^2u^2 - 2ayu^2 - bu^2} \\ \frac{2y^2u + 1 - 2bu - au^2}{2y + y^2u^2 - 2ayu^2 - bu^2} \end{pmatrix}. \quad (1.42)$$

Clearly the exceptional divisor is mapped to the line at infinity, by the map

$$\begin{pmatrix} 0 \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ 1/(2y) \end{pmatrix}. \quad \square \quad (1.43)$$

The points $\mathbf{q}_i, \mathbf{p}_j$ give us five points of indeterminacy for N_F . Let us see that that is all.

Proposition 1.3.4. *The points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ are points of indeterminacy of N_F , and they are the only ones.*

Proof. This is a straightforward matter of writing the mapping in homogeneous coordinates and seeing where all three coordinate values vanish. It is easier in the form (1.24); we need to find points where

$$x(Bx^2 + 2xy + Ay^2 - xz - Ayz) = 0 \quad (1.44)$$

$$y(Bx^2 + 2xy + Ay^2 - Bxz - yz) = 0 \quad (1.45)$$

$$z(2Bx^2 + 4xy + 2Ay^2 - (2 + B)xz - (2 + A)yz + z^2) = 0. \quad (1.46)$$

To list the solutions to these equations, note that:

- The value $z = 0$ leads to $Bx^2 + 2xy + Ay^2$, giving the two points of indeterminacy at infinity $\mathbf{p}_1, \mathbf{p}_2$;
- The value $x = 0$ leads to $z = Ay$, the point \mathbf{q}_1 ;
- The value $y = 0$ leads to $z = Bx$, the point \mathbf{q}_2 ;
- The last point is found by setting $Bx^2 + 2xy + Ay^2 - xz - Ayz = Bx^2 + 2xy + Ay^2 - Bxz - yz = 0$, which leads to $x(1 - B) = y(1 - A)$, and the point \mathbf{q}_3 . \square

Observe the following rather remarkable coincidences:

Lemma 1.3.5. *The polar curve P_F passes by the following 11 distinguished points:*

- (a) *The two points $\{\mathbf{p}_1, \mathbf{p}_2\}$ at infinity on both parabolas;*
- (b) *The six centers of the line segments joining the roots;*
- (c) *The three intersections $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ of lines joining disjoint pairs of roots.*

Proof. Part (a) is immediate, but (b) and (c) are rather hard to show in this form. If you switch to the normalization (1.24), the polar locus is given by the equation

$$2Bx^2 + 4xy + 2Ay^2 - (2 + B)x - (2 + A)y + 1. \quad (1.47)$$

It is straightforward to show that the points $\begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1/B \\ 0 \end{pmatrix}$ satisfy this equation; this is enough. \square

1.4. The critical locus of N_F

In the form (1.34), the critical locus of the Newton map N_F can easily be computed from Lemma 1.1.4. Indeed, the quadratic terms of F are simply

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^2 \\ y^2 \end{pmatrix}, \quad (1.48)$$

so that the singular directions are at every point the directions of the axes, and by Lemma 1.1.4 the critical points of N_F are those points $\begin{pmatrix} x \\ y \end{pmatrix}$ at which the Newton vector

$$\left[DF \begin{pmatrix} x \\ y \end{pmatrix} \right]^{-1} F \begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{4xy - 1} \begin{bmatrix} 2y & 1 \\ 1 & 2x \end{bmatrix} \begin{pmatrix} x^2 - y + a \\ y^2 - x + b \end{pmatrix} \quad (1.49)$$

points in the direction of one of the axes. This gives the following result.

Proposition 1.4.1. *The critical locus of N_F is the union of the two cubic curves C_1 and C_2 of equation*

$$2xy^2 - x^2 + 2xb - y + a = 0 \quad \text{and} \quad 2x^2y - y^2 + 2ay - x + b = 0. \quad (1.50)$$

Remark. The fact that the critical locus is reducible is rather surprising, and is a direct consequence of the fact that $[D^2F]$ is constant. This will usually not happen when we intersect curves of higher degree, and correspondingly the critical locus will then usually not be reducible. \triangle

In dynamics, the critical values are often more important than the critical points; here they are especially simple.

Lemma 1.4.2. (a) *The images of the curves C_1 and C_2 are the parabolas Γ_1 and Γ_2 of equation respectively*

$$x = y^2 + b \quad \text{and} \quad y = x^2 + a. \quad (1.51)$$

(b) *The mappings $N_F : C_1 \rightarrow \Gamma_1$ and $N_F : C_2 \rightarrow \Gamma_2$ are given by*

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} y^2 + b \\ y \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ x^2 + a \end{pmatrix} \quad (1.52)$$

respectively.

In particular, we observe that the parabolas we chose to intersect are precisely the critical value locus.

Proof. This is a straightforward computation from Equation (1.34). \square

This now allows a fairly detailed description of C_1 and C_2 . They pass through the roots, and they are non-singular if the roots are simple, i.e., if the two parabolas we are intersecting are not tangent. They also go through the three points $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ of Lemma 2.4,(c), for the following reason. They must intersect each invariant line in three points (counted with multiplicity, of course), and two transverse intersections are at the roots. Where can the third point be? Not at infinity, since this is a fixed point of N_F and is not on Γ_1 or Γ_2 . Not at any point of the invariant line other than the roots where N_F is well defined. There is only the point of indeterminacy left.

We can use this information to describe the lines L_i , images of the exceptional divisors obtained by blowing up \mathbf{q}_i .

Lemma 1.4.3. *From \mathbf{q}_i , draw the parallels to the axes of Γ_1 and Γ_2 . These lines intersect Γ_1 and Γ_2 respectively at a single point; L_i is the line joining these points. This line is tangent to both parabolas.*

Proof. The exceptional divisor $E_{\mathbf{q}_i}$ intersects each critical cubic in a single point, so the image L_i of $E_{\mathbf{q}_i}$ is a line that intersects each critical value parabola in a single point. But that means that it is a line tangent to both Γ_1 and Γ_2 . There are exactly three of these double tangents. But we know that C_1 maps “vertically” to Γ_1 , and C_2 maps “horizontally” to Γ_2 ; the description of L_i follows. \square

The cubics C_1, C_2 intersect the line at infinity as follows: C_1 is tangent to the x -axis at ∞ and tangent to the line at ∞ on the y -axis, and C_2 is tangent to the y -axis at ∞ and tangent to the line at ∞ on the x -axis. Note that this accounts for the three points at infinity of both C_1 and C_2 , and also for the six intersections of each with P_F : there is one double intersection at infinity, one simple intersection at infinity, and three intersections at the points $\mathbf{q}_1, \mathbf{q}_2$ and \mathbf{q}_3 , which must therefore be transversal.

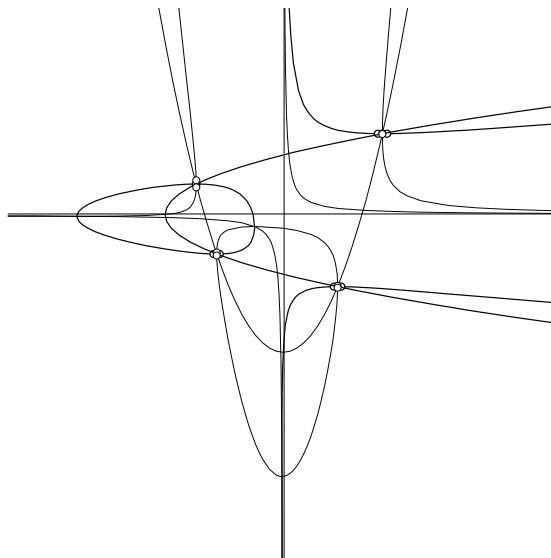


FIGURE 9. The two critical cubics and the critical value parabolas, when $a = -4$ and $b = -3$. We have also drawn in the polar curve of equation $4xy = 1$. The horizontal parabola and its inverse image are drawn dark; the vertical parabola and its inverse image are drawn light.

1.5. The topological degree of the Newton mapping

We will see in this section that the Newton map N_F is of topological degree 4. Written out, the equation $N_F(\mathbf{x}) = \mathbf{u}$ becomes

$$\begin{pmatrix} x \\ y \end{pmatrix} - \frac{1}{4xy - 1} \begin{bmatrix} 2y & 1 \\ 1 & 2x \end{bmatrix} \begin{bmatrix} x^2 - y + a \\ y^2 - x + b \end{bmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}. \quad (1.53)$$

The inverse image of any line is a cubic; more specifically, the inverse image of the line of equation $cx + dy = e$ is the cubic of equation

$$c(2x^2y + y^2 - 2ay - b) + d(2xy^2 + x^2 - 2bx - a) = e(4xy - 1). \quad (1.54)$$

We will need to understand these cubics. Most are non-singular, but in the 2-dimensional space of lines in \mathbb{P}^2 , there are curves, forming a set of codimension 1, where the cubics degenerate, and points, forming a locus of codimension 2, where the cubics degenerate further.

Proposition 1.5.1. *The inverse image of a line is a non-singular cubic in \mathbb{P}^2 , unless the line is tangent to one or both of the critical value parabolas, or passes through a root. The five exceptional cases are:*

- **Codimension 1 exceptional lines**

If a line is tangent to exactly one of the critical value parabolas, not at one of the roots, then its inverse image is reducible: it consists of a conic and the line through the point of tangency parallel to the axis of the parabola. These two curves intersect transversally on the critical cubic, at the two inverse images of the point of tangency.

If the line passes through a root, and it is not tangent to one of the critical value parabolas at the root, then its inverse is a cubic of genus 0, with a double point at the root.

• **Codimension 2 exceptional lines**

If a line is tangent to both critical value parabolas, its inverse image is the union of three lines. Two of these lines are parallel to the axes of the parabolas. These two intersect at one of the points of indeterminacy, and the third line joins the other two points of indeterminacy.

If a line is tangent to a critical parabola at a root, its inverse image is the union of a line and a conic, both passing through the root and tangent there. The line is parallel to the axis of the parabola.

If a line passes through two roots, its inverse image is reducible, consisting of itself (it is an invariant line), and a conic that intersects it transversally at the roots.

The various possibilities are illustrated in Figures 10, 11, and 12.

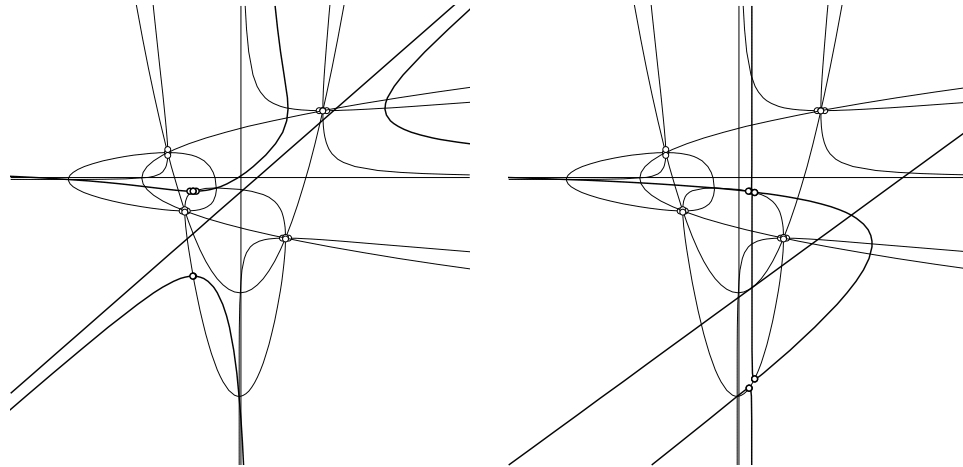


FIGURE 10. Left: a generic line and its inverse image, which is a non-singular cubic. Right: the inverse image of a line tangent to one of the critical value parabolas Γ_1 ; its inverse image is the union of a line parallel to the axis of the parabola (vertical in this case) through the point of tangency, and a conic, which intersects the line at the points where the line intersects the critical curve C_1 .

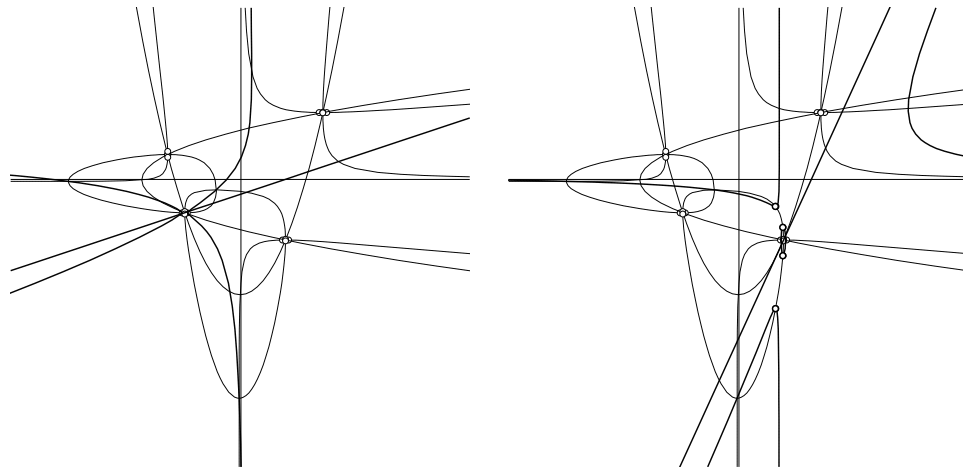


FIGURE 11. Left: a line through a root and its inverse image. Right: a line tangent to a critical value parabola at a root, and its inverse image.

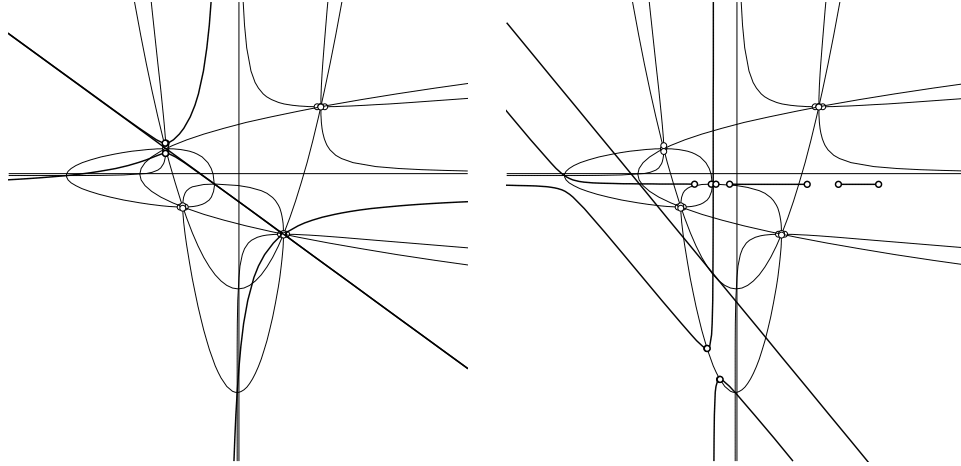


FIGURE 12. Left: a line through two roots and its inverse image; one component of its inverse image is itself. Right: a line tangent to both critical value parabolas, and its inverse image, which consists of three lines.

Proof. The inverse image of a line l can be singular only at critical points of N_F or at points of indeterminacy (which are transverse intersections of the critical curves C_1 and C_2). If l is transverse to a critical value parabola, then a straightforward computation shows that the inverse images of such a point are non-singular points of $N_F^{-1}(l)$, but are critical points of the restriction of N_F to $N_F^{-1}(l)$.

Each of the points \mathbf{q}_i “maps” to a line L_i by Lemma 1.2.4, and the inverse image of $l \cap L_i$ is at \mathbf{q}_i . Unless l is not one of the L_i , it intersects L_i at a single point transversally, and we have that $N_F^{-1}(l)$ passes through $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ and is smooth at these points. A similar argument shows that $N_F^{-1}(l)$ is smooth at infinity. This proves the first part.

If the line l is tangent to a critical value parabola, then its inverse image must have an ordinary double point at both of the inverse images in the critical cubic. But an irreducible cubic cannot have two double points, so the inverse image is necessarily a union of a line m_i and a conic. The conic is smooth, as otherwise it would degenerate to two lines, which would generate a third singular point. The line must join the two double points, so it is the line through the point of tangency parallel to the axis of the critical value parabola to which l is tangent, by Equation (1.52). The conic must then pass through all of the points $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$.

If a line l passes through a root and is not tangent to a critical value parabola there, a local computation shows that its inverse image has an ordinary double point there. If l does not pass through any other root, this is the only singular point on $N_F^{-1}(l)$, which is then a cubic of genus 0. This proves the “codimension 1” part.

A line tangent to both parabolas is one of the L_i . Passing to the limit in the argument about lines tangent to a critical value parabola, we see that the inverse image of a double tangent must contain two lines m'_i and m''_i parallel to the axes of the parabolas through the points of tangency; these intersect the point \mathbf{q}_i by Lemma 1.4.3. A cubic that contains two lines is the union of three lines, so $N_F^{-1}(L_i)$ must contain a third line m'''_i , which must go through the other two points of indeterminacy. In fact, m'''_i must intersect m'_i and m''_i on their respective critical cubics, so these intersections and the two points of indeterminacy must be aligned.

The case of a line tangent to a parabola at a root follows by continuity from the case where the line is tangent to a critical value parabola at a generic point.

The inverse image of an invariant line l is evidently itself and a conic, which must pass through the roots on l and be transverse to l there. \square

Remark. Proposition 1.5.1 applies to the cases where the roots are distinct, not on parallel lines. The degenerate Newton maps are slightly different; we will not describe the differences here. \triangle

We can now compute the degree of the Newton map.

Theorem 1.5.2. *The mapping N_F has topological degree 4.*

Proof. There are several ways of seeing this. One is to take a generic line l , and its inverse image $C_l = N_F^{-1}(l)$, which has genus 1. Then N_F induces a mapping $C \rightarrow l$, of some degree d , with critical points at the inverse images of $l \cap (\Gamma_1 \cup \Gamma_2)$, which consists of four points. Each point gives rise to two critical points, so there are eight of them, and they are all ordinary critical points. Now apply the Riemann-Hurwitz formula:

$$\chi(C_l) = 0 = d\chi(l) - (\text{number of critical points}) = 2d - 8, \quad \text{i.e., } d = 4. \quad \square \quad (1.55)$$

Remark. Another way to calculate the degree is to intersect the cubic curves $C_{1,u}$ and $C_{2,v}$ of equation

$$(4xy - 1)(x - u) - 2y(x^2 - y + a) - (y^2 - x + b) = 0 \quad (1.56)$$

and

$$(4xy - 1)(y - v) - 2x(y^2 - x + b) - (x^2 - y + a) = 0. \quad (1.57)$$

These curves must intersect at nine points, and we know that they intersect at $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{p}_1$, and \mathbf{p}_2 . It is not too hard to see that they are transverse at these points, and the other four points of intersection are the inverse image of $\begin{pmatrix} u \\ v \end{pmatrix}$. \triangle

Theorem 1.5.3. *For the Newton map N_F , the basin of each root is connected.*

Proof. The mapping is locally four-to-one near the roots. So we can choose a connected neighborhood U_0 of a root such that $U_1 = N_F^{-1}(U_0)$ is connected. Define $U_k = N_F^{-1}(U_{k-1})$; we must prove that all the U_k are connected. Suppose U_k is the first disconnected one, choose $\mathbf{x} \in U_k$, and choose a path γ in U_{k-1} connecting $N_F(\mathbf{x})$ to a point of U_0 . By a small perturbation, we may assume that γ does not intersect the critical value locus $\Gamma_1 \cup \Gamma_2$, or the three double tangents L_1, L_2, L_3 . Then the inverse image of γ consists of four arcs, all ending at points of U_1 . One arc must lead to \mathbf{x} in U_k . \square

1.6. The one-variable rational functions associated to the roots

At each root \mathbf{a} of F , i.e., at each fixed point of N_F , the quadratic terms of N_F induce a rational function $G_{\mathbf{a}}$ on the projective line $\mathbb{P}_{\mathbf{a}}$ associated to the tangent space to \mathbb{C}^2 at \mathbf{a} . This projective line can also be understood as the exceptional divisor one obtains if the root is blown up.

Our two normalizations of Newton's method give this rational function in two different forms. In the representation (1.24), it is easy to see the three fixed points of the rational function; they are the points of $\mathbb{P}_{\mathbf{a}}$ corresponding to the three invariant lines through \mathbf{a} . In particular, for the

root at the origin they are the points of $\mathbb{P}_{\mathbf{a}}$ corresponding to the axes, and to the line of equation $\alpha y = \beta x$.

In the representation (1.34), on the other hand, the points of $\mathbb{P}_{\mathbf{a}}$ corresponding to vertical and horizontal lines are the critical points of $g_{\mathbf{a}}$. Thus we will naturally find $g_{\mathbf{a}}$ in the form

$$z \mapsto \frac{z^2 + c}{z^2 + d}, \quad (1.58)$$

with critical points at 0 and ∞ .

Comparing these two normalizations, we see that in the normalization (1.34), the rational function $g_{\mathbf{a}}$ is never a polynomial, or conjugate to a polynomial. Indeed, polynomials have a fixed critical point, and the vertical or horizontal lines are never invariant lines of N_F in the form (1.34). In the form (1.24), polynomials do arise, precisely when the roots are on pairs of parallel lines.

Since we know where the roots are, it is easiest to deal with the form (1.24). The quadratic terms at the origin give the rational function

$$\frac{z^2 + Bz}{Az + 1} \quad (1.59)$$

with fixed points

$$\begin{aligned} 0, & \quad \text{with multiplier } B; \\ \infty, & \quad \text{with multiplier } A; \\ \frac{\beta}{\alpha}, & \quad \text{with multiplier } \alpha + \beta. \end{aligned} \quad (1.60)$$

Note that these multipliers do satisfy the Fatou relation [Mi2]

$$\sum \frac{1}{m_i - 1} = 1, \quad (1.61)$$

where the sum is over the fixed points, and the m_i are the multipliers of the fixed points. In the case where one of the multipliers is 1, a more elaborate relation is required, but a look at equations (1.60) shows that this case does not arise: it would require $\alpha = 0$, or $\beta = 0$, or $\alpha + \beta = 1$, i.e., that three roots be on a line, and we have seen that in that case, there is no Newton map, even degenerate.

From this we can easily read off the multipliers of the other roots. The affine map that permutes the roots as follows:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} \quad (1.62)$$

is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ -x - Ay + 1 \end{pmatrix} \quad (1.63)$$

and in particular $\begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \begin{pmatrix} 1/\beta \\ (\alpha + \beta - 1)/\beta \end{pmatrix}$.

Thus the multipliers of the fixed points previously at $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, now at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, are

$$A' = \frac{\beta - 1}{\alpha + \beta - 1}, \quad B' = 1 - \alpha, \quad \text{and} \quad \alpha' + \beta' = \frac{\alpha + \beta}{\beta}. \quad (1.64)$$

One can of course continue this way for the other roots.

Figure 13 illustrates the multipliers.

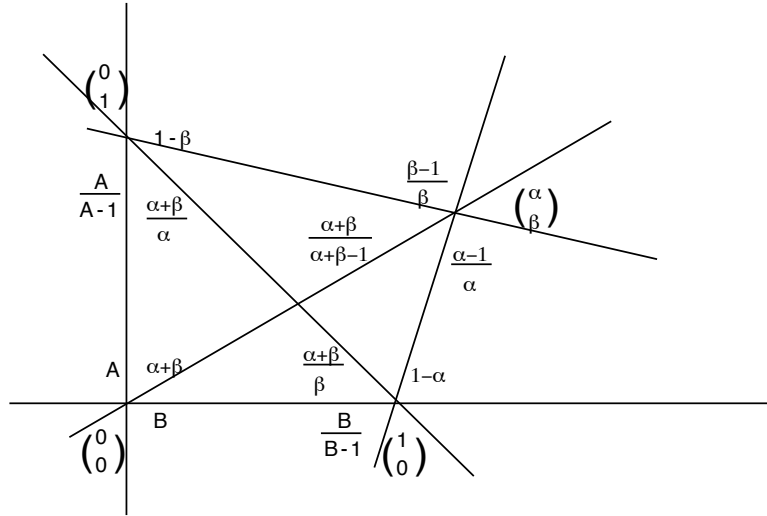


FIGURE 13. Each invariant line corresponds to a fixed point of a rational function at each root it passes through; this picture represents the corresponding multipliers.

There is a way to condense this information into a cleaner algebraic form (which still gives no relation between the dynamics of the various rational maps). Note that one way of parametrizing rational functions with a distinguished fixed point is by the multiplier m of that fixed point, and the product μ of the multipliers at the other two fixed points. We can then summarize the information in Figure 13 by the following statement.

Proposition 1.6.1. *Let L be an invariant line of N_F , and suppose m, μ and m', μ' are the coordinates of the two rational functions corresponding to roots on L , with m and m' being the multipliers of of the fixed points corresponding to L . Then*

$$m' = \frac{m}{m-1} \quad \text{and} \quad \mu = \mu'. \tag{1.65}$$

2

Invariant 3-manifolds associated to invariant circles

The most striking feature of computer drawings for Newton's method is the apparent existence of smooth 3-manifolds in the dynamical plane, which make up most of the apparent boundary of the basins. We will prove in Theorem 2.4.1 that under appropriate circumstances, there actually are invariant 3-manifolds in \mathbb{P}^2 , which do belong to the boundaries of the basins.

In computer pictures, these manifolds look remarkably smooth. The pictures are misleading; sections of these manifolds by complex lines are quasi-circles, or at least quasi-arcs, which are non-differentiable on a dense subset (though they may be differentiable almost everywhere).

These manifolds are some sort of "stable manifolds" of invariant circles. But their existence does not follow from any variant of the stable manifold theorem, as there are always parts of them that are not attracted to the circles. In the final analysis, we control the topological structure of these manifolds using a completely different tool, the λ -lemma of [MSS] concerning holomorphic motions.

We also prove, in Theorem 2.3.4, that every point at infinity except \mathbf{p}_1 and \mathbf{p}_2 has an unstable manifold. This is much less delicate than the previous result, and explains most of the self-similarity one sees in computer pictures: the points that appear to be centers of self-similarity are simply inverse images of points at infinity.

In Sections 2.1 and 2.2 we will write our Newton map in the form (1.24).

2.1. The circles in the invariant lines

In each invariant line, the bissector of the roots is invariant under N_F , and if we add the point at infinity, this bissector becomes a circle, in which N_F is angle-doubling. More specifically, consider the x -axis. The bissecting line is parametrized by

$$\theta \mapsto \frac{1}{2} + \frac{i}{2} \cot \theta, \quad \text{and} \quad N_F \left(\frac{1}{2} + \frac{i}{2} \cot \theta \right) = \frac{1}{2} + \frac{i}{2} \cot 2\theta. \quad (2.1)$$

Of course, N_F is expanding along that line; we will compute how it behaves in the normal direction. Since the x -direction is invariant, the derivative is triangular, and the derivative in the normal direction is given by the linear terms in y of the second component of N_F , i.e., by

$$\frac{Bxy - Bx^2y}{-2Bx^2 + Bx + 2x - 1} = \frac{Bx(1-x)}{(2x-1)(1-Bx)}y. \quad (2.2)$$

If we set $x = (1 + it)/2$, and recall that we want to compute the average with respect to the invariant measure

$$\frac{1}{\pi} |d\theta| = \frac{1}{\pi} \frac{|dt|}{(1+t^2)}, \quad (2.3)$$

we see that the logarithm of the average is given by

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \log \left| \frac{B((1+it)/2)((1-it)/2)}{it(1-B(1+it)/2)} \right| \frac{dt}{1+t^2}. \quad (2.4)$$

Theorem 2.1.1. *The Lyapunov exponents of the circle $Re x = 1/2, y = 0$ are $\lambda_1 = 2$, corresponding to the circle itself, and*

$$\lambda_2 = \begin{cases} |B| & \text{if the point of indeterminacy } 1/B \text{ is in the basin of } 1 \\ \left| \frac{B}{B-1} \right| & \text{if the point of indeterminacy } 1/B \text{ is in the basin of } 0. \end{cases} \quad (2.5)$$

The rational functions $\mathbf{g} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\mathbf{g} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ corresponding to the roots on the x -axis each have a fixed point corresponding to the x -axis itself, with multipliers respectively B and $B/(B-1)$. Moreover, one of these roots attracts the point of indeterminacy $\mathbf{q}_1 = \begin{pmatrix} 1/B \\ 0 \end{pmatrix}$, unless $Re(1/B) = 1/2$.

So we see that Theorem 2.1.1 can be restated without reference to the x -axis, so it applies to the invariant circles in any invariant line.

Theorem 2.1.1'. *The second Lyapunov exponent λ_2 of the invariant circle in an invariant line l is the absolute value of the multiplier of the fixed point corresponding to l of the rational function at the root in l that does not attract the point of indeterminacy in l .*

The second Lyapunov exponent λ_2 above always satisfies $|\lambda_2 - 1| \leq 1$, so we do have $0 \leq \lambda_2 \leq \lambda_1$.

If the point of indeterminacy is on the circle, then $|B| = |B/(1-B)|$, and in that case this number is λ_2 .

Proof. By the ergodic theorem, the second Lyapunov exponent exists almost everywhere, and is an invariant function on the circle. Angle doubling is an ergodic map of the circle to itself, hence this function is constant (almost everywhere), and equal to its space average. In other words, we have

$$\begin{aligned} \lambda_2 &= \exp \left(\frac{1}{\pi} \int_{-\infty}^{\infty} \log \left| \frac{B((1+it)/2)((1-it)/2)}{it(1-B(1+it)/2)} \right| \frac{dt}{1+t^2} \right) \\ &= \exp \left(\frac{1}{\pi} \int_{-\infty}^{\infty} \log \left| \frac{1+t^2}{2it(2-B(1+it))} \right| \frac{dt}{1+t^2} \right). \end{aligned} \quad (2.6)$$

Evaluating this integral is an entertaining exercise in the calculus of residues. Write it as the sum of

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \left(\log |B| - \log 2 + \log |t+i| + \log |t-i| - \log |t| \right) \frac{dt}{1+t^2} = \log |B| - \log 2 + \log 2 + \log 2 + 0 \quad (2.7)$$

and

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \log \left| \frac{1}{2-B(1+it)} \right| \frac{dt}{1+t^2} = \begin{cases} -\log |1-B| - \log 2 & \text{if } Re(1/B) < 1/2 \\ -\log 2 & \text{if } Re(1/B) > 1/2. \end{cases} \quad (2.8)$$

The only terms requiring care in the first integral are the ones involving $\log |t \pm i|$. To evaluate them we observe that the function $\log(z+i)$ has an analytic branch in the upper half-plane, so we can close the interval $[-R, R]$ with a half-circle in the upper half-plane, to form a closed curve Γ_R bounding a half-disc containing the pole of $1+z^2$ at i . Applying the residue formula, we find

$$\int_{\Gamma_R} \frac{\log(z+i)dz}{z^2+1} = 2\pi i \frac{\log 2}{2i} = \pi \log 2. \quad (2.9)$$

In the standard way, let $R \rightarrow \infty$, so that the contribution from the half-circle tends to 0.

The third integral is evaluated by residues also; we need to avoid the zero of $1 - B(1 + iz)/2$, which is in the upper half-plane or the lower half-plane depending on the sign of $\operatorname{Re}(B - 1/2)$, so we integrate around a semi-disc in the other half-plane.

In the final analysis, the average normal expansion on the circle is

$$\begin{cases} \left| \frac{B}{1-B} \right| & \text{if } \operatorname{Re}(1/B) < 1/2 \\ |B| & \text{if } \operatorname{Re}(1/B) > 1/2. \end{cases} \quad (2.10)$$

This makes sense: one of the two roots on that line is distinguished, the one that attracts the point of indeterminacy $1/B$. What we are finding depends on the non-distinguished point, and is the multiplier of the fixed point of the rational function associated to the invariant line itself.

Note that the multiplier is always a number in $[0, 2]$; the value 0 occurs when $B = 0$, i.e., $\beta = 1$, which means that the roots are on the two parallel lines $y = 0$ and $y = 1$. The value 2 is realized exactly if $\operatorname{Re}(1/B) = 1/2$, the case omitted above. This happens for instance for the case $\alpha = \beta = -1$, which also corresponds to solving the two equations

$$x^2 = y, \quad y^2 = x. \quad (2.11)$$

Pesin theory and stable manifolds

Pesin theory, together with Theorem 2.1.1, implies that if $\lambda_2 < 1$, then the invariant circle is an “ergodic attractor” [PS]. In particular, almost every point has a stable manifold, which is a Riemann surface in \mathbb{C}^2 made up of points attracted to the circle, and which is part of the separator of the basins. This gives some sort of a “measure theoretic” 3-dimensional real manifold in the separator for each circle with $|\lambda_2| < 1$. But we have no precise idea what the topological structure of such an object is. Later, we will show that under appropriate circumstances its closure is a genuine topological manifold.

2.2. Repelling cycles on invariant circles

The map $\phi : \mathbb{C} \rightarrow \mathbb{P}^2$ given by

$$\phi(u) \mapsto \left(\frac{u}{u_0 - 1} \right) \quad (2.12)$$

conjugates $u \mapsto u^2$ to Newton’s method. In particular, it maps the unit circle to the invariant circle in the x -axis. Set $\zeta_{m,l} = e^{2\pi il/(2^m - 1)}$, so that $\zeta_{m,l}^{2^m} = \zeta_{m,l}$. Suppose that $\zeta_{m,l}^{2^{m'}} \neq \zeta_{m,l}$ for all m' with $1 < m' < m$. Then the points

$$\phi(\zeta_{m,l}), \phi(\zeta_{m,l}^2), \dots, \phi(\zeta_{m,l}^{2^{m-1}}) \quad (2.13)$$

are precisely the cycles of length m on the invariant circle in the x -axis. The derivative of N_F at $\phi(\zeta)$ has eigenvalues

$$2 \quad \text{and} \quad \frac{B\zeta}{(\zeta + 1)(1 + B(\zeta - 1))}. \quad (2.14)$$

Set $\zeta = \zeta_{m,l}$. The second eigenvalue of the corresponding cycle is

$$M_{m,l} = \prod_{k=1}^{m-1} \frac{B\zeta^{2^k}}{(\zeta^{2^k} + 1)(1 + B(\zeta^{2^k} - 1))} = \frac{B^m}{\prod_{k=0}^{m-1} (1 + \zeta^{2^k}(B - 1))}. \quad (2.15)$$

We want to analyze for what m, l we have $|M_{m,l}| < 1$. The normal derivative at the point $x = z/(z - 1)$, $y = 0$ can be understood from Figure 14. The point $1 + z(B - 1)$ is the point z , rotated by $\arg z$ around 1. Thus the ratio $B/(1 + z(B - 1))$ is easy to visualize: it is < 1 outside the emphasized arc from B to \overline{B} , but > 1 inside the arc. The arc disappears when $B \in (0, 1)$, so in that case (and, as we will see, nearby ones) all cycles are saddles. But when B moves away from the real axis, some cycles become repelling.

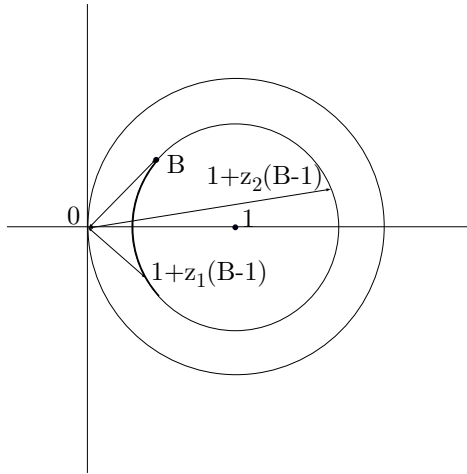


FIGURE 14. The multiplier along a cycle is the product of ratios $B/(1 + z(B - 1))$, i.e., ratios of the magnitude of two points on the circle of radius $|B - 1|$ centered at 1.

A first result is the following.

Proposition 2.2.1. *If B is real and $0 < B < 1$, we have $|M_{m,l}| < 1$ for all $m \geq 2$, and all l .*

Proof. A geometric picture of the product formula shows why this is obvious.

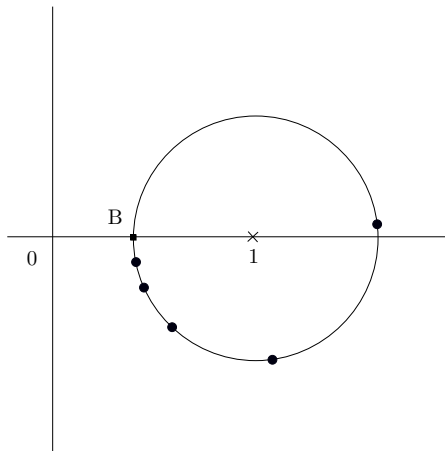


FIGURE 15. When $0 < B < 1$, the derivative along every cycle is a product of ratios, every one of which is < 1 .

The product we are considering is

$$\frac{B^m}{\prod_{k=0}^{m-1} (1 + \zeta^{2^k} (B - 1))} = \prod_{k=0}^{m-1} \frac{B}{1 + \zeta^{2^k} (B - 1)}. \quad (2.16)$$

The denominator is exactly B rotated by ζ^{2^k} around 1; in Figure 15 these points might look like the heavy dots on the circle centered at 1 and passing through B . In any case, all these points have absolute value greater than B , so each term in the product has absolute value < 1 . In particular, the product is < 1 . \square

Remark. There is one cycle to which the argument above does not apply: the fixed point at infinity. \triangle

The argument above can be refined to show that there is a neighborhood of the interval $(0, 1)$ in \mathbb{C} such that for B in that neighborhood, all the products have absolute value < 1 .

Theorem 2.2.2. *When B is in the region defined by the inequalities*

$$\begin{aligned} \left| \frac{B-1}{B} \arg(1-B) \right| < \frac{1}{3}, \quad \left| B - \frac{1}{2} \right| < \frac{1}{2}, \\ \left| \frac{B^2}{(2-B)(1-|1-B|)} \exp \left(4 \left| \operatorname{Im} \left(\frac{B-1}{B} \arg(1-B) \right) \right| \right) \right| < 1. \end{aligned} \quad (2.17)$$

all cycles on the invariant circle in the x -axis are saddles, except for the fixed point at infinity.

Remark. The second inequality is almost a consequence of the first: the relation of the two regions is represented in Figure 16. \triangle

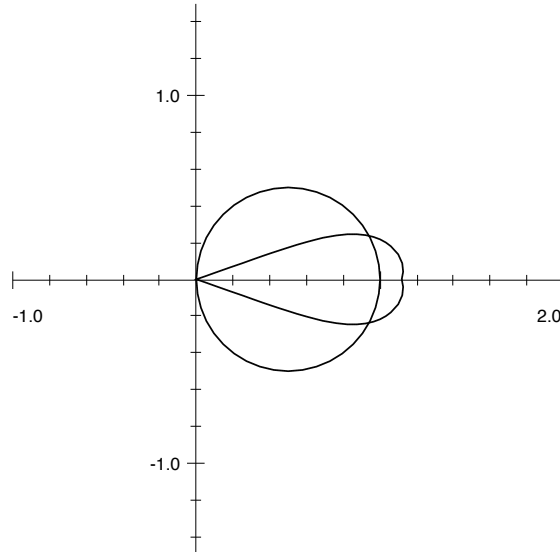


FIGURE 16. The interior of the wing-shaped curve corresponds to the region given by the first of the inequalities (2.17); it is almost contained in the region specified by the second inequality.

Proof. Choose B with $|B| < 1$, and $|1 - B| < 1$, and consider the circle C passing by B and centered at 1. Call C_0 the arc bounded by B and \overline{B} , contained in the unit disc centered at the

origin (see Figure 14). Call $z_k = 1 + \zeta^{2^k}(B - 1)$. Note that $z \in C_0$ if and only if $|B/z| > 1$; if $|B/z| < 1$, it is in the complementary arc. So the points $z_k \in C_0$ are the ones to worry about.

We will consider all the maximal subsets $z_{p-1}, z_p, \dots, z_{p+q}$ with $z_p, \dots, z_{p+q} \in C_0$, so that $z_{p-1} \notin C_0$.

What makes the theorem hard to prove is that there is no bound on q ; there can be arbitrarily long subchains of the cycle in C_0 . But it turns out that

$$\prod_{k=0}^q \left| \frac{B}{z_{p+k}} \right| \tag{2.18}$$

is bounded independently of q , because for any θ , the infinite product

$$\prod_{k=0}^{\infty} \left| \frac{B}{1 + e^{i\theta/2^k}(B - 1)} \right| \tag{2.19}$$

is convergent. We will want not just that it is convergent, we will want an explicit bound; Lemma 2.2.3 is a first step in that direction.

Lemma 2.2.3. *If $|c| < 1/3$ and $\operatorname{Re} c \leq 0$, then*

$$\prod_{k=0}^{\infty} \left| 1 + \frac{c}{2^k} \right| \geq e^{4\operatorname{Re} c}. \tag{2.20}$$

Proof of Lemma 2.2.3. Take logarithms, to find

$$\begin{aligned} \log \prod_{k=0}^{\infty} \left| 1 + \frac{c}{2^k} \right| &= \sum_{k=0}^{\infty} \log \left| 1 + \frac{c}{2^k} \right| \\ &\geq \sum_{k=0}^{\infty} \log \left(1 + \frac{\operatorname{Re} c}{2^k} \right) \geq \sum_{k=0}^{\infty} 2 \frac{\operatorname{Re} c}{2^k} = 4\operatorname{Re} c \end{aligned} \tag{2.21}$$

where we have used $\log(1 + x) \geq 2x$ when $-1/3 \leq x \leq 0$. Exponentiation gives the result. \square

The next step is to replace our product with one of the form studied in Lemma 2.2.3. In order to do this, let us take $u_0 = z_{p+q}, u_1 = z_{p+q-1}$ and more generally

$$u_k = 1 + (B - 1) \left(\frac{u_0 - 1}{B - 1} \right)^{1/2^k}, \tag{2.22}$$

where the power is the principal branch, well defined since B/u_0 is close to 1. Unlike the z 's, the u_k are defined (by Equation (2.22)) for all $k \geq 0$, and hence well adapted to taking infinite products.

Even the points u_k are a bit awkward, and we will replace them by better points w_k , whose definition is suggested by Figure 17.

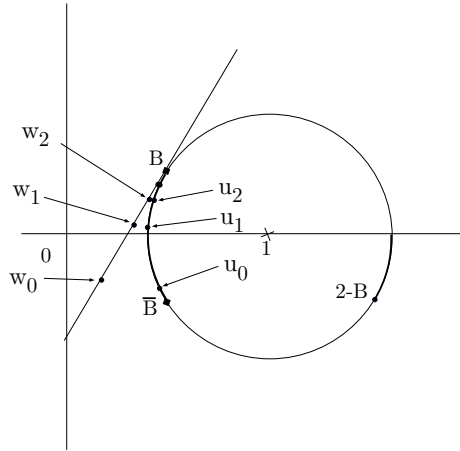


FIGURE 17. This illustrates how to find lower bounds for the multipliers. The w 's are closer to 0 than are the u 's.

Lemma 2.2.4. *If $|\arg B| < \pi/3$, and*

$$w_k = B + i(B - 1) \arg \frac{u_k - 1}{B - 1}, \tag{2.23}$$

then

$$\left| \frac{B}{u_k} \right| \leq \left| \frac{B}{w_k} \right|. \tag{2.24}$$

Proof. This is easy but fussy plane geometry. It is easiest to make the transformations $z \mapsto \frac{z-1}{B-1}$, to obtain Figure 18.

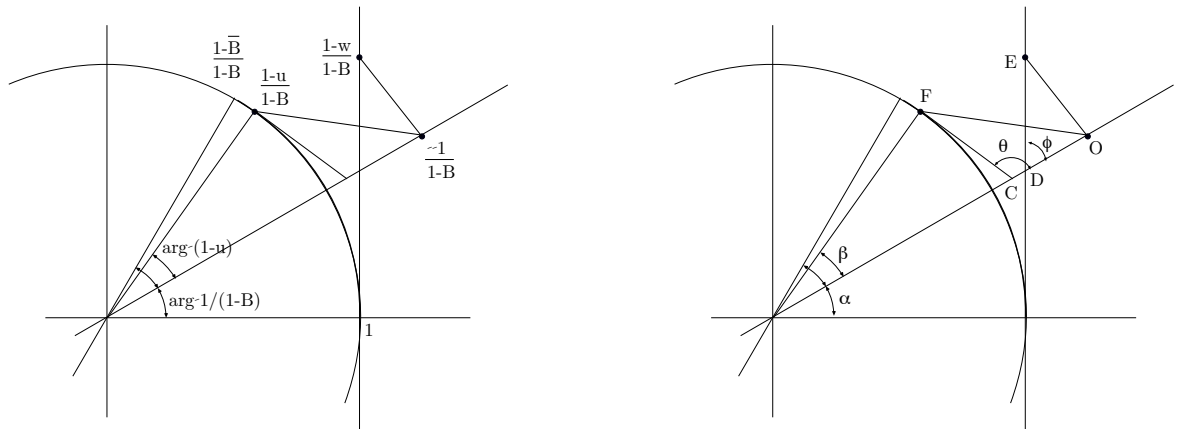


FIGURE 18. We have made the change of variables $z \mapsto (z - 1)/(B - 1)$. On the right, we have written arbitrary labels in standard “Euclidean geometry” style.

The first thing to observe is that the requirement $|B - 1/2| < 1/2$ translates into saying that the point labeled 0 is in the half-plane where the real part is > 1 . The requirement $|\arg B| < \pi/3$ guarantees that the point E in the figure is above the point D . There are actually two cases to

consider, $\arg(1 - u) > 0$ and $\arg(1 - u) < 0$. We will do only the first here, and leave the other, which is easier, to the reader. The object is to show $|Ow| < |Ou|$.

Consider the two triangles (ODE) and (OCF) . They satisfy the following properties:

$$|OD| < |OC|, \quad |ED| < |FC|, \quad \widehat{ODE} < \frac{\pi}{2}, \quad \widehat{OCF} > \frac{\pi}{2}. \quad (2.25)$$

The first and third are obvious, and the second follows from

$$|ED| = \alpha + \beta - \tan \alpha < \beta < \tan \beta = |FC|. \quad (2.26)$$

The result now follows from

$$|EO|^2 < |ED|^2 + |OD|^2 < |OC|^2 + |FC|^2 < |OF|^2, \quad (2.27)$$

where the first and last inequalities come from the law of cosines and the fact that θ is acute and ϕ is obtuse. \square

It follows that

$$\prod_{k=0}^{\infty} \left| \frac{B}{u_k} \right| \leq \prod_{k=0}^{\infty} \left| \frac{B}{w_k} \right|. \quad (2.28)$$

We can now bound the product:

$$\begin{aligned} \prod_{k=0}^{\infty} \left| \frac{B}{u_k} \right| &\leq \left| \frac{B}{u_0} \right| \prod_{k=1}^{\infty} \left| \frac{B}{w_k} \right| \\ &\leq \left| \frac{B}{u_0} \right| \prod_{k=1}^{\infty} \left| \frac{B}{B + i(B-1) \arg \frac{u_k-1}{B-1}} \right| \\ &= \left| \frac{B}{u_0} \right| \prod_{k=1}^{\infty} \left| \frac{1}{1 + i \frac{B-1}{B} \arg \frac{u_k-1}{B-1}} \right| \\ &= \left| \frac{B}{u_0} \right| \prod_{k=0}^{\infty} \left| \frac{1}{1 + i \frac{B-1}{2^k} \arg \frac{u_1-1}{B-1}} \right| \\ &\leq \left| \frac{B}{u_0} \right| \exp \left| 4 \operatorname{Im} \left(\frac{B-1}{B} \arg \frac{u_1-1}{B-1} \right) \right| \\ &\leq \frac{|B|}{1 - |1-B|} \exp \left| 4 \left(\operatorname{Im} \frac{B-1}{B} \arg(B-1) \right) \right|. \end{aligned} \quad (2.29)$$

In the process of writing this sequence of inequalities, we used Lemma 2.2.3 in the fifth line, and to go from the fifth to the sixth line, we used the fact that the product is monotone as a function of $\arg(u_1 - 1)$ with the maximum realized when $u_1 = 1 - |B - 1|$, whereas $|u_0|$ is bounded below by $1 - |B - 1|$. If we had started the product with u_0 , we would not have known that the product was monotone, which is why we split off the zeroth term.

Now we will include the point z_{p-1} of the cycle. It is in the arc from $1 + |1 - B|$ to $2 - B$, as shown in Figure 17. So we have

$$\left| \frac{B}{z_{p-1}} \right| \leq \frac{|B|}{1 + |1 - B|}, \quad (2.30)$$

so the whole block $z_{p-1}, z_p, \dots, z_{p+q}$ of the cycle contributes at most

$$\frac{|B|}{1 + |1 - B|} \exp \left| 4 \operatorname{Im} \left(\frac{B-1}{B} \arg(1-B) \right) \right| \quad (2.31)$$

to the multiplier of the cycle. Thus there is a neighborhood of the real axis where every cycle has one eigenvalue < 1 (except the fixed point at ∞); more specifically, this neighborhood contains the locus

$$\begin{aligned} \left| \frac{B-1}{B} \arg(1-B) \right| < \frac{1}{3}, \quad \left| z - \frac{1}{2} \right| < \frac{1}{2}, \\ \left| \frac{B^2}{(2-B)(1-|1-B|)} \exp \left(4 \left| \operatorname{Im} \left(\frac{B-1}{B} \arg(1-B) \right) \right| \right) \right| < 1. \quad \square \end{aligned} \quad (2.32)$$

Remark. There are definitely values of B where $|B| < 1$ and $|1 - B| < 1$, so that the invariant circle in the x -axis is on average attracting, but for which this circle contains infinitely many repelling cycles. We will see examples of such things in Figures 29 and 30. \triangle

2.3. Unstable manifolds at infinity

In this section we will write our Newton's method in the form (1.34).

We have seen that with the exception of the points $\mathbf{p}_1, \mathbf{p}_2$ on the axes of the critical value parabolas, the line at infinity is pointwise fixed, and the derivative at each point has eigenvalues 1 and 2, with the line of infinity the eigendirection for the eigenvalue 1, of course.

This suggests that each point should have an unstable manifold tangent to the other eigendirection. This is true, and we will prove somewhat more.

Theorem 2.3.1. *There exists a neighborhood U of $l_\infty - \{\mathbf{p}_1, \mathbf{p}_2\}$ in \mathbb{P}^2 , a neighborhood V of the y -axis with 0 removed in \mathbb{C}^2 , and an analytic isomorphism $\Phi = (\phi_1, \phi_2) : U \rightarrow V$ such that*

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \Phi = \Phi \circ N_F \quad (2.33)$$

on $U \cap N_F^{-1}(U)$.

Any Ψ satisfying the same property on a neighborhood of any open subset of $l_\infty - \{\mathbf{p}_1, \mathbf{p}_2\}$ is of the form $\Psi = (\psi_1, \psi_2)$, where

$$\psi_2 = \alpha(\phi_2), \text{ and } \psi_1 = \beta(\phi_2)\phi_1 \quad (2.34)$$

for appropriate analytic functions α, β , where β does not vanish.

In particular, we see that the curves $\phi_1 = \text{constant}$ are mapped to themselves by Newton's method, providing a dynamically natural foliation of a neighborhood of $l_\infty - \{\mathbf{p}_1, \mathbf{p}_2\}$ by invariant Riemann surfaces. How these can be extended is an obviously interesting problem, which will be examined in Section 2.5.

Proof. The first step is to find an appropriate neighborhood of $l_\infty - \{\mathbf{p}_1, \mathbf{p}_2\}$. Let

$$U_C = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \left| \sqrt{\frac{|x|}{C}} < |y| < C|x|^2 \right. \right\} \cup (l_\infty - \{\mathbf{p}_1, \mathbf{p}_2\}), \quad (2.35)$$

as shown in Figure 19.

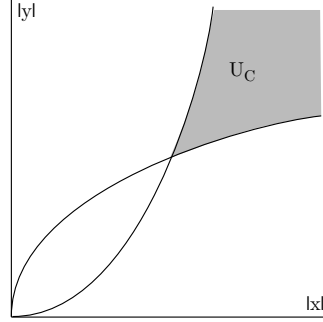


FIGURE 19. The region U_C is a neighborhood of the line at infinity with the points $\mathbf{p}_1, \mathbf{p}_2$ removed.

It will become clear in Chapter 3 that such a domain is optimal.

Lemma 2.3.2. *For $C > 0$ sufficiently small, the mapping*

$$N_F : U_C \cap N^{-1}(U_C) \rightarrow U_C \tag{2.36}$$

is an analytic isomorphism.

Proof. We will show that the restriction of N_F to $U_C \cap N^{-1}(U_C)$ is proper and a local homeomorphism. It is then a finite sheeted covering map, of some degree. But the points at infinity have only themselves as inverse images in U_C , so the degree is 1, and the map is a homeomorphism.

Clearly when C is sufficiently small U_C will not intersect either of the critical value parabolas or contain any of $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$; the statement about N_F being a local homeomorphism follows.

It remains to show that the restriction of N_F is proper. Let us denote $\tilde{\mathbb{P}}_I^2$ the projective plane blown up at the five points of indeterminacy, and $\tilde{N}_F : \tilde{\mathbb{P}}_I^2 \rightarrow \mathbb{P}^2$ the lifted Newton map, which is now well defined by Lemmas 1.2.4 and 1.3.3, and is evidently proper. Let \mathbf{x}_n be a sequence in U_C converging to $\mathbf{a} \in U_C$, and \mathbf{x}'_n be a sequence of preimages in $U_C \cap N_F^{-1}(U_C)$; we need to show that we can extract a subsequence of the \mathbf{x}'_n that converges in $U_C \cap N_F^{-1}(U_C)$. Since \tilde{N}_F is proper, we may assume that \mathbf{x}'_n converges to $\mathbf{a}' \in \tilde{\mathbb{P}}_I^2$; of course we have $\tilde{N}_F(\mathbf{a}') = \mathbf{a}$.

Let us see where \mathbf{a}' can be. One possibility is that it might be at infinity. In that case, it is fixed, contradicting $\mathbf{a} \in U_C$ and $\tilde{N}_F(\mathbf{a}') = \mathbf{a}$. This is true even if $\mathbf{a}' \in E_{\mathbf{p}_1} \cup E_{\mathbf{p}_2}$, as the sequence \mathbf{x}'_n would approach \mathbf{p}_1 or \mathbf{p}_2 tangentially to the line at infinity, and the corresponding points of the exceptional divisors are fixed.

Another possibility is that $\mathbf{a}' \in \partial U_C$. That is excluded by Lemma 2.3.3; points in the boundary of U_C are not mapped to points in the closure of U_C .

Lemma 2.3.3. *For C sufficiently small, we have*

$$\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_1 \right| < \frac{3}{4}|x| \quad \text{and} \quad \left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_2 \right| < \frac{3}{4}|y| \tag{2.37}$$

when $(x, y) \in U_C \cap \mathbb{C}^2$, and the indices 1 and 2 indicate the first and second coordinates.

Proof. This is a straightforward matter of seeing what terms dominate in the formula for N_F . Recall that

$$N_F \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{4xy - 1} \begin{pmatrix} 2x^2y + y^2 - 2ay - b \\ 2xy^2 + x^2 - 2xb - a \end{pmatrix} \tag{2.38}$$

and we have

$$\left| \frac{2x^2y + y^2 - 2ay - b}{4xy - 1} - \frac{x}{2} \right| \in o(|x| + |y|), \quad \left| \frac{2xy^2 + x^2 - 2ax - b}{4xy - 1} - \frac{y}{2} \right| \in o(|x| + |y|). \quad \square \quad (2.39)$$

A third thing that could happen is that $\mathbf{a}' \in U_C \cap \partial N_F^{-1}(U_C)$. But at such a point N_F is a local homeomorphism, mapping boundary to boundary, so that is excluded also. Thus $\mathbf{a}' \in U_C \cap N_F^{-1}(U_C)$. \square Lemma 2.3.2

Next, we give a general “unstable manifold theorem” under a relaxation of the present hypotheses, where we assume that one eigenvalue is 1, and there is a corresponding analytic “center manifold”.

Consider an open subset $U \subset \mathbb{C}$, a neighborhood $V \subset \mathbb{C}$ of 0, and a mapping $U \times V \rightarrow \mathbb{C}^2$ of the form

$$F : \begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} u + vf(u, v) \\ v(\lambda(u) + vg(u, v)) \end{pmatrix}, \quad (2.40)$$

where f, g are analytic functions on $U \times V$, and $|\lambda(u)| < 1$ on U .

Proposition 2.3.4. *There exist local coordinates (\tilde{u}, \tilde{v}) on a neighborhood of $U \times \{0\}$, with $u = \tilde{u}$ on $U \times \{0\}$, such that in these coordinates the mapping F is given by*

$$F : \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} \mapsto \begin{pmatrix} \tilde{u} \\ \lambda(\tilde{u})\tilde{v} \end{pmatrix}. \quad (2.41)$$

Proof. This is a matter of putting parameters into the standard proof of the linearization theorem for saddles with 1-dimensional stable manifolds, or more precisely, the existence of a parametrization of the stable manifold that linearizes the mapping in the stable manifold.

Choose $R_0 > 0$, and M such that

$$|f(u, v)| \leq M \quad , \quad |g(u, v)| < M \quad \text{when} \quad |u|, |v| \leq R_0. \quad (2.42)$$

Set

$$\Lambda = \sup_{|u - u_0| \leq R_0} |\lambda(u)| \quad \text{and choose } A \text{ so that } \Lambda < A < 1. \quad (2.43)$$

Finally choose $R \leq R_0$ so small that

$$\frac{RM}{1 - A} \leq R_0 \quad \text{and} \quad \Lambda + RM \leq A. \quad (2.44)$$

We claim that if $|v_0| \leq R$, then the sequence defined inductively by

$$(u_{n+1}, v_{n+1}) = F(u_n, v_n) \quad (2.45)$$

satisfies the inequalities

$$|u_n - u_0| \leq RM \frac{1 - A^n}{1 - A} \quad \text{and} \quad |v_n| \leq A^n |v_0|. \quad (2.46)$$

This is evidently true for $n = 0$, and by induction

$$\begin{aligned} |u_{n+1} - u_0| &\leq |u_{n+1} - u_n| + |u_n - u_0| \leq |v_n| |f(u_n, v_n)| + RM \frac{1 - A^n}{1 - A} \\ &\leq |v_0| M A^n + RM(1 + A + \cdots + A^{n-1}) \leq RM \frac{1 - A^{n+1}}{1 - A} \end{aligned} \quad (2.47)$$

and

$$|v_{n+1}| = |v_n| |\lambda(u_n) + v_n g(u_n, v_n)| \leq A^n |v_0| (\Lambda + MR) \leq A^{n+1} |v_0| \quad (2.48)$$

It follows immediately from these inequalities that the sequence u_n converges; we can now define

$$\tilde{u}(u_0, v_0) = \lim_{n \rightarrow \infty} u_n. \quad (2.49)$$

Clearly \tilde{u} is an analytic function on some neighborhood of $U \times \{0\}$, and $\tilde{u}(u, 0) = u$ and $\tilde{u}(F(u, v)) = \tilde{u}(u, v)$.

Now define \tilde{v} by the formula

$$\tilde{v}(u_0, v_0) = \lim_{n \rightarrow \infty} \frac{v_n}{(\lambda(\tilde{u}(u_0, v_0)))^n}. \quad (2.50)$$

We will show that when v_0 is small enough, this is a Cauchy sequence. To simplify notation, we will write $\lambda(\tilde{u}(u_0, v_0)) = \tilde{\lambda}$. We find

$$\left| \frac{v_{n+1}}{\tilde{\lambda}^{n+1}} - \frac{v_n}{\tilde{\lambda}^n} \right| = \frac{|v_n|}{|\tilde{\lambda}|^n} \left| \frac{\lambda(u_n) + v_n g(u_n, v_n)}{\tilde{\lambda}} - 1 \right|. \quad (2.51)$$

Now to bound these quantities:

$$|u_n - \tilde{u}| \leq C_1 A^n \quad \text{gives} \quad |\lambda(u_n) - \tilde{\lambda}| \leq C_2 A^n, \quad (2.52)$$

which together with $|v_n g(u_n, v_n)| \leq MRA^n$ gives

$$\left| \frac{v_{n+1}}{\tilde{\lambda}^{n+1}} - \frac{v_n}{\tilde{\lambda}^n} \right| \leq C_3 \left(\frac{A^2}{|\tilde{\lambda}|} \right)^n \quad (2.53)$$

for some constants C_1, C_2, C_3 which could be computed from R, M , and the error in the linear approximation to λ around \tilde{u} .

Unfortunately, $A^2/|\tilde{\lambda}|$ is not smaller than 1 unless we do something to make this happen. If we make R very small, then $\lambda(u_n)$ is very nearly constant, and we can choose A so that $A^2 < |\lambda(u_n)| < A$. With R this small, the series defining \tilde{v} converges. \square

To prove Theorem 2.3.1, apply Proposition 2.3.4 to the mapping $N_F^{-1} : U_C \rightarrow U_C$, which is guaranteed to exist by Lemma 2.3.2. \square

2.4. The invariant manifolds of circles

Throughout this section, we will write our Newton maps in the form (1.24), and we will focus on the invariant line l that is the x -axis, mapped to itself by $x \mapsto x^2/(2x - 1)$. This line contains the invariant circle S_l of equation $\operatorname{Re} x = 1/2$ (and the point at infinity), and if we set $\phi(z) = \begin{pmatrix} z/(z-1) \\ 0 \end{pmatrix}$ we find that S_l corresponds to the unit circle, mapped to itself by $z \mapsto z^2$, since

$$\phi^{-1} \circ N_F \circ \phi(z) = z^2. \quad (2.54)$$

In Section 2.1, we computed the average expansion of the invariant circle S_l , and found it to be either $|B|$ or $|B/(B - 1)|$ depending on which root attracts the point of indeterminacy $(1/B, 0)$. In this section, we will assume two things:

- that $|1 - B| < 1$, which as we saw implies that the root $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ attracts the point of indeterminacy, and that the multiplier of the invariant circle is $|B|$;
- that $|B| < 1$, so that S_l is “attracting.”

In Section 2.2, we found conditions under which all the cycles in S_l are saddles, repelling in the direction of l , of course, and attracting in the “normal” direction.

In this section, we will examine whether S_l has a stable manifold. This, if true, is much better than just knowing that the cycles are individually saddles; it is saying that the stable manifolds of these cycles fit together so that their closure is a 3-dimensional manifold.

The problem with this program is that it obviously does not work. At $l \cap l_\infty$, we can clearly see the splitting of the tangent space, into the direct sum of the tangent space to l (the direction of the base), and the tangent space to l_∞ (the direction of the fiber). These are the only invariant subspaces. But the Newton map is the identity on l_∞ , (except at the two points of indeterminacy, but $l \cap l_\infty$ is not a point of indeterminacy). Thus no point of l_∞ can be attracted to S_l . Neither can any point of the polar curve, or any point of its inverse images, etc. The inverse images of $l \cap l_\infty$ are dense in S_l , so there is a dense set of points that do not have a stable manifold in any classical sense of the word.

Despite this, we will show that a sort of “center-stable” manifold of S_l exists under appropriate conditions. It does not have the smoothness one would expect if S_l were hyperbolic, and the techniques used to show that it exists at all are quite different from the standard tools: one key ingredient is the theory of holomorphic motions.

Theorem 2.4.1. *Suppose the point (α, β) is chosen so that $B = (1 - \beta)/\alpha$ is real and satisfies $0 < B < 1$. Then there exists a neighborhood $\Omega \subset \mathbb{C}^2$ of (α, β) such that if $(\alpha', \beta') \in \Omega$, then there exists a 3-dimensional topological manifold $V_l \subset \mathbb{P}^2$, together with an inclusion $S_l \subset V_l$ and a projection $p : V_l \rightarrow S_l$ whose fibers are Riemann surfaces homeomorphic to discs, such that*

$$p|_{S_l} = id \quad \text{and} \quad N_{\alpha', \beta'}(V_l) \subset V_l, \quad (2.55)$$

and such that

$$p \circ N_{\alpha', \beta'} = N_{\alpha', \beta'} \circ p \quad (2.56)$$

on V_l .

A section of V_l by a complex line sufficiently close to l is a topological circle. Our proof will do better; it will show that such a section is a quasi-circle. This is probably optimal. Although computer pictures show 3-dimensional manifolds separating basins that appear to be smooth, this is probably false; these apparently smooth manifolds are probably not in fact C^1 on any open subset. Following the proof, we will give examples to demonstrate both why they look smooth, and why they are not.

Proof. The proof of Theorem 2.4.1 is long (more than four pages) and requires a number of intermediate results. The beginning of the proof looks rather like the graph transform argument, except that although a graph transform map is well defined, it is not contracting.

Without loss of generality, we will assume throughout that the invariant line we are considering is the x -axis. The first thing to do is to write the derivative of N along l .

Lemma 2.4.2. *At any point $x \in l$, the derivative of N is*

$$\begin{bmatrix} \frac{2x(x-1)}{(2x-1)^2} & \frac{Ax(1-x)}{(Bx-1)(1-2x)^2} \\ 0 & \frac{Bx(1-x)}{(Bx-1)(1-2x)} \end{bmatrix} = \begin{bmatrix} \alpha(x) & \beta(x) \\ 0 & \gamma(x) \end{bmatrix}. \quad (2.57)$$

The proof is left to the reader.

The coordinate x is not a good coordinate to use on S_l , since the fixed point on S_l corresponds to $x = \infty$. And actually, y is not a good coordinate for vertical tangent vectors either, since a finite vector tangent to l_∞ at $l \cap l_\infty$ is infinite in that coordinate. We will instead work in the coordinates

$$z = \frac{x}{x-1}, \quad v = \frac{y}{x}. \quad (2.58)$$

Lemma 2.4.3. *In the variables (z, v) , the derivative of N becomes*

$$\begin{bmatrix} 2z & \frac{Az^2(z-1)^2}{1+z(B-1)} \\ 0 & \frac{B}{1+z(B-1)} \end{bmatrix}. \quad (2.59)$$

Again, the proof is left to the reader.

The key property of this matrix is that if $0 < B < 1$, then S_l is on average attracting, and, more to the point at the moment, the entry $\frac{B}{1+z(B-1)}$ satisfies

$$\sup_{|z|=1} \left| \frac{B}{1+z(B-1)} \right| = 1 < 2. \quad (2.60)$$

The first step in constructing the graph transform map is to find an invariant family of cones.

Lemma 2.4.4. *There exist a neighborhood \tilde{W} of S_l and a (small) constant Θ such that if $C \begin{pmatrix} z \\ v \end{pmatrix}$ is the subset*

$$\left\{ \begin{bmatrix} \xi \\ \eta \end{bmatrix} \in T \begin{pmatrix} z \\ v \end{pmatrix} \mathbb{P}^2 \mid |\eta| > \Theta |\xi| \right\} \quad (2.61)$$

then

$$\left(DN \begin{pmatrix} z \\ v \end{pmatrix} \right)^{-1} \left(C_N \begin{pmatrix} z \\ v \end{pmatrix} \right) \subset C \begin{pmatrix} z \\ v \end{pmatrix} \quad (2.62)$$

for all $\begin{pmatrix} z \\ v \end{pmatrix}$ in \tilde{W} .

Proof. First, let us see this when $|z| = 1$ and $v = 0$. There it follows from the fact that the second diagonal term of the matrix (2.59) is smaller than the first, and the off-diagonal term is bounded. We can choose a single Θ that works for all points, since S_l is compact. By continuity, the result will hold in a neighborhood of S_l . \square

We will now choose a compact neighborhood $W \subset \tilde{W}$ of S_l adapted to $N_{\alpha,\beta}$, and depending on parameters $\epsilon_1, \epsilon_2, \delta$. Set $W = W_1 \cup W_2$, where

$$\begin{aligned} W_1 &= \left\{ \begin{pmatrix} z \\ v \end{pmatrix} \mid |\log |z|| \leq \delta, |v| \leq \epsilon_1, |\arg z| \geq \delta \right\} \\ W_2 &= \text{the union of the unstable manifolds of points in } l_\infty \text{ where } |v| \leq \epsilon_2, \\ &\quad \text{intersected with the region } |\arg z| \leq \delta, |\log |z|| \leq \delta. \end{aligned} \quad (2.63)$$

We will consider the variable z as “horizontal” and the variable v as “vertical,” and correspondingly we will set

$$\begin{aligned}
 \partial_H W_1 &= \text{the part of the boundary where } |v| = \epsilon_1 \\
 \partial_V W_1 &= \text{the part of the boundary where } |z| = e^{\pm\delta} \\
 \partial_H W_2 &= \text{unstable manifolds of points where } |v| = \epsilon_2 \\
 \partial_V W_2 &= \text{the part of the boundary where } |z| = e^{\pm\delta}.
 \end{aligned}
 \tag{2.64}$$

We will also consider the part of the boundary where $\arg z = \delta$ as part of the horizontal boundary. Figure 20 shows these regions, with z drawn as 2-dimensional and v drawn as 1-dimensional; Figure 21 is an attempt to draw their images (drawn separately for W_1 on the left and W_2 on the right.)

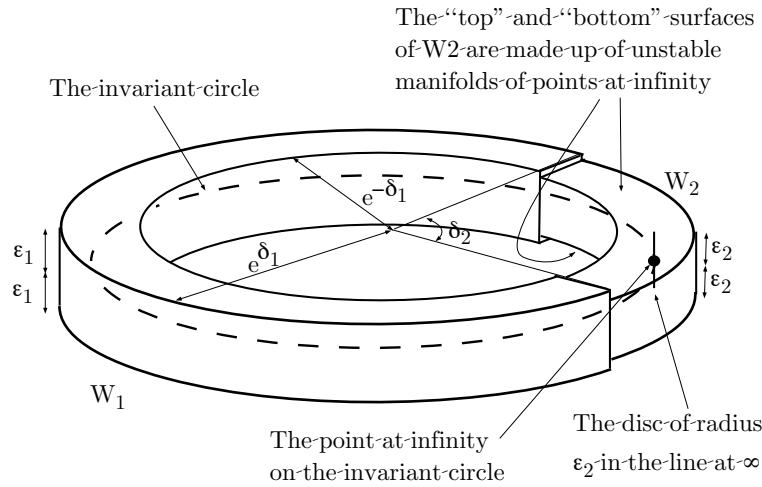


FIGURE 20. A sketch of the regions W_1 and W_2 . Of course, these are really 4-dimensional; we have represented the coordinate v as 1-dimensional, whereas z is represented as a thickened circle in a plane, which it indeed really is.

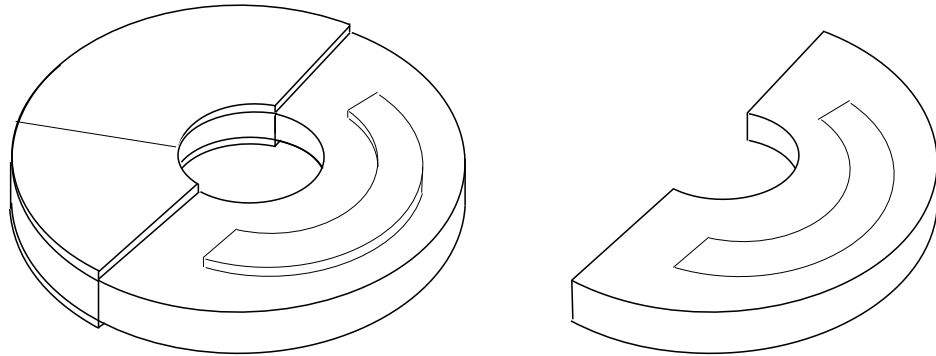


FIGURE 21. The images of W_1 and W_2 , with the image of W_1 on the left and W_2 on the right. The space W_1 really maps “hyperbolically,” contracting in the vertical direction and expanding in the horizontal direction. But W_2 does not: the map does expand in the horizontal direction, but the horizontal boundary is contained in the image of the horizontal boundary, since it is made up of unstable manifolds. With the convention of the drawing, i.e., that v is 1-dimensional, one should imagine the top and bottom of W_2 as each a single unstable manifold, which is simply “spread out” over itself by $z \mapsto z^2$.

We are now in a position to state our principal lemma, which allows us to set up a sort of “graph transform.”

Lemma 2.4.5. *We can choose $\epsilon_1, \epsilon_2, \delta$ so that if $D \subset W$ is an analytic embedded disk with*

$$\begin{aligned} T_w D \subset C_w \text{ for every } w \in D & & (D \text{ is almost vertical}); \\ \partial D \subset \partial_H(W) & & (D \text{ crosses } W \text{ from top to bottom}); \\ D \cap \{|z| = 1, v = 0\} \text{ is a single point} & & (D \text{ crosses the circle}); \end{aligned}$$

then $N^{-1}(D) \cap W$ has exactly two components that have the same properties.

Proof. Most of the difficulty in the proof is that we don’t have much control on where the unstable manifolds of the points at infinity are: all we know is that the map Φ of Theorem 2.3.1 is analytic. Fortunately that is enough. It will be convenient to work with the parameter $\zeta = \log z$ near $z = 1$. By the implicit function theorem, the unstable manifold through the point of coordinate $w = y/x$ at infinity is the graph of a function σ_w , i.e., it has equation $v = \sigma_w(\zeta)$, at least for $|w|$ sufficiently small.

Lemma 2.4.6. *There exist $\epsilon_0 > 0, \delta_0 > 0$, and K such that*

$$|\sigma_w(\zeta) - w| \leq K|\zeta||w| \tag{2.65}$$

when $|\zeta| < \delta_0, |w| < \epsilon_0$.

Proof. Since the implicit function theorem applied to analytic data gives analytic implicit functions, the functions $\sigma_w(\zeta)$ are analytic functions that depend analytically on w , i.e., they are given by convergent power series:

$$\sigma_w(\zeta) = w + a_1(w)\zeta + a_2(w)\zeta^2 + \dots \tag{2.66}$$

Moreover, $a_i(0) = 0$ for all i , since $\sigma_0(\zeta) = 0$, reflecting the fact that l is the unstable manifold of the point at infinity on l . The result is now standard. \square

Having found K , we can now choose $\delta_1 > 0$ sufficiently small that

$$B < \frac{1 - K\delta_1}{1 + K\delta_1}, \tag{2.67}$$

and set $\delta = \delta_1/3$, and ϵ_1, ϵ_2 satisfying

$$\epsilon_1 > \epsilon_2(1 + K\delta_1), \quad \epsilon_1 < \epsilon_2(1 - K\delta_1), \tag{2.68}$$

which is possible by Equation (2.67). We will require that $\epsilon_1, \epsilon_2, \delta_1$ have all been chosen sufficiently small that $W \subset \tilde{W}$, so that our invariant cone fields exist. From here on we will not change δ again, but we may need to further shrink ϵ_1, ϵ_2 , making sure that they always satisfy Equation (2.68). The requirements we want are:

- for all $(z, v) \in W_1$, the second coordinate of $N_F(z, v)$ is smaller than $|v|$;
- the cone $C \begin{pmatrix} z \\ 0 \end{pmatrix}$ intersects only the horizontal boundary of W .

We claim that with these requirements, Lemma 2.4.5 is true. This is mainly a matter of seeing that the image of the interior of W does not intersect the horizontal boundary of W . This should be clear if both the point (z, v) and its image (z_1, v_1) are in W_1 , since N_F contracts strictly in the vertical direction in W_1 , so $|v_1| < |v|$. It should also be true if the point and its image are in W_2 : the point is on some stable manifold, its image is on the same stable manifold, and this manifold

is not on the boundary of W_2 . The case where the point is in W_2 and its image is in W_1 is slightly more complicated: it is on the manifold of a point w with $|w| < \epsilon_2$, and so is its image. Moreover, $|\log z| < \delta_1/2$, so $\log z_1 < \delta_1$, and

$$|\sigma_w(z_1)| < |w| + K\delta < \epsilon_2 + K\delta < \epsilon_1. \quad (2.69)$$

The hardest case is if a point $(z, v) \in W_1$ is mapped to a point (z_1, v_1) with $|\arg z_1| < \delta$ and $e^{-\delta} < |z| < e^\delta$; in this case, $(z_1, v_1) \in W_2$, and is not on the horizontal boundary of W_2 . The hypothesis implies that z_1 is close to -1 , and the normal multiplier in the entire left half-plane is at most $|B|$. So $|v_1| < |Bv|$, and the unstable manifold through (z_1, v_1) will intersect the line at infinity at a point v_2 with $|v_2 - v_1| < K\delta$, hence with $|v_2| < B\epsilon_1 + K\delta < \epsilon_2$.

The conditions we have set up will still be satisfied if we replace $N_{\alpha, \beta}$ by $N_{\alpha', \beta'}$, with the point (α', β') chosen sufficiently close to (α, β) ; W_1 is kept precisely as it was; and W_2 is chosen as the union of the unstable manifolds of the same points $z = 1, |v| < \epsilon_2$ for the new map.

Now suppose that $z_1^2 = z_2$, and that D crosses the circle at z_2 . There is then a unique component D' of $N^{-1}(D) \cap W$ that contains z_1 , and D' is an analytic disc in \mathbb{P}^2 . Our first requirement guarantees that D intersects only the horizontal boundary of W , so the thing to check is that no part of $\partial D'$ is in the interior of W . If a sequence in D' converges to a point of the boundary of D' that is an interior point of W , its image also converges to an interior point of W , and this is a contradiction. \square

The two lemmas together show that the graph transform map is well defined.

Set X_1 to be the polar locus, and $X_{n+1} = N_F^{-1}(X_n)$. Let $Y_{n,k}$ be the component of $X_n \cap W$ that contains the dyadic root of unity $\zeta_{n,k} = e^{2\pi ik/2^n}$. Moreover, let $\epsilon = \epsilon_2 - K\delta$.

Corollary 2.4.7. *Each $Y_{n,k}$ is the graph $\Gamma_{f_{n,k}}$ for some $f_{n,k} : D_\epsilon \rightarrow \mathbb{C}$ expressing z as a function of $v \in D_\epsilon$, with $f_{n,k}(0) = \zeta_{n,k}$.*

Proof. This follows immediately from Lemma 2.4.5. \square

Let $\Delta \subset S_l$ be the set of dyadic angles.

Corollary 2.4.8. *The map $h : \Delta \times D_\epsilon \rightarrow \mathbb{C}$ given by*

$$h : (\zeta_{n,k}, v) \mapsto f_{n,k}(v) \quad (2.70)$$

is a holomorphic motion parametrized by D_ϵ .

Proof. All the conditions for a holomorphic motion are clearly satisfied: the map is clearly injective and analytic with respect to v . \square

We can now prove Theorem 2.4.1. By the λ -lemma ([MSS]), the map h extends to a map $\tilde{h} : S_l \times D_\epsilon \rightarrow \mathbb{C}$, quasiconformal with respect to z and analytic with respect to v . The set V_l is the image of \tilde{h} , parametrized by $S_l \times D_\epsilon$ where $u = h(\zeta, v)$ is a 3-dimensional manifold, and N maps V_l to itself. Since the diagram

$$\begin{array}{ccc} V_l & \longrightarrow & V_l \\ \downarrow & & \downarrow \\ S_l & \longrightarrow & S_l \end{array} \quad (2.71)$$

commutes on the set parametrized by $\Delta \times D_\epsilon$, it commutes on $S_l \times D_\epsilon$. Since the fibers of the projection pr_1 are graphs of analytic functions, they are analytic discs. \square Theorem 2.4.1

In Figures 22–24 we give examples to demonstrate both why the boundaries of the basins look smooth, and why they are not. If a bundle V_l as above exists, we should see it as a line separating the red and green basins.

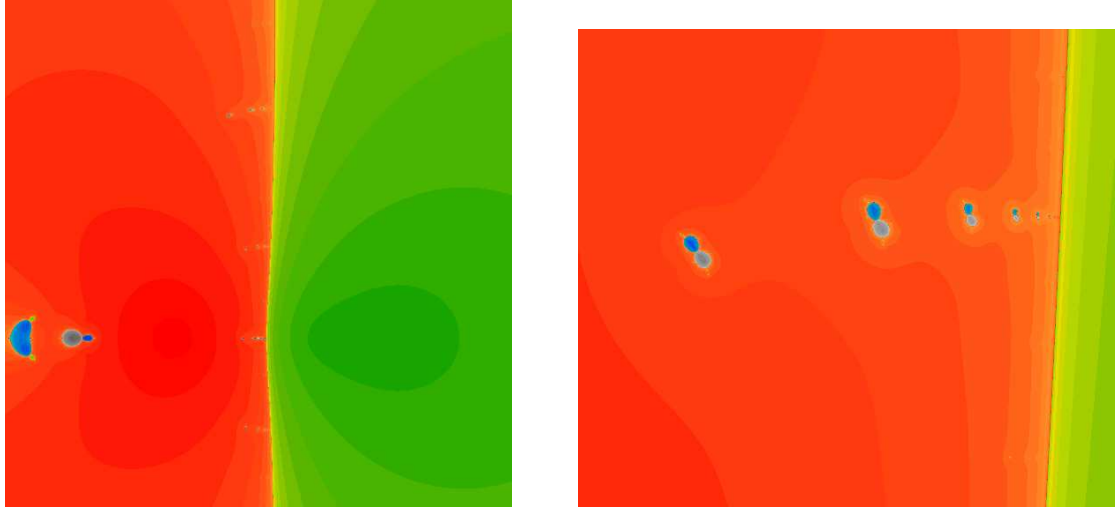


FIGURE 22. This represents the line $y = .1x$, as parametrized by x , where we have $-1 \leq \operatorname{Re} x, \operatorname{Im} x \leq 2$, for the Newton map where $\alpha = \beta = -1$. The circle in the x -axis (which is the straight line $\operatorname{Re} x = 1/2$) has multiplier $B/(B - 1) = 2/3$, so it is attracting, and in fact has a stable manifold by Theorem 2.4.1. You would expect this stable manifold to cut the line $y = .1x$ along a quasi-circle close to the line $\operatorname{Re} x = 1/2$. Indeed, this is the case. Note the complicated decorations in the basin of the origin, red in the picture; one bubble sequence is blown up at right. This is also as one would expect: the bubbles arise because there is a point of indeterminacy on the x -axis, and they are in the red basin because the point of indeterminacy is in the red basin. They present a dyadic structure, as they should.

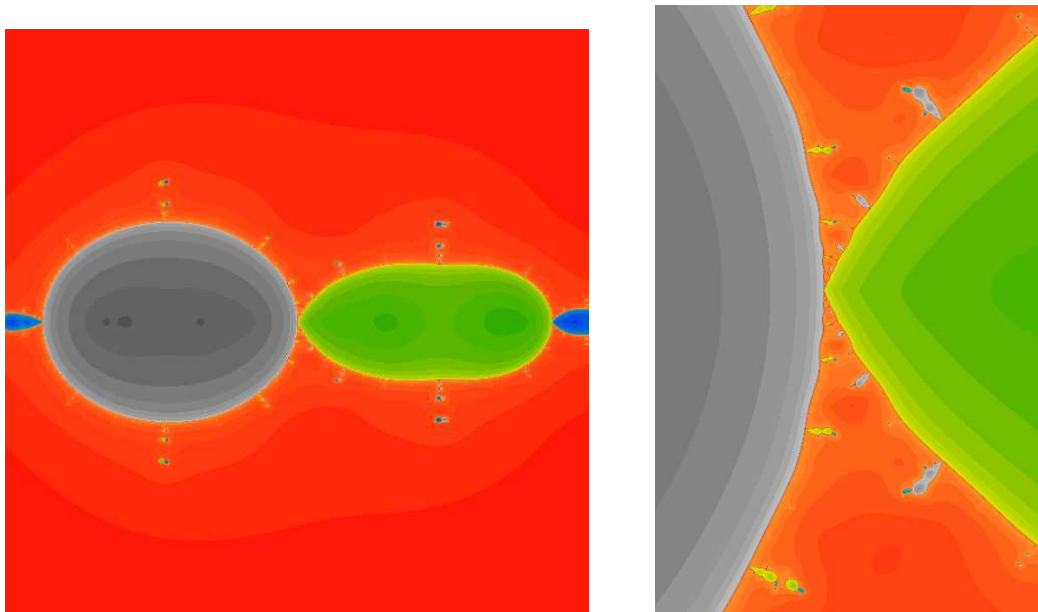


FIGURE 23. This represents the line of equation $y = .6x$, for the same Newton's method as above. This time, we are plotting $1/x$, i.e., near the point at infinity. The undulations of the quasi-circles are very apparent.

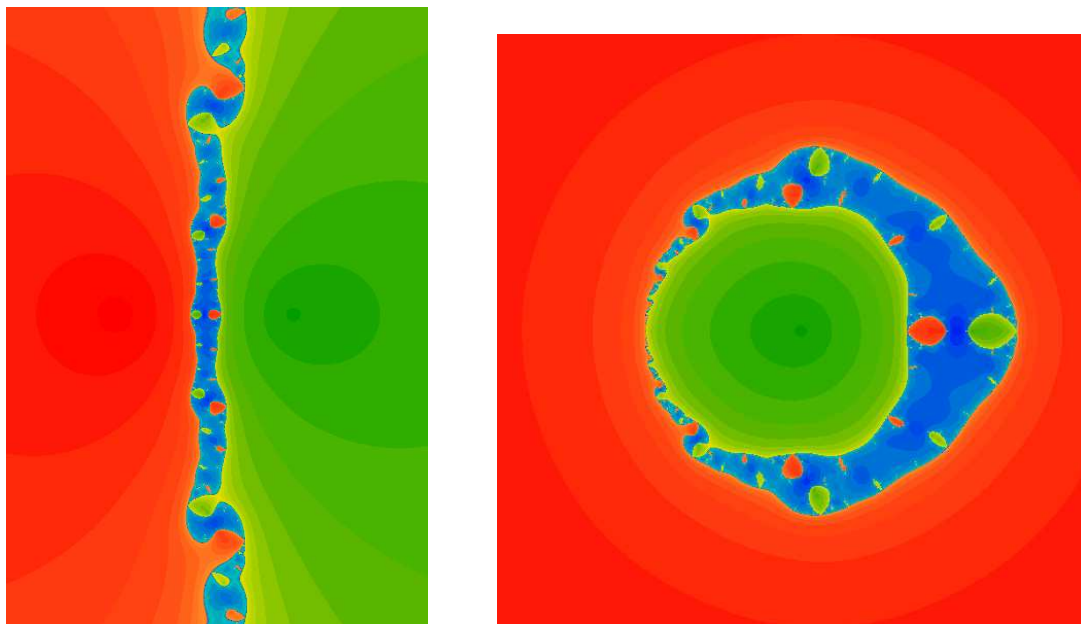


FIGURE 24. This represents the line $y = .1x$, for the Newton's method with $\alpha = -1, \beta = 3$. This is actually conjugate to the one above, but this time the circle in the x -axis is repelling (with multiplier 2, since the point of indeterminacy is on it). When we perturb the slope away from 0, we would not expect a curve to separate the red and green basins, and indeed there is no such curve; instead, a zone of blue appears. On the right, we see the same line, but this time near infinity. Since the red and green basins touch along the line at infinity, they must touch along its inverse images, which have a dyadic structure. That is what one observes.

2.5. The extension of Φ and the origin of “bubbles.”

The mapping

$$\psi_m(w) = \Phi^{-1}(w, m) \quad (2.72)$$

is *a priori* defined, analytic, and injective on some small disc D_r , and satisfies $\psi_m(0) = [1 : m : 0]$ and $\psi_m(2w) = N_F(\psi_m(w))$ for $w \in D_{r/2}$. In other words, it is a linearizing parametrization of the unstable manifold of $[1 : m : 0]$; moreover, it depends analytically on $m \in \mathbb{C}^*$.

Proposition 2.5.1. *For each $m \in \mathbb{C}^*$, the mapping ψ_m extends to an analytic mapping $\psi_m : \mathbb{C} \rightarrow \mathbb{P}^2$.*

Proof. The extension is defined by the formula $\psi_m(2w) = N_F(\psi_m(w))$. The only difficulty one can run into is when $\psi_m(w)$ is a point of indeterminacy, and even then there is no real difficulty. On any disc $D \subset \mathbb{C}$, the map $\psi_m : D \rightarrow \mathbb{P}^2$ lifts to a mapping $\tilde{\psi}_m : D \rightarrow \tilde{\mathbb{P}}_I^2$, and $N_F : \tilde{\mathbb{P}}_I^2 \rightarrow \mathbb{P}^2$ is well defined. \square

As we will see, the extended ψ_m is no longer injective; in fact, we expect the image to be in general very wild, perhaps dense in all of \mathbb{C}^2 .

Another unpleasant feature is that the extended ψ_m does not depend continuously on m , and this gives rise to one of the characteristic features of drawings in complex slices: *bubble sequences*. We will now explore this phenomenon.

There are some values of m for which ψ_m can be completely understood, and in fact an explicit formula can be given, namely those that are the slopes of the invariant lines. Within these invariant lines the point at infinity is a repelling fixed point, and it is well known that at a repelling fixed point of a rational function, the linearizing map extends to an entire meromorphic function.

For the map $z \mapsto z^2$ and the fixed point 1, this extended linearizing map is $w \mapsto e^w$. In the form (1.24), the x -axis is an invariant line, mapped to itself by N_F by the formula

$$N_F \begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} x^2/(2x - 1) \\ 0 \end{pmatrix}. \tag{2.73}$$

If we set $z = x/(x - 1)$, this becomes $z \mapsto z^2$. For this slope we have

$$\psi_m(w) = \frac{e^w}{e^w - 1}. \tag{2.74}$$

In the w -plane, we see the imaginary axis mapping as a universal cover to the invariant circle $\operatorname{Re} x = 1/2$, and separating the inverse images of the basins of 0 and 1 (the roots are “at $\pm\infty$ ”).

This picture is a bit misleading. Somewhere on the invariant line is the point of indeterminacy $1/B$, probably in one of the basins but maybe on the invariant circle, and its inverse image under ψ_0 is a geometric sequence with increment $2\pi i$. The inverse images of these points give a “dyadic structure” on a neighborhood of the imaginary axis, on one side only, as shown in Figure 25.

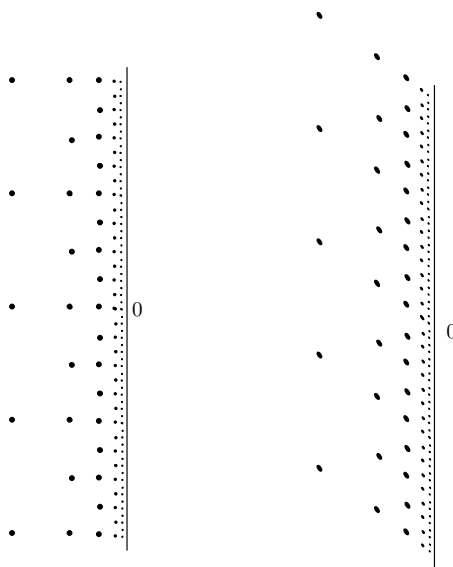


FIGURE 25. The dyadic structure of inverse images of points of indeterminacy. If the point generating the structure is real, the dyadic pattern of points is as on the left, but if it is not real the pattern will look skewed, as on the right. The right-hand side of Figure 29 shows such a distorted dyadic pattern.

Proposition 2.5.2. *If $\psi_0(w_0) = \begin{pmatrix} 1/B \\ 0 \end{pmatrix}$, then $\psi_m(w)$ is discontinuous at the point $m = 0, w = 2w_0$.*

Proof. There is no difficulty in showing that Φ^{-1} can be extended analytically to the region $|\operatorname{Re} w| < \operatorname{Re} w_0 + \epsilon$, $|m| < \epsilon$ for $\epsilon > 0$ sufficiently small. If $r < \epsilon$, and $m \neq 0$ satisfies $|m| < \epsilon$, then the points $\psi_m(w)$ form a disc that passes very near the point of indeterminacy \mathbf{q}_1 . If we blow this point up, the lift of this disc will come very close to the entire exceptional divisor, and its image will come very close to the image of the exceptional divisor, which is a straight line (a line tangent to both parabolas in the normalization (1.34).) But the original invariant line does not come close to the image of the exceptional divisor. \square

Thus we see something like the Picard-Lefschetz formula appearing: there is a “vanishing cycle” on the unstable manifolds ψ_m , whose images, as m tends to 0, tend to the union of the invariant line and the image of the exceptional divisor. In the disc D_r above, and at all its forward and inverse images, i.e., at all the points $2^n(w_0 + 2k\pi i)$, $n, k \in \mathbb{Z}$, we should see all of the dynamics of the exceptional divisor. But for m sufficiently small, we should see them only in a small neighborhood of the dyadic grid; at every other point w , $\psi_m(w)$ is analytic as a function of m for m sufficiently small. This is the source of the bubbles.

What should one see in a bubble? Mainly a third basin. The point of indeterminacy is the intersection of two invariant lines. The disc $\psi_m(D_r)$ above will intersect the other invariant line somewhere, and a neighborhood of that point will be in the basin of the root to which the point of indeterminacy is attracted in the other invariant line. Usually, the bubble is primarily made up of that basin, but with complications around the edges, which can range from other bubble sequences in the simplest cases to horribly complicated things in other cases.

Interestingly enough, it appears that there are values of the parameters such that the images of the ψ_m enter only three basins, not all four, for m sufficiently small. For instance, when $\alpha = \beta = 2$, we have been unable to find any points of the basin of $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ in the unstable manifolds parametrized by ψ_m for $|m| < .1$. This has consequences for the Russakovskii-Shiffman measure (discussed in Section 3.3): it is then not true that the entire line at infinity is in the support of the measure.

Figure 26 represents the unstable manifold $\psi_m(w)$ for $m = .05(1 + i)$, and $\alpha = \beta = 2$.

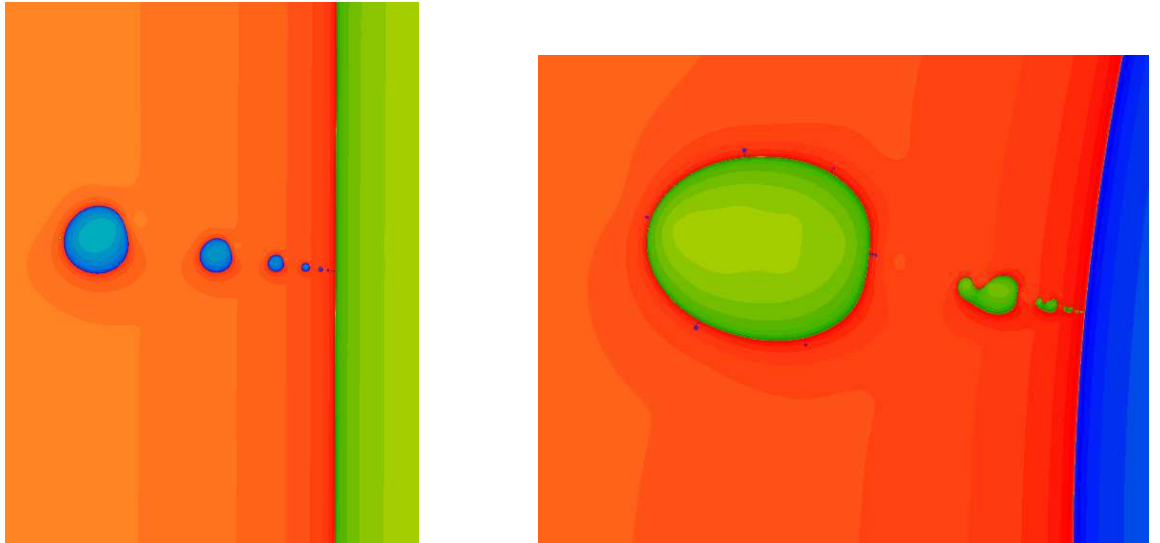


FIGURE 26. Left: a sequence of blue bubbles (blue = basin of $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$) in the red basin (red = basin of $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$). Right: a blow-up shows what the green dot off the left-most blue bubble really looks like.

In this case, the mapping is real, and we can easily visualize the unstable manifold and the bubbles.

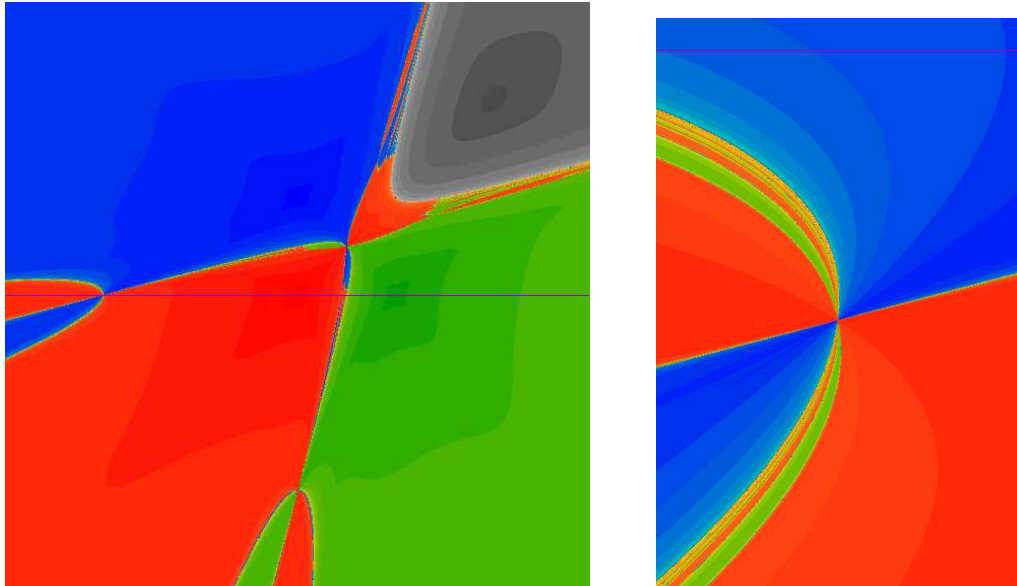


FIGURE 27. The real picture of Newton's method when $\alpha = \beta = 2$. A line $y = mx$ with m small will be mainly red and green, with blue interruptions in the red, near the point of indeterminacy and its inverse images. The blow-up at right shows red and green stripes near the point of indeterminacy; these stripes correspond to the green balls in the red basin in Figure 26, right. There does not appear to be any gray in either the real or the complex picture.

The examples above are misleadingly simple. Figure 28 shows what happens when $\alpha = i, \beta = (1 + i)/2$, in the line $y = .02x$. The decorations around the boundaries of the bubbles are now much more complicated, and all four colors are present in the picture.

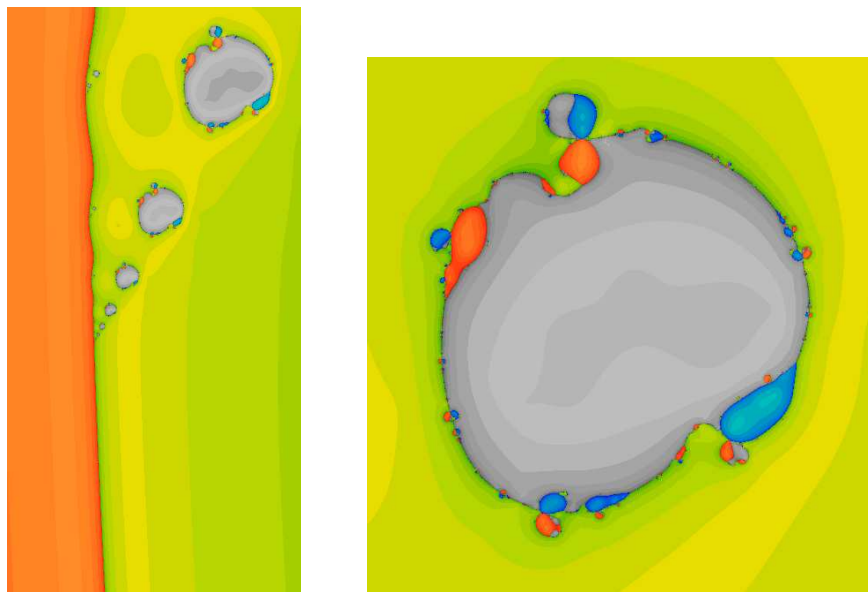


FIGURE 28. More complicated bubbles

This picture brings out several interesting features. One is the undulations of the quasi-circle making the boundary between the red and the green basin. Another is the dissymmetry between the part above the origin (which corresponds to the point at infinity) and the part beneath. The discussion in Section 2.2 shows that such lack of symmetry is to be expected when B is not real, and indeed here $B = (1 + i)/2$ is not real.

Attracting circles with repelling cycles

It is perfectly possible for an invariant circle to be attracting on average, but to contain repelling cycles. In that case, Pesin theory guarantees that the circle has a stable manifold almost everywhere; but of course the repelling cycle cannot have a stable manifold, and neither can its inverse images, which are dense in the circle. No neighborhood of the invariant circle like the set W above can exist, and neither can there be a “center-stable” manifold as above with sections that are quasi-circles. It isn’t quite clear what to expect instead.

Our best guess is something like “Arnold tongues,” open tongues emanating from the repelling cycle, and making islands in which one should see all sorts of forward behavior, leaving a Cantor set of large measure; this set is the “measure-theoretic stable manifold.”

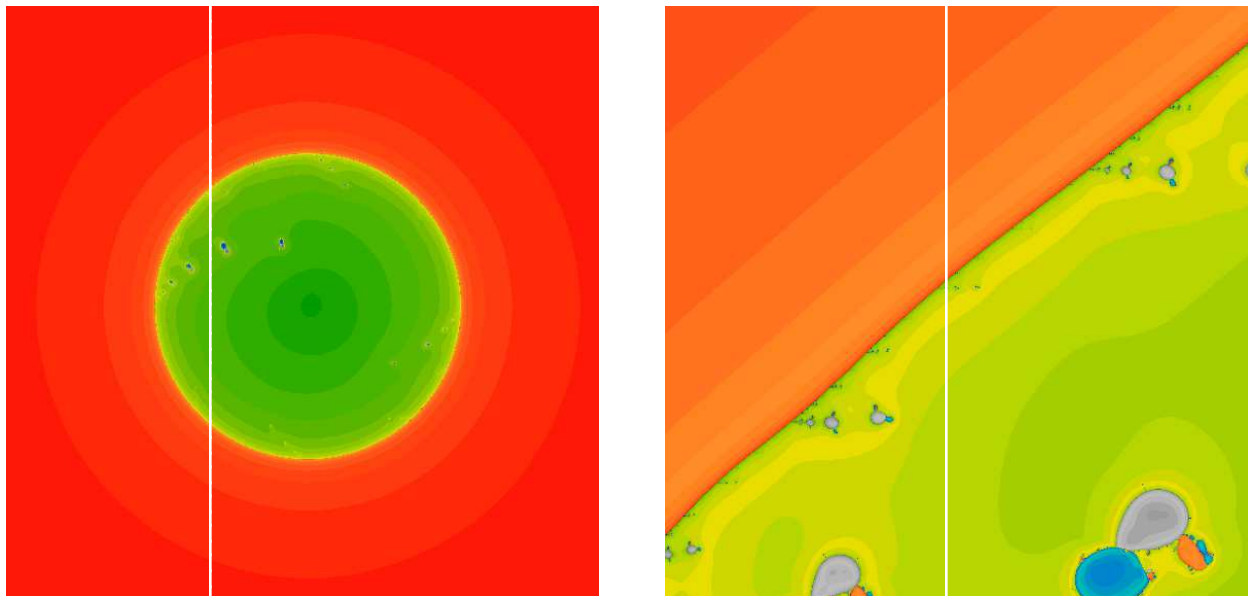


FIGURE 29. On the left, we see what looks like a stable manifold, separating the green and red basins, with bubbles in the green basin exhibiting the normal sort of dyadic behavior. The vertical line has equation $\operatorname{Re} z = 1 - \cos 2\pi/7 \approx .3765$, and should intersect the “circle” near the points $\zeta_{3,1}$ (at the top) and $\zeta_{3,6}$ (at the bottom), since angles are computed on the circle viewed as the boundary of the outside (the basin of the origin). The cycle containing $\zeta_{3,1}$ is repelling, as we saw, and the cycle containing $\zeta_{3,6}$ is attracting, with multiplier of absolute value $\approx .4103$. On the right, we see a first blow-up at the top, with many more bubble sequences, but the circle still looks smooth. This region is blown up further in Figure 30.

To construct an example, notice that the multiplier of the 3-cycle $\phi(\zeta_{3,k})$ is

$$\frac{B}{1 + \omega(B-1)} \frac{B}{1 + \omega^2(B-1)} \frac{B}{1 + \omega^4(B-1)} = \frac{B^3}{B + (B-1)(\omega + \omega^2 + \omega^4) + (B-1)^2(\omega^3 + \omega^5 + \omega^6)},$$

where $\omega = e^{2\pi i/7}$. For the value $B = .8 + .4i$, the invariant circle in the x -axis is attracting (with average multiplier $|B| = \sqrt{.89}$), but the 3-cycle $\phi(\zeta_{3,k})$ is repelling, with multiplier of absolute value ≈ 2.556 . We will look in the line of equation $y = .05x$, and in the local coordinate $z = 1/x$. Moreover, we will draw the region $-1 \leq \operatorname{Re} z \leq 3$, $|\operatorname{Im} z| < 2$. In the x -axis itself, the invariant circle is the circle of radius 1 centered at 1, with the basin of the origin outside and the basin of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ inside. At that scale, in the line $y = .05x$ we see the “circle” shown in the left-hand side of Figure 29. However, the apparent circle does not really exist. If we perform three further blow-ups, we find the structure on the right of Figure 30. Similar complications are dense on the entire circle, but at very small scales at most places.

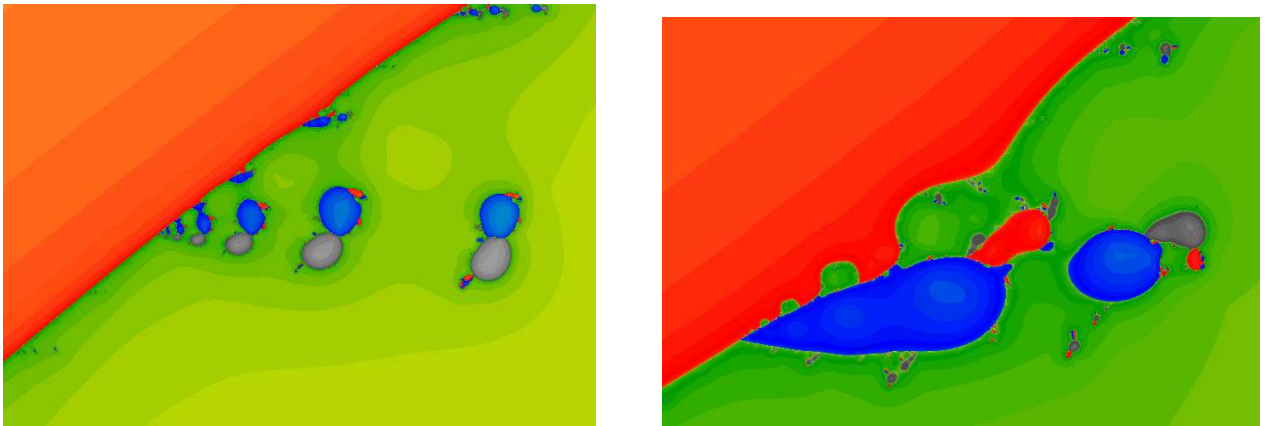


FIGURE 30. On the left, a further blow-up of the right-hand side of Figure 29 shows problems with the proposed “circle,” presumably the appearance of a “tongue” emanating from the repelling cycle. A further blow-up gives some details of what one should see within such a tongue.

3

The behavior at infinity when $a = b = 0$

In this chapter we will study Newton's method as applied to the two equations

$$y = x^2, \quad x = y^2. \tag{3.1}$$

The associated Newton's method is

$$N : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{4xy - 1} \begin{pmatrix} 2x^2y + y^2 \\ 2xy^2 + x^2 \end{pmatrix}. \tag{3.2}$$

The dynamics of this map $N = N_{0,0}$ is probably not substantially simpler than any other $N_{a,b}$, but the formulas are simpler. In Chapter 6 we will show that all the same phenomena occur for all $N_{a,b}$.

The specific mapping N was studied by Russakovskii and Shiffman. In [RS], they study rational mappings $P : \mathbb{P}^n \rightsquigarrow \mathbb{P}^n$, whose topological degree is greater than the degree of $P^{-1}(W)$, where W is a generic hyperplane. This is the case for the mapping above, with topological degree 4 and algebraic degree 3. They prove that in this case there is an invariant measure; but in the case of the mapping N they could not decide whether this measure charges the points of indeterminacy at infinity. As an example of the untangling allowed by Farey blow-ups, we answer this in Section 3, showing that the points of indeterminacy are not charged, and in fact giving a fairly detailed description of the measure.

3.1. The primitive space

Consider in \mathbb{P}^2 the union of the axes and the line at infinity. We will blow up the double points at infinity, then all the double points created, then all the double points created, etc., creating compact algebraic surfaces $X_0 = \mathbb{P}^2, X_1, X_2, \dots$, together with mappings $\pi_n : X_n \rightarrow X_{n-1}$. Call $D_0 \subset \mathbb{P}^2$ the line at infinity, and $D_n = \pi_n^{-1}(D_{n-1})$, $n \geq 1$. For each n , D_n is a union of rational curves, intersecting each other transversally, and will be called the *divisor at infinity* of X_n .

We call the projective limit $X_P = \varprojlim_n (X_n, \pi_n)$ of these surfaces the *primitive space*, for the following reason.

Consider finite sequences P_i of elements of \mathbb{Z}^2 defined as follows:

$$\begin{aligned} P_0 &= \left\{ \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}, \\ P_1 &= \left\{ \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}, \\ P_2 &= \left\{ \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}, \\ &\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \end{aligned} \tag{3.3}$$

and so forth, making P_{n+1} by intercalating $\begin{pmatrix} p_1 + p_2 \\ q_1 + q_2 \end{pmatrix}$ between each pair of consecutive entries $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$ and $\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}$ of P_n . The P_n are made up of points whose coordinates are coprime. Two vectors

$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}, \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} \in \mathbb{Z}^2$ are consecutive in some P_n exactly if $|p_1q_2 - p_2q_1| = 1$; we will call such vectors neighbors.

Set $P = \cup_n P_n$; this set is a lot like the set of rationals: the ratios p/q enumerate all the rational numbers, except that each negative rational slope corresponds to two points, and one point corresponds to ∞ (namely, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$). The order in which the vectors appear is like the Farey sequence as a way of enumerating the rationals.

Proposition 3.1.1. *The irreducible components of D_n correspond to the elements of the sequence P_n , in the sense that if $\begin{pmatrix} p \\ q \end{pmatrix} \in P_n$, then the curve in \mathbb{C}^2 parametrized by*

$$t \mapsto \begin{pmatrix} t^p \\ t^q \end{pmatrix} \tag{3.4}$$

tends, as $t \rightarrow \infty$, to a point of D_n that is an ordinary point of an irreducible component $L\begin{pmatrix} p \\ q \end{pmatrix}$, and every irreducible component is reached in this way.

We will prove the stronger result (Proposition 3.1.2) below, providing curves that tend to all the points of D_n , allowing us to parametrize these components. The underlying reason our construction works is that the function x^qy^{-p} extends to $L\begin{pmatrix} p \\ q \end{pmatrix}$; this extension provides an isomorphism $L\begin{pmatrix} p \\ q \end{pmatrix} \rightarrow \mathbb{P}^1$. We will use this as follows: since p and q are coprime, there exist integers k and l such that $lq - kp = 1$.

Proposition 3.1.2. *The curve $x^qy^{-p} = m$ is parametrized by*

$$t \mapsto \begin{pmatrix} m^l t^p \\ m^k t^q \end{pmatrix}, \tag{3.5}$$

and the mapping

$$m \mapsto \lim_{t \rightarrow \infty} \begin{pmatrix} m^l t^p \\ m^k t^q \end{pmatrix} \tag{3.6}$$

exists for $m \in \mathbb{C}^$ and is an isomorphism $\mathbb{C}^* \rightarrow L^*\begin{pmatrix} p \\ q \end{pmatrix}$, where $L^*\begin{pmatrix} p \\ q \end{pmatrix}$ is the complement of the two double points on $L\begin{pmatrix} p \\ q \end{pmatrix}$. The limit point will be referred to as the point of coordinate m on $L\begin{pmatrix} p \\ q \end{pmatrix}$.*

Proof. We will prove this by induction. Thus, we will assume that the statement is true for X_n , and show it for X_{n+1} . Let $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}, \begin{pmatrix} p_2 \\ q_2 \end{pmatrix}$ be two neighbors in P_n ordered so that $p_1q_2 - p_2q_1 = -1$.

Then the relations

$$m_1 = x^{q_1}y^{-p_1}, \quad m_2 = x^{q_2}y^{-p_2} \tag{3.7}$$

express x and y in terms of m_1 and m_2 :

$$x = m_1^{p_2}m_2^{-p_1}, \quad y = m_1^{q_2}m_2^{-q_1}. \tag{3.8}$$

This shows that we can use (m_1, m_2) as coordinates. The blow-up is being performed at the point $m_1 = \infty$, $m_2 = 0$, so coordinates on the blow-up are provided by

$$m_3 = m_1 m_2 \quad \text{and} \quad m_2. \quad (3.9)$$

In terms of x and y , we have $m_3 = x^{q_1+q_2} y^{-(p_1+p_2)}$, and the curve

$$t \mapsto \begin{pmatrix} t^{p_1+p_2} \\ t^{q_1+q_2} \end{pmatrix} \quad \text{is the curve} \quad m_3 = 1, \quad m_2 = 1/t, \quad (3.10)$$

So it does indeed lead to a point of $L \begin{pmatrix} p_1+p_2 \\ q_1+q_2 \end{pmatrix}$. \square

3.2. Newton's method and the primitive space

The reason this primitive space is of interest to us is that it almost represents a natural compactification of \mathbb{C}^2 to which Newton's method almost extends.

Part of the reason for this is that the mapping N takes the x -axis to the y -axis, which is mapped back to the x -axis, by the formula

$$\begin{pmatrix} t \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ -t^2 \end{pmatrix} \mapsto \begin{pmatrix} t^4 \\ 0 \end{pmatrix}. \quad (3.11)$$

In Chapter 6, we will see how the primitive space must be modified when a and b are arbitrary; in that case the axes have no particular dynamical significance, and we will need to use other curves in their place.

Proposition 3.2.1. *The Newton mapping extends as a rational map to the primitive space, giving $N_P : X_P \rightsquigarrow X_P$, with four points of indeterminacy:*

- the point of coordinate -2 in $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and the point of coordinate $-1/2$ in $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$;
- the point of coordinate 4 in $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and the point of coordinate $1/4$ in $L \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Note that the points of the first pair are at the intersections of D_∞ with the parabolas $x+2y^2=0$ and $y+2x^2=0$; each of these parabolas is one component of the inverse image of the x -axis and the y -axis (the axes are tangent to a critical value parabola at a root, so the inverse images are in agreement with Proposition 1.5.1). The points of the second pair are at the intersection of D_∞ with the polar hyperbola of equation $4xy=1$.

Proof. The curve parametrized by

$$t \mapsto \begin{pmatrix} m^l t^p \\ m^k t^q \end{pmatrix}, \quad (3.12)$$

maps to the curve

$$t \mapsto \frac{1}{4m^{l+k} t^{p+q} - 1} \begin{pmatrix} 2m^{2l+k} t^{2p+q} + m^{2k} t^{2q} \\ 2m^{l+2k} t^{p+2q} + m^{2l} t^{2p} \end{pmatrix}. \quad (3.13)$$

We need to show that as $t \rightarrow \infty$, this curve has a well defined limit in X_P . This is a matter of making sure what terms in the expressions above are dominant. To keep track of the various terms, let us consider the (p, q) -plane (in which one should remember that only the points with coprime

integer coordinates, and at least one coordinate positive, represent components of D_∞), broken up as shown in Figure 31 into five zones.

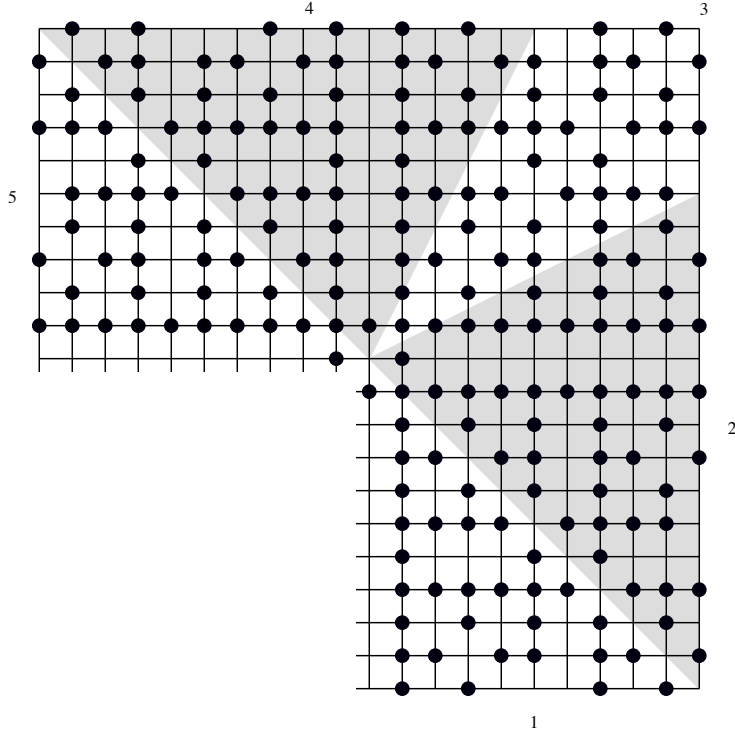


FIGURE 31. The marked points of the grid correspond to components of D_∞ ; the five zones are distinguished by the shading.

In the different zones, different terms are dominant, and the four points on the boundaries of zones correspond to the components of D_∞ where there are competing terms.

The mapping in zone 1: $p + q < 0$ and $p > 0$

In zone 1, the dominant terms are

$$t \mapsto \frac{1}{-1} \begin{pmatrix} 2m^{2l+k}t^{2p+q} \\ m^{2l}t^{2p} \end{pmatrix}. \tag{3.14}$$

This breaks up into two cases, depending on whether q is even or odd.

If q is even, say $q = 2q_1$, then the image component is $L \begin{pmatrix} p+q_1 \\ p \end{pmatrix}$, and the coordinate in that component is

$$x^p y^{-(p+q_1)} = (-1)^{q_1} 2^p m^{-1}. \tag{3.15}$$

Thus we see that the component $L \begin{pmatrix} p \\ q \end{pmatrix}$ is mapped by an isomorphism to the component $L \begin{pmatrix} p+q_1 \\ p \end{pmatrix}$.

However, the even power of t in both coordinates means that this component is a critical curve: a small circle around such a component is mapped to a circle that wraps twice around the image component, much closer than the original circle. This will be a key consideration when understanding invariant measures.

If q is odd, the situation is quite different. Then the image component is $L\left(\begin{smallmatrix} 2p+q \\ 2p \end{smallmatrix}\right)$, and the coordinate within this component is

$$x^{2p}y^{-(2p+q)} = -2^{2p}m^{-2}. \quad (3.16)$$

Thus we see that the domain component is mapped by a rational map of degree 2 to its image, with ramification at $m = 0$ and $m = \infty$.

The case of $\left(\begin{smallmatrix} p \\ q \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ (on the boundary of zones 1 and 2)

The line $L\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ is parametrized by $t \mapsto \left(\begin{smallmatrix} t/m \\ t^{-1} \end{smallmatrix}\right)$. This curve maps to

$$\frac{1}{4/m - 1} \left(\begin{smallmatrix} 2t/m + t^{-2} \\ 2m^{-1}t^{-2} + t^2/m^2 \end{smallmatrix} \right). \quad (3.17)$$

Clearly the first term dominates in the first coordinate and the second term dominates in the second coordinate, so $L\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ maps to $L\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$, with the point of coordinate m mapping to the point of coordinate

$$x^2y^{-1} = \frac{1}{4/m - 1} \frac{4}{m^4}(m^2) = \frac{4}{m(4 - m)}, \quad (3.18)$$

which is a rational function of degree 2.

Let us look a bit more carefully at this rational function. The point of coordinate $m = 4$ in $L\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ is one of the two points where the polar curve of equation $4xy = 1$ intersects D_∞ (the other is in $L\left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right)$). Since the polar curve maps to the line at infinity, which is $L\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ in the present notation, and $L\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ does not intersect $L\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, this will create a problem, and this point $m = 4$ will be a point of indeterminacy of the extended Newton map.

How about critical points? The rational function $\frac{4}{m(4-m)}$ has two critical points, at ∞ and at $m = 1/2$. The latter is one of the intersections of the critical cubic C_2 with D_∞ , and maps to the point where the critical value parabola of equation $y = x^2$ intersects D_∞ in $L\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$. The other critical point is accounted for, like all the critical points at 0 and ∞ , by the fact that many components of D_∞ are themselves critical.

The case of zone 2: $p + q > 0$ and $2q < p$

In zone 2, the parametrized curve

$$t \mapsto \left(\begin{smallmatrix} m^l t^p \\ m^k t^q \end{smallmatrix} \right) \quad (3.19)$$

is mapped to the parametrized curve

$$t \mapsto \frac{1}{4m^{l+k}t^{p+q} - 1} \left(\begin{smallmatrix} 2m^{2l+k}t^{2p+q} + m^{2k}t^{2q} \\ 2m^{l+2k}t^{p+2q} + m^{2l}t^{2p} \end{smallmatrix} \right). \quad (3.20)$$

In the denominator the term t^{p+q} dominates over the constant -1 , otherwise again the first term dominates in the first coordinate function and the second dominates in the second coordinate function, so altogether we find as principal terms

$$\begin{pmatrix} m^l t^p / 2 \\ m^{l-k} t^{p-q} / 4 \end{pmatrix}. \tag{3.21}$$

Thus the image of $L \begin{pmatrix} p \\ q \end{pmatrix}$ is the line $L \begin{pmatrix} p \\ p-q \end{pmatrix}$, and the point with coordinate m is mapped to the point of coordinate

$$x^{p-q} y^{-p} = 2^{p+q} m^{(p-q)l} t^{p(p-q)} m^{-p(l-k)} t^{-p^2+pq} = 2^{p+q} m^{-1}. \tag{3.22}$$

We see that the line $L \begin{pmatrix} p \\ q \end{pmatrix}$ is mapped to the line $L \begin{pmatrix} p \\ p-q \end{pmatrix}$ by an isomorphism.

The case of the line $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ (on the boundary of zones 2 and 3)

A computation just like the ones above shows that the parametrized curve

$$t \mapsto \begin{pmatrix} mt^2 \\ t \end{pmatrix}, \tag{3.23}$$

leading to the point with coordinate m on $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, maps to the curve with principal terms

$$t \mapsto \frac{1}{4mt^3} \begin{pmatrix} 2m^2 t^5 \\ 2mt^4 + m^2 t^4 \end{pmatrix} = \begin{pmatrix} mt^2/2 \\ (2+m)t/4 \end{pmatrix}, \tag{3.24}$$

with the two terms in the second coordinate in competition. Thus this line maps to itself, with the point of coordinate m mapping to the point of coordinate

$$xy^{-2} = \frac{mt^2}{2} \left(\frac{4}{(2+m)t} \right)^{-2} = \frac{8m}{(2+m)^2}. \tag{3.25}$$

This rational fraction plays an important role in the dynamics of Newton's method at infinity. It is dynamically quite simple, similar to the Chebychev polynomial $z^2 - 2$. The Julia set is a line segment, but instead of the complement being attracted to a superattractive fixed point, it is attracted to a linearly attracting fixed point.

More specifically, the critical points are at $m = \pm 2$. The one at -2 maps to infinity, which maps to 0, which is a repelling fixed point. The one at $m = 2$ is attracted to the attracting fixed point at $2(\sqrt{2} - 1)$. The Julia set is the negative real axis, and its complement is entirely attracted to the attracting fixed point.

This is the dynamics of the rational map, but doesn't quite jibe with the 2-dimensional dynamics of N . More specifically, the critical point $m = 2$ is on the critical cubic C_1 , and it maps to the point $m = 1$ on the critical value parabola of equation $x = y^2$. But the other critical point is much more troublesome. It is on the parabola of equation $x = -2y^2$, which maps to the x -axis, and the point $m = 0$ on $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ does not intersect the x -axis. As such, this point $m = -2$ is a point of

indeterminacy of the extended Newton map. We will need to blow it up again, and a lot of the complications of the Newton map are hidden in this point.

The case of zone 3. In zone 3, the map is especially simple. There the first term dominates in each coordinate, and the $4xy$ dominates in the coefficient, so the principal terms of the image of $\begin{pmatrix} x \\ y \end{pmatrix}$ are simply $\begin{pmatrix} x/2 \\ y/2 \end{pmatrix}$. Thus every line in zone 3 is mapped to itself by the identity. But notice the 2 's in the denominator; they indicate that in X_P , all these lines are repelling. Of course, we knew all about that in the case of $L\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, from Section 2.3.

The cases of zones 4 and 5 and of the points $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, are completely parallel to those already covered, and we only give the results.

The case of the line $L\begin{pmatrix} 2 \\ 1 \end{pmatrix}$. This line is mapped to itself, by the rational function $m \mapsto (2m+1)^2/(8m)$, obtained from the case of $L\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ by conjugating by $m \mapsto 1/m$.

The case of zone 4. The line $L\begin{pmatrix} p \\ q \end{pmatrix}$ is mapped to the line $L\begin{pmatrix} q-p \\ q \end{pmatrix}$ by the isomorphism $m \mapsto 2^{-(p+q)}m$.

The case of the line $L\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. This line is mapped to $L\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, by the rational map of degree 2 $m \mapsto (4m-1)/(4m^2)$.

The case of zone 5. Again there are two cases. If p is even, say $p = 2p_1$, then the image of $L\begin{pmatrix} p \\ q \end{pmatrix}$ is $L\begin{pmatrix} p \\ q+p_1 \end{pmatrix}$, and the mapping is given by $m \mapsto (-1)^{q_1}2^{-p}m^{-1}$. Again in this case the $\begin{pmatrix} p \\ q \end{pmatrix}$ -component is critical. If p is odd, the image of $L\begin{pmatrix} p \\ q \end{pmatrix}$ is $L\begin{pmatrix} 2p \\ 2q+p \end{pmatrix}$, and the mapping within this component is $m \mapsto -2^{-2p}m^{-2}$.

3.3. The asymptotic behavior of the Russakovskii-Shiffman measure

Although after all these blow-ups our Newton map still has points of indeterminacy, we have blown up enough to untangle much of the dynamics at infinity. Russakovskii and Shiffman [RS] have constructed a measure μ on \mathbb{P}^2 which describes the distribution of inverse images of points in the finite plane \mathbb{C}^2 . More specifically, they prove that if $f : \mathbb{P}^2 \rightsquigarrow \mathbb{P}^2$ is a rational mapping, whose topological degree d is larger than the algebraic degree, then there is a pluripolar exceptional set $E \subset \mathbb{P}^2$ such that if $x \in \mathbb{P}^2 - E$, then the sequence of probability measures

$$\mu_n = \frac{1}{d^n} \sum_{y \in f^{-1}(x)} \delta_y \quad (3.26)$$

converges to a limit independent of x .

The mapping considered here is one of their examples, but they could not determine whether this measure charges the points of indeterminacy \mathbf{p}_1 and \mathbf{p}_2 , though their computer experiments suggest that it does not. We will prove that this is indeed the case.

The difficulty is that near these points, Newton's method displays several different kinds of behavior, both attracting and repelling. The blow-ups we have performed in Sections 3.1 and 3.2

Clearly, we can choose K large enough that any finite number of these inverse images are well-behaved, but knowing that all the inverse images are well-behaved requires proof.

Lemma 3.3.1. *For K sufficiently large, each M_i cuts U_K into two pieces, one containing all the M_j , $j < i$, and the other containing all M_j , $j > i$.*

Proof. Choose K large enough that the statement is true for $i = -2, -1, 1, 2$. For all the others, near the component of D_0 adherent to M_i , the Newton mapping is either of the form

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} * \\ -x^2 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -y^2 \\ * \end{pmatrix}. \quad (3.27)$$

Thus in this region, U_K maps well inside itself, and is locally separated by M_{1-i} . It follows that the inverse image M_i separates U_K . \square

The manifolds M_i now allow us to build a ‘‘Markov partition’’ of U_K ; we label A_k the closures of the components of the complement of the M_i , ordered so that

A_i is bounded by M_i and M_{i+1} when $i > 0$,

A_i is bounded by M_i and M_{i-1} when $i < 0$,

A_0 is bounded by M_1 and M_{-1} .

Moreover, we define $A_k(C) \subset A_k$ to be the subset where $\| \begin{pmatrix} x \\ y \end{pmatrix} \| \leq C$; in particular, if $C < K$, then $A_k(C) = A_k(K) = A_k$.

Theorem 3.3.2. *There exist numbers $\lambda_k > 0$, $k \in \mathbb{Z}$, such that $\sum_{k \in \mathbb{Z}} \lambda_k < \infty$, and a cutoff $K > 0$, such that if we set*

$$\mu_k = \begin{cases} \lambda_k \frac{|dt|}{t^2} & \text{if } k \text{ is even} \\ \lambda_k \frac{|dt|}{t^{3/2}} & \text{if } k \text{ is odd,} \end{cases} \quad (3.28)$$

then whenever a probability measure ν on \mathbb{C}^2 satisfies

$$\nu(A_k(C)) \leq \int_C^\infty \mu_k(t) dt \quad (3.29)$$

for any $C > K$, the probability measure $\frac{1}{4}N^*\nu$ also satisfies this inequality.

Corollary 3.3.3. *The Russakovskii-Shiffman measure σ does not charge the points \mathbf{p}_1 and \mathbf{p}_2 .*

Proof of Corollary 3.3.3. The measure σ is obtained by giving weight 1 to a generic point (off an exceptional pluripolar set), then giving weight $1/4$ to its inverse images, then weight $1/4^2$ to its inverse images, etc. This defines a sequence of measures $\sigma_0, \sigma_1, \dots$; Shiffman and Russakovskii prove that the sequence converges.

Clearly we can choose σ_0 so that it satisfies Equation (3.29), for instance by choosing the initial point outside of U_K . Then all σ_n also satisfy (3.29), and so does σ . But for any $\epsilon > 0$, there exists K' such that any measure satisfying (3.29) assigns measure $< \epsilon$ to the region $U_{K'}$. \square

Proof of Theorem 3.3.2. The proof will take six pages, so before giving it we will outline what is really happening. The main content of the theorem is that on average, the initial points of indeterminacy \mathbf{p}_1 and \mathbf{p}_2 are attracting. The problem is that this isn't quite true: there are many

inverse images near $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, and these are repelling. But they are only repelling *linearly*, whereas elsewhere, more particularly near $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $L \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, the map is *superattracting*, at least with an exponent $3/2$. We chose our domains A_k carefully so that this would be the case; the point of having a domain like A_2 at all was to keep A_3 away from $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, which is again repelling. Then it is not surprising that the superattracting behavior wins out over the linearly repelling behavior; this fact is quite specifically illustrated by the $K^{1/4}$ in formula (3.38), which is really the key to why the proof works at all.

We need to choose the numbers λ_k so that the pull-back of the measure μ itself is at most four times itself. This is fairly easy for $|k| \geq 3$, and for all k even, but choosing $\lambda_{-3}, \lambda_{-1}, \lambda_1, \lambda_3$ is quite difficult: because of the points of indeterminacy, there are a lot of inverse images near these points.

We will find inequalities which express that the theorem is true in each A_k ; this is a different problem in each A_k . We will start with the hardest case, A_1 . Let ν be a probability measure on \mathbb{C}^2 such that $\nu(A_k(C)) \leq \int_C^\infty \mu_k(t) dt$ for all $k \in \mathbb{Z}$.

Lemma 3.3.4. *For any $\epsilon > 0$, there exists K such that when $C \geq K$, the measure $N^*\nu$ will satisfy*

$$\frac{1}{4}N^*\nu(A_1(C)) \leq \int_C^\infty \mu_1(t) dt \tag{3.30}$$

if

$$(2\sqrt{1-\epsilon} - \sqrt{2})\lambda_1 > \sqrt{2} \sum_{k=1}^\infty \lambda_{2k+1} + \frac{1}{\sqrt{K(1-\epsilon)}} \sum_{k=1}^\infty \lambda_{2k} + \frac{\sqrt{2K(1-\epsilon^2)}}{2}. \tag{3.31}$$

Proof. The region A_1 maps two-to-one to the union of all the $A_k, k > 1$, and in this region the x -coordinate of $N \begin{pmatrix} x \\ y \end{pmatrix}$ is $x/2 + o(|x|)$. In particular, if we choose K sufficiently large, we will have

$$\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_1 \right| \geq \frac{|x|(1-\epsilon)}{2}, \tag{3.32}$$

where the index 1 on the left-hand side means the first coordinate. Since $N(A_1(K))$ is strictly larger than $\cup_{k>0} A_k(K)$, we must allow for the possibility that the pull-back measure may pick up mass from outside U_K ; this mass is at most $1/2$, since ν is a probability measure, and a point has at most two of its inverse images in A_1 . We will represent this mass, rather wastefully, as a_1/\sqrt{C} , where a_1 must be chosen sufficiently large that

$$\frac{a_1}{\sqrt{2(K+\epsilon)}} \geq \frac{1}{2}, \quad \text{i.e.,} \quad a_1 = \frac{\sqrt{2K(1+\epsilon)}}{2}. \tag{3.33}$$

Thus we find

$$\begin{aligned}
\frac{1}{4}(N^*\nu)(A_1(C)) &\leq \frac{1}{2} \sum_{k=1}^{\infty} \nu \left(A_k \left(\frac{C(1-\epsilon)}{2} \right) \right) + \frac{a_1}{\sqrt{C}} \\
&\leq \frac{1}{2} \left(\sum_{k=0}^{\infty} \lambda_{2k+1} \int_{C(1-\epsilon)/2}^{\infty} \frac{dt}{t^{3/2}} + \sum_{k=1}^{\infty} \lambda_{2k} \int_{C(1-\epsilon)/2}^{\infty} \frac{dt}{t^2} \right) + \frac{a_1}{\sqrt{C}} \\
&= \sqrt{\frac{2}{C(1-\epsilon)}} \sum_{k=0}^{\infty} \lambda_{2k+1} + \frac{1}{C(1-\epsilon)} \sum_{k=1}^{\infty} \lambda_{2k} + \frac{a_1}{\sqrt{C}} \\
&\leq \sqrt{\frac{2}{C(1-\epsilon)}} \sum_{k=0}^{\infty} \lambda_{2k+1} + \frac{1}{(1-\epsilon)\sqrt{CK}} \sum_{k=1}^{\infty} \lambda_{2k} + \frac{a_1}{\sqrt{C}}
\end{aligned} \tag{3.34}$$

which will indeed be smaller than

$$\lambda_1 \int_C^{\infty} \frac{dt}{t^{3/2}} = \frac{2\lambda_1}{\sqrt{C}} \tag{3.35}$$

if

$$(2\sqrt{1-\epsilon} - \sqrt{2})\lambda_1 > \sqrt{2} \sum_{k=1}^{\infty} \lambda_{2k+1} + \frac{1}{\sqrt{K(1-\epsilon)}} \sum_{k=1}^{\infty} \lambda_{2k} + a_1\sqrt{1-\epsilon}. \quad \square \tag{3.36}$$

Now let us deal with with A_3 , which maps to A_{-1} and to a part of A_0 ; which part depends on K .

Lemma 3.3.5. *For any $\epsilon > 0$, there exists K such that the measure $N^*\nu$ will satisfy*

$$\frac{1}{4}(N^*\nu)(A_3(C)) \leq \mu_3(A_3(C)) \tag{3.37}$$

for $C \geq K$ if

$$\frac{\lambda_0}{K^{5/2}(1-\epsilon)^3} + \frac{\lambda_{-1}}{K^{1/4}(1-\epsilon)^{3/4}} \leq \lambda_3. \tag{3.38}$$

Proof. The region A_3 maps into $A_{-1} \cup A_0$; moreover, in the part of A_3 that maps into A_1 , the second coordinate of the image is $|x|^{3/2} + o(|x|^{3/2})$; see Figure 33. Thus for K sufficiently large we have

$$\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_2 \right| > |x|^{3/2}(1-\epsilon) \quad \text{when} \quad |x| \geq K. \tag{3.39}$$

(The index 2 on the left-hand side indicates the second coordinate.) This map is 2 to 1 to its image.

In the part of A_3 that maps to A_0 , which is a neighborhood of the polar hyperbola, the situation is better yet. For one thing, this part of A_3 maps to the part of A_0 where $|y| > |x|$; moreover, the second coordinate is $|x|^3/8 + o(|x|^3)$, so when K is sufficiently large we have

$$\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_2 \right| > \frac{|x|^3(1-\epsilon)}{8} \quad \text{when} \quad |x| \geq K. \tag{3.40}$$

This map is 1 to 1 to its image.

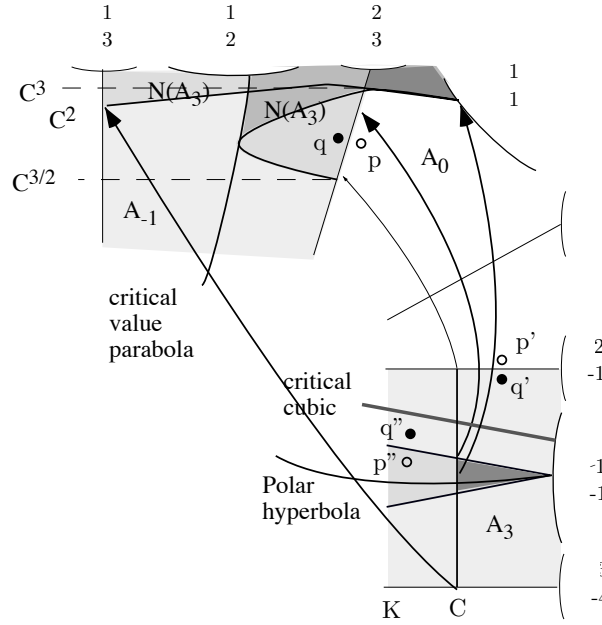


FIGURE 33. A schematic drawing of the mapping N in A_3 . To understand its behavior in the real, we must remember that one of the critical cubics passes through A_3 , and that the image of A_3 is bounded in part by the critical value parabola $y = x^2$. The part of the neighborhood of the polar hyperbola where $|x| > C$, shown in dark, maps into A_0 , also shown in dark. The points p, q and their inverse images p', p'', q', q'' are drawn to help understand how the mapping can be 2-to-1 outside the neighborhood of the polar hyperbola, and 1-to-1 inside it.

We can now argue as above:

$$\begin{aligned}
 \frac{1}{4}(N^*\nu)(A_3(C)) &= \frac{1}{2}\nu(N(A_3(C)) \cap A_{-1}) + \frac{1}{4}\nu(N(A_3(C)) \cap A_0) \\
 &\leq \frac{1}{2}\nu(A_{-1}(C^{3/2}(1-\epsilon)^{3/2})) + \frac{1}{4}\nu\left(N\left(A_0\left(\frac{C^3}{8}\right)\right)\right) \\
 &\leq \frac{\lambda_{-1}}{2} \int_{(C(1-\epsilon))^{3/2}}^{\infty} \frac{dt}{t^{3/2}} + \frac{\lambda_0}{4} \int_{\frac{(C(1-\epsilon))^3}{8}}^{\infty} \frac{dt}{t^2} \\
 &\leq \frac{2\lambda_{-1}}{(C(1-\epsilon))^{3/4}} + \frac{\lambda_0}{4} \frac{8}{(C(1-\epsilon))^3} \\
 &\leq \frac{2\lambda_{-1}}{K^{1/4}C^{1/2}(1-\epsilon)^{3/4}} + \frac{2\lambda_0}{K^{5/2}C^{1/2}(1-\epsilon)^3}
 \end{aligned} \tag{3.41}$$

which will indeed be smaller than

$$\lambda_3 \int_C^{\infty} \frac{dt}{t^{3/2}} = \frac{2\lambda_3}{C^{1/2}} \quad \text{if} \quad \frac{\lambda_0}{K^{5/2}(1-\epsilon)^3} + \frac{\lambda_{-1}}{K^{1/4}(1-\epsilon)^{3/4}} \leq \lambda_3. \quad \square \tag{3.42}$$

Of course, exactly the same results hold for λ_{-1} and λ_{-3} .

Now let us deal with λ_0 , $\lambda_{\pm 2}$, which is much easier.

Lemma 3.3.6. (a) For K sufficiently large, the inequality

$$\frac{1}{4}(N^*\nu)(A_0(C)) \leq \mu_0(A_0(C)) \quad (3.43)$$

is satisfied when $C \geq K$ if

$$\lambda_0 > \frac{K(1-\epsilon)^2}{1+2\epsilon} \quad (3.44)$$

(b) For any $\epsilon > 0$, there exists K such that the inequality

$$\frac{1}{4}N^*\nu(A_2(C)) \leq \mu_2(A_2(C)) \quad (3.45)$$

is satisfied for $C \geq K$ if

$$\lambda_2 > \frac{\lambda_0}{1-\epsilon} + K(1+\epsilon). \quad (3.46)$$

Proof. For part (a), observe that A_0 maps 1 to 1 to A_0 , and that when K is sufficiently large, we have

$$\frac{1}{2} \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\| (1+\epsilon) \geq \left\| N \begin{pmatrix} x \\ y \end{pmatrix} \right\| \geq \frac{1}{2} \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\| (1-\epsilon) \quad \text{when} \quad \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\| \geq K. \quad (3.47)$$

Again, we need to allow for the possibility that $N^*\nu$ picks up some mass from the mass of ν outside U_K . We will represent this mass as a_0/C , where $a_0 = K(1+\epsilon)/2$. Then for any $C < 2K(1+\epsilon)$, i.e., for any C such that $N(A_0(C))$ may fail to be contained in U_K , the term a_0/C is at least $1/4$, which is the maximum mass $N^*\nu/4$ can pick up, since ν is a probability measure.

Thus

$$\frac{1}{4}N^*\nu(A_0(C)) = \frac{1}{4}\nu\left(A_0\left(\frac{C(1-\epsilon)}{2}\right)\right) + \frac{a_0}{C} \leq \frac{\lambda_0}{4} \int_{C(1-\epsilon)/2}^{\infty} \frac{dt}{t^2} + \frac{a_0}{C} = \frac{\lambda_0}{2C(1-\epsilon)} + \frac{a_0}{C}, \quad (3.48)$$

which is indeed smaller than

$$\lambda_0 \int_C^{\infty} \frac{dt}{t^2} = \frac{\lambda_0}{C} \quad \text{when} \quad \lambda_0 > \frac{2a_0(1-\epsilon)}{1+2\epsilon}. \quad (3.49)$$

The second part is a bit more delicate, and requires cutting A_2 in two so that the part A'_2 with $|y| \leq 1/2 + o(1)$ maps to the part of A_0 with $|y| > |x|$, and the part A''_2 with $|y| \geq 1/2 + o(1)$ maps to the part of A_0 where $|x| > |y|$. In both cases, the map is 1-to-1, and in both the image spills out beyond U_K . We will represent the mass that $N^*\nu$ picks up from the mass of ν outside U_K by a_2/C , where a_2 is chosen so that

$$\frac{a_2}{4K(1+\epsilon)} \geq \frac{1}{4} \quad \text{and} \quad \frac{a_2}{2K(1+\epsilon)} \geq \frac{1}{4} \quad \text{i.e.,} \quad a_2 = K(1+\epsilon). \quad (3.50)$$

In A'_2 , we have $\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_2 \right| \geq \frac{|x|(1-\epsilon)}{4}$. Thus we find

$$\begin{aligned} \frac{1}{4}N^*\nu(A'_2(C)) &= \frac{1}{4}\nu\left(N(A'_2(C))\right) \leq \frac{1}{4}\nu\left(A_0\left(\frac{C(1-\epsilon)}{4}\right)\right) + \frac{a_2}{C} \\ &= \frac{\lambda_0}{4} \int_{C(1-\epsilon)/4}^{\infty} \frac{dt}{t^2} + \frac{a_2}{C} = \frac{\lambda_0}{C(1-\epsilon)} + \frac{a_2}{C}. \end{aligned} \quad (3.51)$$

This is smaller than

$$\lambda_2 \int_C^\infty \frac{dt}{t^2} = \frac{\lambda_2}{C} \quad \text{when} \quad \lambda_2 \geq \frac{\lambda_0}{1-\epsilon} + a_2. \quad (3.52)$$

In A_2'' , we have $\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_2 \right| \geq \frac{|x|(1-\epsilon)}{2}$. Thus we find

$$\begin{aligned} \frac{1}{4} N^* \nu(A_2'(C)) &= \frac{1}{4} \nu(N(A_2''(C))) \leq \frac{1}{4} \nu \left(A_0 \left(\frac{C(1-\epsilon)}{2} \right) \right) + \frac{a_2}{C} \\ &= \frac{\lambda_0}{4} \int_{C(1-\epsilon)/2}^\infty \frac{dt}{t^2} + \frac{a_2}{C} = \frac{\lambda_0}{2C(1-\epsilon)} + \frac{a_2}{C}. \end{aligned} \quad (3.53)$$

This is smaller than

$$\lambda_2 \int_C^\infty \frac{dt}{t^2} = \frac{\lambda_2}{C} \quad \text{when} \quad \lambda_2 \geq \frac{\lambda_0}{2(1-\epsilon)} + a_2, \quad (3.54)$$

a fortiori when

$$\lambda_2 \geq \frac{\lambda_0}{(1-\epsilon)} + a_2. \quad \square \quad (3.55)$$

Finally, we come to the conditions on the λ_k , $|k| > 3$. These present no difficulty: A_k maps well inside A_{2-k} , there is no spillover to worry about.

Lemma 3.3.7. *For any ϵ , there exists K such that the inequality*

$$\frac{1}{4} N^* \nu(A_k(C)) \leq \int_C^\infty \mu_k dt \quad (3.56)$$

is satisfied for $C \geq K$ if

$$\begin{cases} \lambda_k \geq \frac{\lambda_{2-k}}{\sqrt{K(1-\epsilon)}} & \text{if } k \text{ is odd} \\ \lambda_k \geq \frac{\lambda_{2-k}}{K(1-\epsilon)^2} & \text{if } k \text{ is even.} \end{cases} \quad (3.57)$$

Proof. In all A_k with $k > 3$, the significant term of the Newton map is $\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_2 \right| = |x|^2 + o(|x|^2)$; more precisely, there exists K such that if $|x| \geq K$, then

$$\left| \left(N \begin{pmatrix} x \\ y \end{pmatrix} \right)_2 \right| \geq (1-\epsilon)|x|^2. \quad (3.58)$$

Thus we find

$$\frac{1}{4} N^* \nu(A_k(C)) = \frac{1}{2} \nu(N(A_k(C))) \leq \frac{1}{2} \nu \left(A_{2-k} \left(C^2(1-\epsilon)^2 \right) \right). \quad (3.59)$$

We now study separately the cases k even and k odd.

If k is even, we have

$$\begin{aligned} \nu \left(A_{2-k} \left(C^2(1-\epsilon)^2 \right) \right) &\leq \lambda_{2-k} \int_{C^2(1-\epsilon)^2}^\infty \frac{dt}{t^2} \\ &= \frac{\lambda_{2-k}}{C^2(1-\epsilon)^2} \leq \frac{\lambda_{2-k}}{KC(1-\epsilon)^2}. \end{aligned} \quad (3.60)$$

This will be smaller than

$$\lambda_k \int_C^\infty \frac{dt}{t^2} = \frac{\lambda_k}{C} \quad \text{if} \quad \lambda_k \geq \frac{\lambda_{2-k}}{K(1-\epsilon)^2}. \quad (3.61)$$

If k is odd, we have

$$\begin{aligned} \nu \left(A_{2-k} \left(C^2(1-\epsilon)^2 \right) \right) &\leq \lambda_{2-k} \int_{C^2(1-\epsilon)^2}^\infty \frac{dt}{t^{3/2}} \\ &= \frac{2\lambda_{2-k}}{C(1-\epsilon)} \leq \frac{2\lambda_{2-k}}{\sqrt{CK}(1-\epsilon)}. \end{aligned} \quad (3.62)$$

This will be smaller than

$$\lambda_k \int_C^\infty \frac{dt}{t^{3/2}} = \frac{2\lambda_k}{\sqrt{C}} \quad \text{if} \quad \lambda_k \geq \frac{\lambda_{2-k}}{\sqrt{K}(1-\epsilon)}. \quad \square \quad (3.63)$$

We are now in a position to prove Theorem 3.3.2: to find K, ϵ and $\lambda_k, k \in \mathbb{Z}$ such that all the inequalities (3.31), (3.38), (3.44), (3.46), (3.57) are satisfied. Let us recopy them here:

$$(2\sqrt{1-\epsilon} - \sqrt{2})\lambda_1 > \sqrt{2} \sum_{k=1}^\infty \lambda_{2k+1} + \frac{1}{\sqrt{K}(1-\epsilon)} \sum_{k=1}^\infty \lambda_{2k} + \frac{\sqrt{2K(1-\epsilon^2)}}{2} \quad (3.31), \text{ again}$$

$$\lambda_3 \geq \frac{\lambda_0}{K^{5/2}(1-\epsilon)^3} + \frac{\lambda_{-1}}{K^{1/4}(1-\epsilon)^{3/4}} \quad (3.38), \text{ again}$$

$$\lambda_0 > \frac{K(1-\epsilon)^2}{1+2\epsilon} \quad (3.44), \text{ again}$$

$$\lambda_2 > \frac{\lambda_0}{1-\epsilon} + K(1+\epsilon) \quad (3.46), \text{ again}$$

$$\lambda_k \geq \begin{cases} \frac{\lambda_{2-k}}{\sqrt{K}(1-\epsilon)} & \text{if } k \text{ is odd, } |k| > 3 \\ \frac{\lambda_{2-k}}{K(1-\epsilon)^2} & \text{if } k \text{ is even, } |k| > 3. \end{cases} \quad (3.57), \text{ again}$$

Let us consider first the associated homogeneous inequalities:

$$(2\sqrt{1-\epsilon} - \sqrt{2})\lambda_1 > \sqrt{2} \sum_{k=1}^\infty \lambda_{2k+1} + \frac{1}{\sqrt{K}(1-\epsilon)} \sum_{k=1}^\infty \lambda_{2k} \quad (3.64)$$

$$\lambda_3 \geq \frac{\lambda_0}{K^{5/2}(1-\epsilon)^3} + \frac{\lambda_{-1}}{K^{1/4}(1-\epsilon)^{3/4}} \quad (3.65)$$

$$\lambda_0 > 0 \quad (3.66)$$

$$\lambda_2 > \frac{\lambda_0}{1-\epsilon} \quad (3.67)$$

$$\lambda_k \geq \begin{cases} \frac{\lambda_{2-k}}{\sqrt{K}(1-\epsilon)} & \text{if } k \text{ is odd, } |k| > 3 \\ \frac{\lambda_{2-k}}{K(1-\epsilon)^2} & \text{if } k \text{ is even, } |k| > 3. \end{cases} \quad (3.68)$$

These can be satisfied as follows: first choose $\epsilon > 0$ small (for instance $1/4$), and K so large that all the asymptotic inequalities for N are true for $\| \begin{pmatrix} x \\ y \end{pmatrix} \| > K$. Choose $\lambda_0 = 1$, then λ_2 so that the fourth inequality is true, then λ_3 so that the *strict* inequality

$$\lambda_3 > \frac{\lambda_0}{K^{5/2}(1-\epsilon)^3} \quad (3.69)$$

is satisfied. Now choose $\lambda_k, |k| > 3$ so that the fifth inequalities are satisfied, and so that the series in the first inequality is convergent. Now choose λ_1 so that the first inequality is true. Next, choose K larger yet so that the $K^{1/4}$ in the denominator in the second inequality allows it to be satisfied too.

We now have numbers that satisfy the homogeneous inequalities. Multiply all the λ 's by a sufficiently large constant so that all the inhomogeneous inequalities are satisfied also. \square

4

The Farey blow-up

In this chapter, we will pin down exactly what operation was performed when we constructed the primitive space. More specifically, we define a new way of “blowing up a surface” at a point (actually, an enriched point), which we call the *Farey blow-up*. The Farey blow-up has some very pleasant naturality properties, and is probably the only way of constructing a natural domain for a rational map with both points of indeterminacy and critical curves. We will proceed to compute the cohomology of the smooth part of the Farey blow-up, and describe the real oriented blow-up of the Farey blow-up along its exceptional divisor.

4.1. Definition of the Farey blow-up

Let X be a complex surface. An *enriched point* of X will be a point $\mathbf{x} \in X$ and two germs of smooth curves C_1, C_2 that intersect transversally at \mathbf{x} . Denote such an enriched point by $\tilde{\mathbf{x}} = (\mathbf{x}, C_1, C_2)$; the information \mathbf{x}, C_1, C_2 will be called *Farey data*.

We define the Farey blow-up $\mathcal{F}(X, \tilde{\mathbf{x}})$ of X at \mathbf{x} with respect to C_1, C_2 to be the pro-algebraic variety obtained as the projective limit of the surfaces X_n , with divisors $E_n \subset X_n$, where

$$X_0 = X, \quad E_0 = C_1 \cup C_2 \tag{4.1}$$

and X_{n+1} is X_n blown up at all the double points of E_n that project to \mathbf{x} .

The mappings of the projective system are the canonical projections $\pi_{n+1} : X_{n+1} \rightarrow X_n$.

The *Farey divisor* $\mathcal{F}(\tilde{\mathbf{x}}) \subset \mathcal{F}(X, \tilde{\mathbf{x}})$, which is the projective limit of the E_n , is a union of infinitely many irreducible components, all of which contain two non-algebraic points.

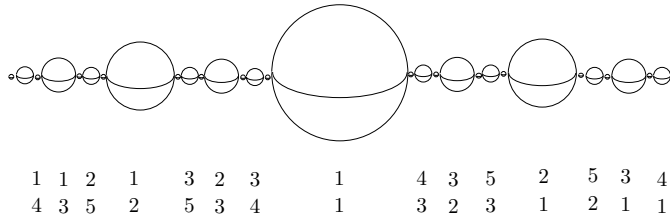


FIGURE 34. The Farey divisor is topologically a chain of beads: spheres that touch on a Cantor set.

Any two smooth curves that intersect at a point can be represented locally as $x = 0, y = 0$ in appropriate local coordinates, so Example 4.1.1 is a model for all Farey blow-ups.

Example 4.1.1. Let us compute the Farey blow-up of \mathbb{C}^2 at the origin, with respect to the x and the y axes; so let $\tilde{\mathbf{0}}$ be the Farey data consisting of the origin and the two axes.

In this case, the appropriate Farey tree is the obvious one, shown in Figure 35:

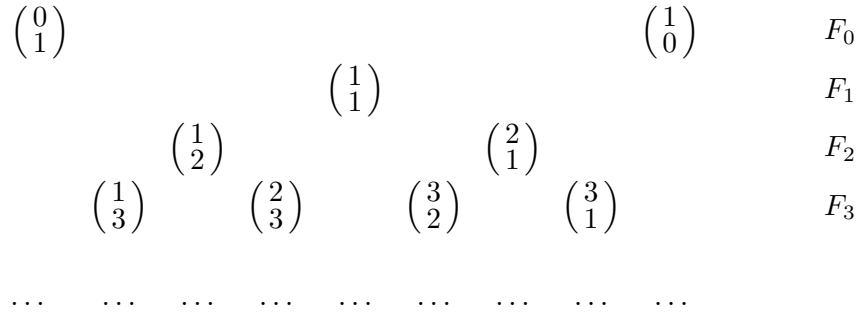


FIGURE 35. To get the entry between two entries, the corresponding coordinates are added; for example, the entry $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ on line F_2 is gotten by adding the entry $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ on line F_0 and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ on line F_1 ; the entry $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ on line F_3 is gotten by adding the “parents” $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ on line F_2 and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ on line F_0 .

Let us denote by F_n the vectors in the n th line (starting at F_0). Evidently the irreducible components of X_n can be labeled by the elements of F_n , as shown in Figure 36.

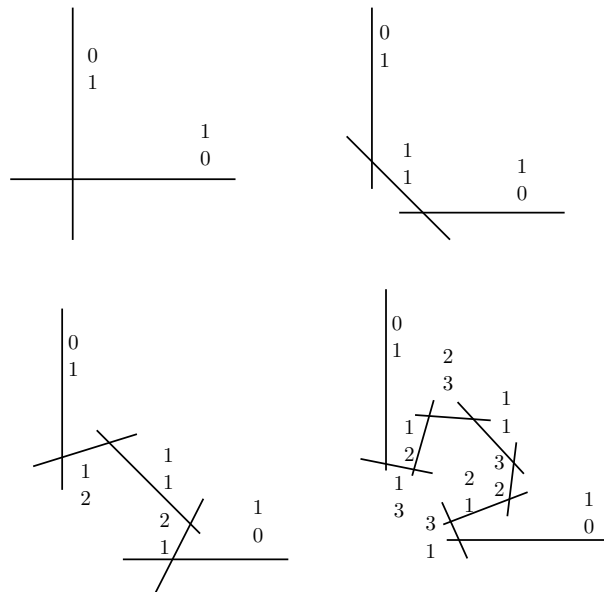


FIGURE 36. The first four stages in constructing the Farey blow-up $\mathcal{F}(\mathbb{C}^2, \vec{0})$. The components (actually rational lines) are represented as line segments, labeled by the pairs of integers appearing in the first four stages of the Farey tree.

Since p and q are coprime, the pair $\begin{pmatrix} p \\ q \end{pmatrix}$ is completely specified by the slope p/q of the ray through it. To lighten the notation, and also to make a distinction with a slightly different construction in Chapter 5, we will label the components of the Farey blow-up $L_{p/q}$.

There is a much deeper relation between a component $L_{p/q}$ of D_∞ and the vector $\begin{pmatrix} p \\ q \end{pmatrix}$ that labels it: as $t \rightarrow 0$, the parametrized curve $t \mapsto \begin{pmatrix} t^p \\ t^q \end{pmatrix}$ tends to a point of the component labeled by $\begin{pmatrix} p \\ q \end{pmatrix}$. As in Section 3.1, this statement can be sharpened.

Since p and q are coprime, there exist integers k and l such that $lq - kp = 1$. Then the limit as $t \rightarrow 0$ of the parametrized curve

$$t \mapsto (m^l t^p, m^k t^q) \quad (4.2)$$

exists in $\mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$ and these limits are all distinct for $m \neq 0$. The number m parametrizes the $\begin{pmatrix} p \\ q \end{pmatrix}$ component of $\mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$.

This provides a fairly complete description of the smooth points of $\mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$, but there are other points. We next give an embedding of the positive irrationals into the Farey blow-up. Given a positive irrational α , it lies between two consecutive elements $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$ and $\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}$ of F_n , in the sense that $p_1/q_1 < \alpha < p_2/q_2$. Set $\mathbf{x}_n(\alpha)$ to be the point

$$\mathbf{x}_n(\alpha) = E\left(\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}\right) \cap E\left(\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}\right). \quad (4.3)$$

Since $\pi_n(\mathbf{x}_n(\alpha)) = \mathbf{x}_{n-1}(\alpha)$, the sequence $\mathbf{x}_0(\alpha), \mathbf{x}_1(\alpha), \mathbf{x}_2(\alpha), \dots$ defines an element $\mathbf{x}(\alpha)$ of the Farey blow-up $\mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$. It should be clear that the map $\mathbb{R}_+ - \mathbb{Q} \rightarrow \mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$, $\alpha \mapsto \mathbf{x}(\alpha)$ is injective, since between any two distinct irrationals there are rationals.

Such points $\mathbf{x}(\alpha)$ are extremely singular. For instance, every neighborhood of such a point $\mathbf{x}(\alpha)$ has infinite-dimensional homology.

Notice that in analogy to the curves $t \mapsto \begin{pmatrix} t^p \\ t^q \end{pmatrix}$, the curve

$$t \mapsto \begin{pmatrix} t \\ t^\alpha \end{pmatrix} \quad (4.4)$$

tends to the point $\mathbf{x}(\alpha)$ of $\mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$.

We have now described almost all the points of $\mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$; the only ones missing are the points of the irreducible components corresponding to $m = 0$ and $m = \infty$. They are also at the end of curves: for instance, the curves

$$t \mapsto (t^p, t^q \log t) \quad \text{and} \quad t \mapsto (t^p \log t, t^q) \quad (4.5)$$

tend to such points.

There is quite a bit of freedom in the order in which the blow-ups involved in a Farey blow-up are performed.

Let X be an analytic surface, and $\tilde{\mathbf{x}} = (\mathbf{x}, C_1, C_2)$ be an enriched point. If you blow up first \mathbf{x} , then some of the double points created, then some of the double points created, etc., finitely many times, you will end up with a divisor D with double points on a surface Y . The components of D form a chain, each intersecting two others, with the first intersecting the proper transform of C_1 and the last intersecting C_2 ; the components can be labeled by pairs of integers as above.

Proposition 4.1.2. (a) *The pairs of integers appearing as successive labels on D are neighbors.*

(b) *The space obtained by performing Farey blow-ups at all the double points of D is precisely $\mathcal{F}(X, \tilde{\mathbf{x}})$.*

(c) *Conversely, the part of a Farey blow-up appearing between components with neighboring labels can be blown down, and the quotient point is a smooth point, where the images of the two components intersect transversally.*

Proof. Part (a) follows from the fact that the Farey sum of two neighbors is a neighbor to both its parents.

Part (b) is more or less obvious: any double point $\mathbf{y} = L_1 \cap L_2$ of D appears at some point in the construction of $\mathcal{F}(X, \tilde{\mathbf{x}})$, since all neighbors are adjacent in some F_n , and the further blow-ups that take place there to construct $\mathcal{F}(X, \tilde{\mathbf{x}})$ are precisely the ones involved in constructing $\mathcal{F}(Y, \tilde{\mathbf{y}})$, where $\tilde{\mathbf{y}}$ is \mathbf{y} enriched by L_1 and L_2 .

Part (c) is exactly undoing part (b). \square

Remark. You cannot blow down any segment of a Farey divisor between components whose labels are not neighbors without creating singularities. \triangle

4.2. Naturality of the Farey blow-up

The ordinary blow-up has bad naturality properties with respect to maps with critical points: the inverse images have to be taken in the sense of schemes, and this introduces complications which the Farey blow-up was designed to avoid.

Example 4.2.1. Consider the mapping $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^2 \\ y \end{pmatrix}$. If we blow up the origin both in the domain and the range, then f does not lift to a map $\tilde{f} : \tilde{\mathbb{C}} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \tilde{\mathbb{C}} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. On the contrary, the

exceptional divisor is collapsed to the point of the exceptional divisor corresponding to the critical locus, except for one point of indeterminacy, which corresponds to the kernel of $Df \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. We must blow up again at that point in order to get a well-defined map, and after blowing up, the new exceptional divisor is critical. A more natural thing is to blow up the subscheme defined by the equations $x^2 = 0, y = 0$. But that blow-up is a singular space, with a cone point; it is what one obtains if one blows down the first exceptional divisor in the first construction, which has self-intersection -2 after the second blow-up, and hence can be blown down, at the expense of creating a singular point.

Theorem 4.2.2. *Let X, Y be complex surfaces, and $f : X \rightarrow Y$ be a finite analytic map; we are particularly thinking of maps with critical curves. Suppose that $\tilde{\mathbf{z}}$ is an enriched point of Y , such that $f^{-1}(\tilde{\mathbf{z}})$ is set-theoretically a union of enriched points of X . Then f extends to a continuous mapping $\mathcal{F}(X, f^{-1}(\tilde{\mathbf{z}})) \rightarrow \mathcal{F}(\tilde{\mathbf{z}})$; this mapping is analytic on the smooth points of $\mathcal{F}(X, f^{-1}(\tilde{\mathbf{z}}))$.*

Proof. The statement is local on X , so we may assume that $f^{-1}(\mathbf{z})$ is a single point. The hypothesis that $f^{-1}(\tilde{\mathbf{z}})$ is set-theoretically a union of enriched points of X is very strong, and we

will see that we may choose coordinates in the domain and in the range so that the Farey data consists of the coordinate axes in both the domain and range.

We will show that near $f^{-1}(\mathbf{z})$ the map can be written

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^k \\ y^l \end{pmatrix} \quad (4.6)$$

for positive integers k, l .

Indeed, without loss of generality, we may assume that both \mathbf{z} and $f^{-1}(\mathbf{z})$ are the origin, and that both sets of Farey data are the axes. If $f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} g(x, y) \\ h(x, y) \end{pmatrix}$, then $g(x, y) = 0$ defines the y -axis, and $h(x, y)$ defines the y -axis, i.e., there exist integers $k, l > 0$ such that we can write $g(x, y) = x^k u_1(x, y)$ and $h(x, y) = y^l u_2(x, y)$ for some functions u_1, u_2 that do not vanish in a neighborhood of the origin. Let $v_1(x, y)$ and $v_2(x, y)$ be analytic functions such that $v_1^k = x^k u_1$ and $v_2^l = y^l u_2$. Clearly (v_1, v_2) form local coordinates in the range of f , and with respect to the coordinates (x, y) in the domain and v_1, v_2 in the range, the map has the desired form (4.6).

In this form, it is easy to define \tilde{f} explicitly on the Farey blow-up: if $t \mapsto \begin{pmatrix} at^p \\ bt^q \end{pmatrix}$ is a parametrized path leading to a point of the Farey blow-up (on the component at rational slope l/k), then its image is the parametrized curve $t \mapsto \begin{pmatrix} a^k t^{pk} \\ b^l t^{lq} \end{pmatrix}$, which tends to a point of the component corresponding to the primitive point on the line of slope $(q/p)(l/k)$. Set $d = \text{GCD}(pk, lq)$; a careful look at the formulas shows that the $\begin{pmatrix} p \\ q \end{pmatrix}$ component is critical if $d > 1$, mapping with normal degree d , and is a covering of its image component of degree kl/d , ramified (if $d < kl$) at the two non-algebraic points. \square

We will denote the mapping $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^k \\ y^l \end{pmatrix}$ by $f \begin{pmatrix} k \\ l \end{pmatrix}$. It is the basic example of a map then can be lifted to the Farey blow-up.

4.3. The real oriented blow-up of the Farey blow-up

The Farey blow-up is *a priori* a rather scary space. It contains an uncountable set of non-algebraic points, for which every neighborhood has infinite-dimensional homology, so it does not have the homotopy type of a CW complex. We will now show that the topological structure of the Farey blow-up is actually much simpler and understandable than one might fear, and actually in some sense very familiar.

Topologically, it is much easier to understand the Farey blow-up after “resolving its singularities,” using real oriented blow-ups as in [HPV]. The Farey blow-up has a real oriented blow-up that is reminiscent of the structures encountered in the Kolmogorov-Arnold-Moser Theorem.

To keep the paper self-contained, we repeat the definition of the real oriented blow-up, which is a device to “open a variety along a subvariety,” i.e., to speak of “the complement of the interior of a tubular neighborhood” without having to choose a tubular neighborhood.

The oriented blow-up \hat{X}_Z : step 1

Suppose X is a real algebraic manifold, and $Z \subset X$ is an algebraic subset, perhaps with singularities. Suppose $U \subset X$ is a coordinate patch in which the ideal $\mathcal{I}(Z)$ of functions vanishing on Z is generated by m functions $f_1, \dots, f_m : U \rightarrow \mathbb{R}$; in other words, $Z \cap U = f^{-1}(0)$, where $f : U \rightarrow \mathbb{R}^m$.

We need to say things in terms of ideals rather than in terms of equations defining the set because, for instance, the origin in \mathbb{R}^2 is defined by the equation $x^2 + y^2 = 0$, but the one function $x^2 + y^2$ generates a much smaller ideal than $\mathcal{I}(Z)$, which is generated by x and y .

We can then define $\hat{U}_Z \subset U \times S^{m-1}$ to be the closure of the set

$$\{(\mathbf{x}, p) \in U \times S^{m-1} \mid f(\mathbf{x}) \neq 0 \text{ and } f(\mathbf{x})/|f(\mathbf{x})| = p\}. \quad (4.7)$$

By contrast, the blow-up is the closure of the set

$$\{(\mathbf{x}, l) \in U \times \mathbb{P}^{m-1} \mid f(\mathbf{x}) \neq 0 \text{ and } f(\mathbf{x}) \in l\}; \quad (4.8)$$

the main difference is that we are now allowing for the orientability of the line $l \in \mathbb{P}^{m-1}$.

This actually defines the oriented blow-up only locally, but a standard result from algebraic geometry [Har] says that the blow-up has a universal property, which allows us to glue the local blow-ups together uniquely. This result goes over to the oriented context, as is shown in [HPV1], allowing us to define a space \hat{X}_Z locally isomorphic to the model above.

The real oriented blow-up, step 2

Already we run into trouble when Z is the union of the axes in \mathbb{R}^2 . The ideal $\mathcal{I}(Z)$ is generated by the single function xy , so the space $\hat{\mathbb{R}}^2_Z \subset \mathbb{R}^2 \times S^0$ is

$$(Q_1 \cup Q_3) \times \{1\} \cup (Q_2 \cup Q_4) \times \{-1\}, \quad (4.9)$$

where Q_i denotes the closed i th quadrant.

This isn't quite what we want: the plane cut along Z , i.e., the disjoint union of the four closed quadrants. We will replace the space \hat{X}_Z by a space $\mathcal{B}_{\mathbb{R}}^*(X, Z)$, which maps to \overline{X}_Z , and whose points above $\mathbf{x} \in \overline{X}_Z$ are the ends of $\overline{X} - Z$ at \mathbf{x} . This does not change the points with a basis of connected neighborhoods, but it does separate the first quadrant from the third and the second from the fourth, as the points of contact of these quadrants each correspond to two ends.

More formally, let X be any topological space, and $Y \subset X$ a closed subset. The endpoint modification $E(X, Y)$ is the space X , where every point $\mathbf{y} \in Y$ has been replaced by the set of ends of $X - Y$ at \mathbf{y} , i.e., by the points of the projective limit

$$\varprojlim \pi_0(V - Y), \quad (4.10)$$

where π_0 is the functor that associates to a space its set of connected components, and the projective limit is taken over all open neighborhoods $V \subset X$ of \mathbf{y} . The space $E(X, Y)$ comes with a natural topology, which we will leave to the reader to define (see [Dou], Vol. II, p.197–198), and there is a canonical map $p : E(X, Y) \rightarrow X$. We will denote by $E(Y) \subset E(X, Y)$ the subset $p^{-1}(Y)$.

Definition 4.3.1. Let X be a real algebraic manifold, and $Z \subset X$ an algebraic subset. The *real oriented blow-up* $\mathcal{B}_{\mathbb{R}}^*(X, Z)$ is the endpoint modification $E(\hat{X}_Z, \hat{Z})$, with the exceptional divisor $\mathcal{B}_{\mathbb{R}}^*(Z) \subset \mathcal{B}_{\mathbb{R}}^*(X, Z)$ being $E(\hat{Z}) \subset E(\hat{X}_Z, \hat{Z})$.

The real oriented blow-up of a Farey blow-up

Let X be a complex surface containing two curves C_1, C_2 intersecting transversally at $\mathbf{x} \in X$, and defining an enrichment $\tilde{\mathbf{x}}$ of \mathbf{x} . Let D be a divisor on X that may contain one, or both, or neither of C_1 and C_2 (it may be empty). If it contains neither, we will assume that $\mathbf{x} \notin D$.

We want to understand the relation between the real oriented blow-up $\mathcal{B}_{\mathbb{R}}^+(X, D)$ and the real oriented blow-up $\mathcal{B}_{\mathbb{R}}^+(\mathcal{F}(X, \tilde{\mathbf{x}}), \mathcal{F}(\tilde{\mathbf{x}}) \cup D)$.

To lighten notation, we will call this object $\mathcal{B}^+\mathcal{F}(X; \tilde{\mathbf{x}}, D)$. It isn't a priori clear what this is, and we must take the projective limits in the proper order. Recall that $\mathcal{F}(X, \tilde{\mathbf{x}})$ is naturally the pro-algebraic space

$$\lim_{\leftarrow} (X_n, \pi_n). \quad (4.11)$$

Each X_n contains a divisor E_n , which with our definition contains both C_1 and C_2 . Let us define D_n to be $E_n \cup D$, from which C_1 , or C_2 , or both have been removed if they do not belong to D . Here, removing a curve means taking the closure of the complement of the proper transform of that curve.

The real oriented blow-ups $\mathcal{B}_{\mathbb{R}}^+(X_n, D_n)$ form projective systems by Theorem 5.9 of [HPV]. We then define

$$\mathcal{B}^+\mathcal{F}(X; \tilde{\mathbf{x}}, D) = \lim_{\leftarrow} \mathcal{B}_{\mathbb{R}}^+(X, D_n). \quad (4.12)$$

To make a model for this object, we first define a way of taking the interval $(0, \infty)$, and ‘‘opening all the rational numbers to intervals.’’ To be more precise, consider the functions $\eta_1, \eta_2 : [0, \infty) \rightarrow \mathbb{R}$ given by

$$\eta_1(t) = \sum_{0 < p/q < t, (p,q)=1} e^{-\sup(p,q)} \quad \text{and} \quad \eta_2(t) = \sum_{0 < p/q \leq t, (p,q)=1} e^{-\sup(p,q)}. \quad (4.13)$$

Then the map $t \mapsto I_t = [\eta_1(t), \eta_2(t)]$ associates to every rational number a closed interval, and to every irrational number a single point. These intervals and points are all disjoint. The limit

$$\lim_{t \rightarrow \infty} \eta_1(t) = \lim_{t \rightarrow \infty} \eta_2(t) = a \quad (4.14)$$

is finite, so we may think of η_1 and η_2 as being defined on $[0, \infty]$, and giving a homeomorphism of ‘‘ $[0, \infty]$ with the positive rationals opened’’ to $J = [0, a]$.

We will put an appropriate ‘‘real oriented blow-up structure’’ on $J \times (\mathbb{R}/\mathbb{Z})^2$, by defining a circle action $*$ on the zones $I_{p/q} \times (\mathbb{R}/\mathbb{Z})^2$. If $t \in I_{p/q}$, we will let the circle \mathbb{R}/\mathbb{Z} act on $\{t\} \times (\mathbb{R}/\mathbb{Z})^2$ by

$$s * (t, (\alpha, \beta)) = (t, (\alpha + ps, \beta + qs)). \quad (4.15)$$

Because the integers (p, q) are coprime, this action is free, and the orbits of the action are (p, q) -torus knots. If p_n/q_n converges to an irrational number ω , then the orbits of the circle action converge to leaves of the irrational foliation of slope ω on the torus $\eta_1(\omega) \times (\mathbb{R}/\mathbb{Z})^2$.

For the remainder of this section, we will assume that C_1 and C_2 are contained in D .

Let us consider the real oriented blow-up $p : \mathcal{B}_{\mathbb{R}}^+(X, D) \rightarrow X$ of X along D . This is a 4-dimensional manifold with boundary; the part of the boundary above $C_1 - \{\mathbf{x}\}$ and above $C_2 - \{\mathbf{x}\}$ carries a circle action, and the fiber above \mathbf{x} is a torus T , which comes with an action of $(\mathbb{R}/\mathbb{Z})^2$, so that the limit of the circle action on the part above C_1 becomes the action of the first factor, and the limit of the circle action on the part above C_2 becomes the action of the second factor, as shown on the left of Figure 37.

Choose a neighborhood $U = U_1 \cup U_2$ of \mathbf{x} in $C_1 \cup C_2$; and identify $p^{-1}(U)$ with $J \times (\mathbb{R}/\mathbb{Z})^2$, so that on the part of boundary above ∂U_1 the action of the first factor of $(\mathbb{R}/\mathbb{Z})^2$ coincides with the circle action there, and so that on the part of boundary above ∂U_2 the action of the second factor of $(\mathbb{R}/\mathbb{Z})^2$ coincides with the circle action there.

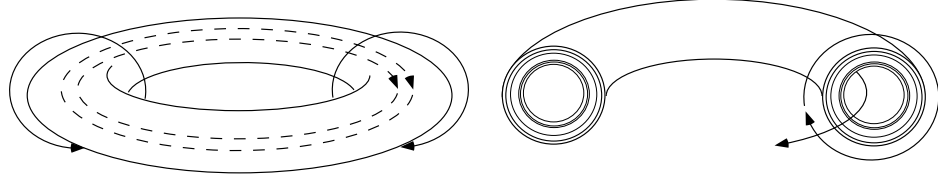


FIGURE 37. Left: a torus, with circle actions on the inside and the outside. Right: the torus now thickened into a Cantor set of concentric tori, with zones between them corresponding to the rational numbers. One should imagine these zones also filled with circle orbits, forming torus knots of all different kinds.

Now let us call Z the set $\mathcal{B}_{\mathbb{R}}^+(X, C_1 \cup C_2)$, with the original circle actions on $\mathcal{B}_{\mathbb{R}}^+(C_1 \cup C_2) - p^{-1}(U)$. Within $p^{-1}(U)$, put the circle actions on the parts corresponding to rational numbers, as described in Equation (4.15).

Theorem 4.3.2. *There is a homeomorphism*

$$h : Z \rightarrow \mathcal{B}^+ \mathcal{F}(X; \bar{\mathbf{x}}, D) \tag{4.16}$$

that is the identity off a neighborhood of the modified T , and which preserves the circle and torus actions.

Proof. Consider the intervals J_N , where all the components of $J - (\cup_{p/q \notin F_N} I_{p/q})$ have been collapsed to points. There is a natural projective system

$$\rightarrow J_N \rightarrow J_{N-1} \cdots \rightarrow J_0 = pt, \tag{4.17}$$

and the projective limit is precisely the interval J .

There are corresponding quotients $J_N \times (\mathbb{R}/\mathbb{Z})^2 \rightarrow J_{N-1} \times (\mathbb{R}/\mathbb{Z})^2$, and using the identification above, we can define spaces Z_N , together with maps forming a projective system

$$\cdots \rightarrow Z_N \rightarrow Z_{N-1} \rightarrow \cdots \rightarrow Z_0 \tag{4.18}$$

with projective limit Z .

So the theorem will follow from giving compatible homeomorphisms

$$\begin{array}{ccccccc} \cdots & \rightarrow & Z_N & \rightarrow & Z_{N-1} & \rightarrow & \cdots \\ & & \downarrow h_N & & \downarrow h_{N-1} & & \\ \cdots & \rightarrow & \mathcal{B}_{\mathbb{R}}^+(X_N, E_N) & \rightarrow & \mathcal{B}_{\mathbb{R}}^+(X_{N-1}, E_{N-1}) & \rightarrow & \cdots \end{array} \tag{4.19}$$

These will be constructed by induction. To start the induction, we choose a neighborhood U' of \mathbf{x} in X , containing U . Then we can find a mapping

$$\mathcal{B}_{\mathbb{R}}^+(X_0, E_0) \rightarrow \mathcal{B}_{\mathbb{R}}^+(X_0, E_0) \tag{4.20}$$

that collapses $p^{-1}(U_1 \cup U_2)$ to the torus $p^{-1}(\mathbf{x}) = (\mathbb{R}/\mathbb{Z})^2$, which is otherwise injective, and the identity outside of $p^{-1}(U')$. This defines a homeomorphism $\mathcal{B}_{\mathbb{R}}^+(X_0, E_0) \rightarrow Z_0$.

Now the construction of h_N from h_{N-1} is precisely what is done in [HPV], Theorem 5.9. \square

There is a straightforward modification of Proposition 4.3.2 in the case where D contains only one of C_1 or C_2 , or neither.

The cases where D contains either C_1 or C_2 are identical. If it contains C_2 , we start with $p : \mathcal{B}_{\mathbb{R}}^+(X, C_2) \rightarrow X$; the fiber $p^{-1}(\mathbf{x})$ is now a circle, and the first thing is to thicken it to a solid

torus, and to identify this solid torus with $D \times \mathbb{R}/\mathbb{Z}$, so that the \mathbb{R}/\mathbb{Z} -action on $\partial D \times \mathbb{R}/\mathbb{Z}$ coincides with the given circle action. Use polar coordinates (ρ, θ) in D , with $\rho \in J$. Then in the regions $\rho \in I_{p/q}$ we can define a circle action as above, and the resulting space is our candidate for a model of $\mathcal{B}^+ \mathcal{F}(X; \tilde{\mathbf{x}}, C_2)$. The main difference is that the fiber above the point where C_1 intersects E_∞ is now a circle, not a torus.

If D contains neither C_1 nor C_2 , the topological model consists of removing a small 4-ball around \mathbf{x} , bounded by a 3-sphere. Then E_1 is a rational curve, and $\mathcal{B}_{\mathbb{R}}^+(E_1) \subset \mathcal{B}_{\mathbb{R}}^+(X_1, E_1)$ maps to E_1 by a map conjugate to the Hopf fibration. Represent E_1 again in polar coordinates, where the radius, which now lives in $[0, \infty]$, should be thought of as belonging to J . Now repeat the construction above; the points where E_∞ intersects C_1 and C_2 both have circles above them, linking in $\mathcal{B}_{\mathbb{R}}^+(E_1)$.

4.4. Naturality and real oriented blow-ups

We saw in Section 4.2 that a map of the form $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^k \\ y^l \end{pmatrix}$ extends to the Farey blow-up $\mathcal{F}(\mathbb{C}^2, \tilde{\mathbf{0}})$, where $\tilde{\mathbf{0}}$ is the origin enriched by the two axes. Here we will see that the mapping also extends to the real oriented blow-ups of the Farey blow-ups, and that the mapping there is much nicer than one might expect: it is a covering map, and only ramified along one or two circles if we decide not to blow up the original enriching data.

Theorem 4.4.1. (a) *Let X, Y be complex surfaces, and $f : X \rightarrow Y$ be a finite analytic map. Suppose that $\tilde{\mathbf{z}}$ is an enriched point of Y , such that $f^{-1}(\tilde{\mathbf{z}})$ is set-theoretically a union of enriched points of X . Let D be a divisor in Y , which contains \mathbf{z} only if it contains C_1 or C_2 or both. Then f lifts to a continuous mapping*

$$\mathcal{B}^+ \mathcal{F}(X; f^{-1}(\tilde{\mathbf{z}}), f^{-1}(D)) \rightarrow \mathcal{B}^+ \mathcal{F}(Y; \tilde{\mathbf{z}}, D). \quad (4.21)$$

(b) *This map is a covering map if D contains both C_1 and C_2 . If C_1 is not in D , then the proper transform of C_1 is bounded by a circle in $\mathcal{B}^+ \mathcal{F}(Y; \tilde{\mathbf{z}}, D)$; the inverse image of that circle is critical, mapping to its image with the same local degree as the degree of f on C_1 . The analogous statement holds for C_2 . These are the only places where the map fails to be a local homeomorphism.*

Proof. The statement is local in the domain, so we can specialize to the case of Example 4.1.1, i.e., to the case where $f = f \begin{pmatrix} k \\ l \end{pmatrix}$.

There is no problem at the smooth points of the Farey divisor: if a component is critical of normal degree d , then circle orbits map to circle orbits with degree d . So let us consider a non-algebraic point, corresponding to ‘‘slope’’ r , in the range. If r is rational, it corresponds to either the upper or the lower end of a component; let us suppose the upper end if that happens. At each level of the Farey tree, the real number r is then bracketed between two rational numbers:

$$r'_N = \frac{p'_N}{q'_N} \leq r < \frac{p''_N}{q''_N} = r''_N. \quad (4.22)$$

The lower bracketing sequence will be eventually constant if r is rational.

A basis of neighborhoods of the non-algebraic point is then provided by the sets

$$\begin{aligned} \left| \frac{x}{y} \right|^{r'_N} < \frac{1}{\epsilon}, & \quad \left| \frac{x}{y} \right|^{r''_N} > \epsilon, \\ |\arg x - \lambda_1| < \epsilon, & \quad |\arg y - \lambda_2| < \epsilon, \quad |x|^2 + |y|^2 < \epsilon. \end{aligned} \quad (4.23)$$

The inverse image of such a neighborhood by the mapping $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^k \\ y^l \end{pmatrix}$ is evidently the union of kl disjoint sets, each of which contains such a set for a smaller ϵ . So the mapping on the real oriented blow-up is continuous, of constant degree kl . \square

4.5. Homology of the Farey blow-up

In this section we will see how the homology of the smooth part of a surface on which a Farey blow-up has been performed at an enriched point is related to the homology of the original surface.

Remark. It isn't clear what the homology of the Farey blow-up itself is (as opposed to its smooth points). This space has the sort of complication present in the Hawaiian earring, so we can expect singular and Čech homology to be different. In fact, it seems more reasonable to study the cohomology of $\mathcal{F}(X, \tilde{\mathbf{z}})$, but we will not consider this problem. Instead, we will develop a notion of completed homology, which is a Hilbert space and hence isomorphic to its dual, and which will serve as both homology and cohomology. \triangle

Let X is a complex surface, and $\tilde{\mathbf{z}}$ be an enriched point of X . We will denote by $\mathcal{F}^*(X, \tilde{\mathbf{z}}) \subset \mathcal{F}(X, \tilde{\mathbf{z}})$ the subset of smooth points. The space $\mathcal{F}(X, \tilde{\mathbf{z}})$ is a projective limit of spaces X_n with divisors E_n . Let us denote X_n^* the surface X_n from which the double points of E_n have been removed. There are then natural inclusions

$$X_0^* \subset X_1^* \subset X_2^* \subset \dots, \tag{4.24}$$

and the increasing union is $\mathcal{F}^*(X, \tilde{\mathbf{z}})$, which is thus represented as an inductive limit.

A first thing to realize (see Propositions 4.5.1 and 4.5.2) is that the homology is naturally a subgroup of

$$H_2(X) \oplus \mathbb{Z}^{\text{Irr } E_\infty} = H_2(X) \oplus \mathbb{Z}^{\mathbb{Q}_+}, \tag{4.25}$$

where the second factor is the product (not the sum) of copies of \mathbb{Z} , i.e., it consists of giving arbitrary integer weights to the components of E_∞ , which are naturally labeled by the positive rationals \mathbb{Q}_+ . Of course, it isn't the entire group, for instance because it is countable, and the product group is uncountable. But neither is it $H_2(X) \oplus \mathbb{Z}^{(\text{Irr } E_\infty)}$, where all but finitely many coefficients vanish.

For the next proposition, it is essential to realize that removing a point (or finitely many points) from a complex surface does not change its second homology group; in particular the inclusion $X_n^* \rightarrow X_n$ induces an isomorphism $H_2(X_n^*) = H_2(X_n)$. This has the consequence that the components $L_{p/q}$ for $\begin{pmatrix} p \\ q \end{pmatrix} \in F_n$ represent homology elements $[L_{p/q}]$ of $H_2(X_n^*)$.

Proposition 4.5.1. *For each n , the homomorphism*

$$H_2(X) \oplus \mathbb{Z}^{F_n} \rightarrow H_2(X_n^*), \quad \left(c, (\alpha_{p/q})_{p/q \in F_n} \right) \mapsto i_*c + \sum_{p/q \in F_n} \alpha_{p/q} [L_{p/q}] \tag{4.26}$$

is an isomorphism.

Proof. This is the content of Proposition 4.6 of [HPV]. \square

Thus any element of the homology is an element of the inductive limit

$$\varinjlim (H_2(X_n^*), i_*) = H_2(X) \oplus \varinjlim \mathbb{Z}^{F_n}. \tag{4.27}$$

Every element of this inductive limit is represented by a sequence $v_m \in \mathbb{Z}^{F_m}, v_{m+1} \in \mathbb{Z}^{F_{m+1}}, \dots$ starting with some m .

Proposition 4.5.2. *The weight assigned by v_k to elements of F_l with $l < k$ is independent of k .*

Proof. This follows from Proposition 4.8 of [HPV]. \square

Proposition 4.5.2 provides an embedding of $H_2(\mathcal{F}^*(X, \tilde{\mathbf{z}})) \rightarrow H_2(X) \oplus \mathbb{Z}^{\text{Irr } E_\infty}$, but we will describe the image much more precisely. Let Q be the first quadrant in \mathbb{R}^2 .

Definition 4.5.3. Let S be the commutative group of functions $\mathbf{f} : Q \rightarrow \mathbb{R}$ that

- are continuous;
- are piecewise-linear, linear except on finitely many rays through the origin with rational slope;
- take on integer values on $Q \cap \mathbb{Z}^2$; and
- vanish on the axes.

This group is a free commutative group on generators $\mathbf{e}_{p/q}, p/q \in \mathbb{Q}_+^*$, one for each positive rational number, defined as follows. Each rational number first appears in the Farey tree as the Farey sum

$$\frac{p}{q} = \frac{p_1 + p_2}{q_1 + q_2} \quad (4.28)$$

of two numbers that appeared earlier (its parents). Let $\mathbf{e}_{p/q} \in S$ be the function that is linear except on the rays of slope through $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix}$, and $\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}$, and which assigns 1 to $\begin{pmatrix} p \\ q \end{pmatrix}$ and 0 to

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}, \begin{pmatrix} p_2 \\ q_2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (4.29)$$

Lemma 4.5.4. *The elements $\mathbf{e}_{p/q}, p/q \in \mathbb{Q}_+^*$ are a basis of S .*

Proof. Consider the subgroup $S_n \subset S$ of functions that are linear except on rays passing through elements of F_n . This is a space of dimension $2^n - 1$, and the $2^n - 1$ functions

$$\mathbf{e}_{p/q}, \begin{pmatrix} p \\ q \end{pmatrix} \in F_n - F_0 \quad (4.30)$$

belong to it. We need to know that they span, or equivalently, that they are linearly independent. Indeed, if a linear combination is 0, the the coefficients corresponding to each element of $F_n - F_{n-1}$ must vanish, since otherwise the linear combination would not be linear at that element of $F_n - F_{n-1}$. But then, for the same reason, the coefficient of each element of $F_{n-1} - F_{n-2}$ must vanish, etc.

Thus the $\mathbf{e}_{p/q}, \begin{pmatrix} p \\ q \end{pmatrix} \in F_n$ are linearly independent in S_n , so that every element of S_n can be written

$$\sum_{\begin{pmatrix} p \\ q \end{pmatrix} \in F_n - F_0} a_{p/q} \mathbf{e}_{p/q}, \quad (4.31)$$

with rational coefficients $a_{p/q}$, and we need to show that these coefficients are all integers. Suppose not, and let $m \leq n$ be the smallest index for which there exists $\binom{p}{q} \in F_m$ with $a_{p/q}$ not an integer. This is a contradiction, since all the basis functions corresponding to indices $m' > m$ vanish at $\binom{p}{q}$, as do as the others of index m , and since all the earlier ones give integers at $\binom{p}{q}$.

Clearly together the $\mathbf{e}_{p/q}$ form a basis of $\lim_{\rightarrow} S_n = S$. \square

There is an evident isomorphism $S_n = \mathbb{Z}^{F_n}$, simply giving the values at the elements of F_n , and this isomorphism induces a canonical projection $S \rightarrow S_n$ given by restricting a function \mathbf{f} to the points of \mathbb{Z}^2 that are in F_n . There is also a natural inclusion $S_n \rightarrow S$, which consists of extending linearly to each sector bounded by rays through points of F_n . This map induces mappings $i_n : S_n \rightarrow S_{n+1}$.

Let X be a complex surface, and $\tilde{\mathbf{z}}$ be an enriched point of X . We define a homomorphism $\Phi : S \rightarrow \mathcal{F}^*(X, \tilde{\mathbf{z}})$ as follows. Choose $p/q \in \mathbb{Q}_+^*$, and suppose $\binom{p}{q} \in F_n - F_{n-1}$. There is then an exceptional divisor $L_{p/q} \subset X_n$ labeled $\binom{p}{q}$. Let $\phi_{p/q} : \mathbb{P}^1 \rightarrow L_{p/q}$ be an isomorphism, and let $\phi'_{p/q} : \mathbb{P}^1 \rightarrow X_n^*$ be a small deformation whose image avoids the double points.

Remark. You can only deform differentiably: analytically, exceptional divisors are rigid. The deformation can take place in an arbitrarily small neighborhood of $L_{p/q}$, but the deformed 2-sphere must intersect $L_{p/q}$ since this curve has self-intersection -1 . \triangle

Then $\phi'_{p/q} : \mathbb{P}^1 \rightarrow \mathcal{F}^*(X, \tilde{\mathbf{z}})$ maps the fundamental class in $H_2(\mathbb{P}^1)$ to a well-defined element $\Phi(\mathbf{e}_{p/q}) \in H_2(\mathcal{F}^*(X, \tilde{\mathbf{z}}))$ (this is the definition of Φ). For Theorem 4.5.5, we need one more generality: the inclusion $X - \{\mathbf{z}\} \rightarrow X$ induces an isomorphism $H_2(X - \{\mathbf{z}\}) \rightarrow H_2(X)$; using its inverse we see that there is a canonical map $H_2(X) \rightarrow H_2(\mathcal{F}^*(X, \tilde{\mathbf{z}}))$.

Theorem 4.5.5. *The mapping Φ and the canonical map $H_2(X) \rightarrow H_2(\mathcal{F}^*(X, \tilde{\mathbf{z}}))$ induce an isomorphism*

$$H_2(X) \oplus S \rightarrow H_2(\mathcal{F}^*(X, \tilde{\mathbf{z}})), \tag{4.32}$$

which we will still call Φ .

Proof. The space $\mathcal{F}^*(X, \tilde{\mathbf{z}})$ is the inductive limit of the X_n^* , and since homology commutes with inductive limits, we see that what we need to prove is that for every n , the map

$$S_n \oplus H_2(X) \rightarrow H_2(X_n^*) \tag{4.33}$$

is an isomorphism, and that the diagrams

$$\begin{array}{ccc} H_2(X) \oplus S_n & \rightarrow & H_2(X_n^*) \\ \downarrow & & \downarrow \\ H_2(X) \oplus S_{n+1} & \rightarrow & H_2(X_{n+1}^*) \end{array} \tag{4.34}$$

all commute.

This is proved by induction, starting with $n = 0$, where it is the isomorphism $H_2(X) = H_2(X - \{\mathbf{z}\})$, and going from step to step using Proposition 4.6 of [HPV]. \square

Give S a quadratic form, by making distinct $\mathbf{e}_{p/q}$ orthogonal, and setting $\mathbf{e}_{p/q}^2 = -1$. The spaces $H_2(X)$ and $\mathcal{F}^*(X, \tilde{\mathbf{z}})$ both carry quadratic forms defined by intersection, and these are all related by the following statement.

Theorem 4.5.6. *If we take $H_2(X) \oplus S$ to be an orthogonal direct sum, then the isomorphism*

$$\Phi : H_2(X) \oplus S \rightarrow H_2(\mathcal{F}^*(X, \tilde{\mathbf{z}})). \quad (4.35)$$

is an isomorphism of spaces with inner products.

Proof. Begin by representing all elements of $H_2(X)$ by cycles that avoid a neighborhood of \mathbf{z} ; that is really the meaning of the isomorphism $H_2(X) = H_2(X - \{\mathbf{z}\})$. Then our first exceptional divisor $L_{1/1}$ is orthogonal to all of $H_2(X)$, and has self-intersection -1 , being an exceptional divisor. Now the statement should be clear from the way Φ was defined. Every homology class, when it is created, is disjoint from all the previous ones, and has self-intersection -1 . Of course, the same thing is true of $\phi'_{p/q}(\mathbb{P}^1)$, thus none of these classes intersect any other class, and they all have self-intersection -1 . \square

4.6. The action of mappings $f \binom{2}{1}$ on homology

The next task is to see how the maps

$$\tilde{f} \binom{k}{l} : \mathcal{F}^*(\mathbb{C}^2, \tilde{\mathbf{0}}) \rightarrow \mathcal{F}^*(\mathbb{C}^2, \tilde{\mathbf{0}}) \quad (4.36)$$

act on the homology. We will work this out only for the mapping $f \binom{2}{1}$, which we will denote simply by f for the rest of this chapter. This turns out quite nicely, but a general proof along the lines of the one given here would be very cumbersome.

Let us denote $f_Q : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ the map

$$f_Q : \binom{p}{q} = \binom{2p}{q}. \quad (4.37)$$

Theorem 4.6.1. *The map induced by f on the homology is*

$$\tilde{f}_* (\mathbf{e}_{p/q}) = 2\mathbf{e}_{p/q} \circ f_Q^{-1}. \quad (4.38)$$

Proof. The idea is to work at a level of blow-up where $\mathbf{e}_{p/q}$ has just appeared. Let the Farey parents of $\binom{p}{q}$ be $\binom{p_1}{q_1}$ and $\binom{p_2}{q_2}$, and assume that q_1 is even and q_2 is odd. (The case q_2 even, q_1 odd is identical, and q_1 and q_2 can't both be even; we will deal with the case where both are odd later.) Observe that $\binom{p_1}{q_1/2}$ and $\binom{2p_2}{q_2}$ are then neighbors, and we can consider the sequence of two blow-ups in the domain and range represented in Figure 38.

Note that the map f extends at the top and the bottom level, but not at the middle level, where the divisor labeled $\binom{p_1 + p_2}{q_1 + q_2}$ is collapsed to the point marked with a circle, except for one point of indeterminacy, marked with a star.

The class $\mathbf{e}_{p/q}$ is represented by the line $L_{p/q}$ after one blow-up, and by $L_{p/q} + L_{(p_1+2p_2)/(q_1+q_2)}$ after the second blow-up. This maps to

$$2L_{(2p_1+2p_2)/(q_1+q_2)} + L_{(p_1+2p_2)/(q_1/2+q_2)}. \quad (4.39)$$

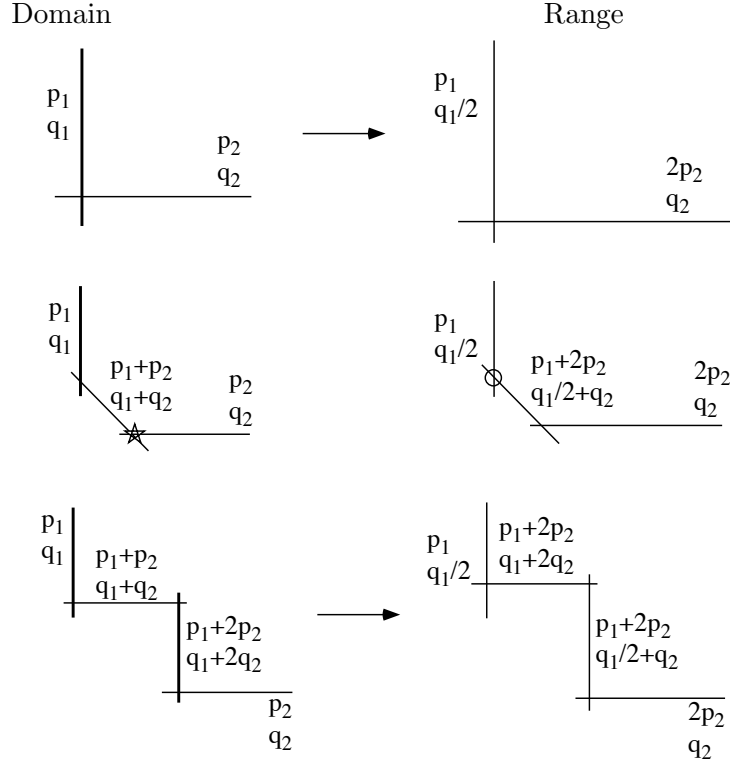


FIGURE 38. The map f can be locally extended before doing any blow-ups (i.e., at the top level), and after doing two blow-ups (i.e., at the bottom level), but it cannot be extended at the middle level (after just one blow-up).

On the other hand, the class $\mathbf{e}_{(p_1+2p_2)/(q_1/2+q_2)}$ is represented by $L_{(p_1+2p_2)/(q_1/2+q_2)}$ after one blow-up in the range, and by

$$L_{(2p_1+2p_2)/(q_1+q_2)} + L_{(p_1+2p_2)/(q_1/2+q_2)} \tag{4.40}$$

after two blow-ups.

Thus the image of $\mathbf{e}_{p/q}$ is represented by

$$\mathbf{e}_{(p_1+2p_2)/(q_1/2+q_2)} + L_{(2p_1+2p_2)/(q_1+q_2)} = \mathbf{e}_{(p_1+2p_2)/(q_1/2+q_2)} + \mathbf{e}_{(2p_1+2p_2)/(q_1+q_2)}, \tag{4.41}$$

since $L_{(2p_1+2p_2)/(q_1+q_2)}$ is the last exceptional divisor.

This is a formula for the action of $\tilde{f} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ on homology (at least in the unramified case), but it isn't quite clear that it is what the statement promised. Before dealing with that, let us find the analogous formula if both q_1 and q_2 are odd.

This is covered by Figure 38 also: both parents of $\begin{pmatrix} p_1 + 2p_2 \\ q_1 + 2q_2 \end{pmatrix}$ have odd second coordinates; we will leave it to the reader to see that all lines with such labels appear in this setting. Here $L_{(p_1+2p_2)/(q_1+2q_2)}$ is an exceptional divisor created by the most recent blow-up, so it represents (after two blow-ups) the class $\mathbf{e}_{(p_1+2p_2)/(q_1+2q_2)}$, and it maps to

$$L_{(p_1+2p_2)/(q_1/2+q_2)} = \mathbf{e}_{(p_1+2p_2)/(q_1/2+q_2)} - \mathbf{e}_{(2p_1+2p_2)/(q_1+q_2)}. \tag{4.42}$$

Now we need to understand these elements of the commutative group S as functions (see Definition 4.5.3). We will restrict our functions to the line $q = 1$ to draw them. The functions $\mathbf{e}_{p/q}$

restrict to tent functions, of course, but the miracle that makes things work is that if $\left(\frac{p'}{q'}\right)$ is a child of $\left(\frac{p}{q}\right)$, then $\mathbf{e}_{p/q} \pm \mathbf{e}_{p'/q'}$ is also a tent function; the slopes are exactly what is needed for this to happen. Thus Figure 39 explains why Theorem 4.6.1 is true.

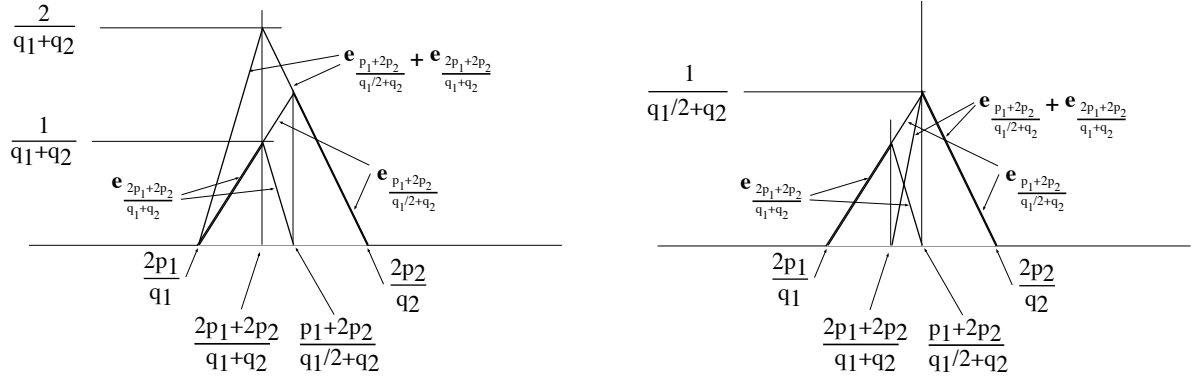


FIGURE 39. Left: the sum $\mathbf{e}_{(p_1+2p_2)/(q_1/2+q_2)} + \mathbf{e}_{(2p_1+2p_2)/(q_1+q_2)}$ is the tent function

$$2\mathbf{e}_{(p_1+p_2)/(q_1+q_2)} \circ f_Q^{-1}.$$

Right: the difference $\mathbf{e}_{(p_1+2p_2)/(q_1/2+q_2)} - \mathbf{e}_{(2p_1+2p_2)/(q_1+q_2)}$ is the tent function

$$2(\mathbf{e}_{(p_1+2p_2)/(q_1/2+q_2)}) \circ f_Q^{-1}.$$

4.7. Homology spaces and Sobolev spaces

We defined the commutative group S in Definition 4.5.3. The vector space $S \otimes_{\mathbb{Z}} \mathbb{C}$ is evidently a (negative definite) pre-Hilbert space, and it is very tempting to complete it with respect to its inner product. This creates the completed homology space

$$\hat{H}_2(\mathcal{F}^*(\mathbb{C}_2, \tilde{\mathbf{0}})) = l_2(\mathbb{Q}_+^*), \quad (4.43)$$

i.e., the space of series

$$\sum_{p/q \in \mathbb{Q}_+^*} a_{p/q} \mathbf{e}_{p/q} \quad \text{with} \quad \sum |a_{p/q}|^2 < \infty. \quad (4.44)$$

This is already a fairly familiar space, but quite surprisingly, it is also naturally an old friend from analysis, the Sobolev space $\mathcal{H}_0^1(I)$ of continuous functions on \mathbf{g} on $[0, 1]$, whose distributional derivative is in L_2 , and such that $\mathbf{g}(0) = \mathbf{g}(1) = 0$. We use a script \mathcal{H} to avoid confusion with homology; the subscript 0 indicates that the function vanishes at the endpoints.

Remark. In one dimension, the functions in \mathcal{H}^1 are automatically continuous, so it is reasonable to evaluate them at points. \triangle

In fact, our tent functions $\mathbf{e}_{p/q}$ form a basis of $\mathcal{H}_0^1(I)$ very similar to the basis provided by the integrals of the Haar functions, where instead of the dyadic partition of the interval, we use the Farey partition.

Let I be the interval joining $\left(\frac{1}{0}\right)$ to $\left(\frac{0}{1}\right)$ in the (p, q) -plane. We will parametrize I by $t \mapsto \left(\frac{t}{1-t}\right)$, and give it the measure $|dt|$, so it has length 1. The restriction to I maps S to a space of functions on I . For future reference, we will specify what the image is.

Lemma 4.7.1. *The image of S is the space of continuous, piecewise-linear functions $\mathbf{g} : I \rightarrow \mathbb{R}$ that vanish at the endpoints, are linear except at finitely many points with rational coordinates, and satisfy the integrality condition that for all positive rational numbers p/q , we have*

$$(p + q)\mathbf{g}(p/q) \in \mathbb{Z}. \tag{4.45}$$

Proof. This should be clear from Lemma 4.5.4: if you extend a function \mathbf{g} as above to be homogeneous of degree 1, you will find a continuous, piecewise linear function on the quadrant q , and its value at $\binom{p}{q}$ will be $(p + q)\mathbf{g}(p/q)$. \square

Theorem 4.7.2. *The map $\mathbf{f} \mapsto \mathbf{f}|_I$ induces an isometry of Hilbert spaces*

$$\Psi : \hat{S} \rightarrow \mathcal{H}_0^1(I). \tag{4.46}$$

Proof. Notice that the functions $\mathbf{f}|_I$ are tent functions. First let us see that $\Psi(\mathbf{e}_{p/q})$ and $\Psi(\mathbf{e}_{p'/q'})$ are orthogonal if $p/q \neq p'/q'$. By their definition, either the supports of $\Psi(\mathbf{e}_{p/q})$ and $\Psi(\mathbf{e}_{p'/q'})$ are disjoint, or the support of one, say $\Psi(\mathbf{e}_{p/q})$, is contained in an interval where the other is linear, hence its derivative is constant. Thus what we need to show is that

$$\int_I \frac{d}{dt} \mathbf{e}_{p/q} \left(\binom{t}{1-t} \right) |dt| = 0. \tag{4.47}$$

(The function $\mathbf{e}_{p/q} \left(\binom{t}{1-t} \right)$ is represented in Figure 40.)

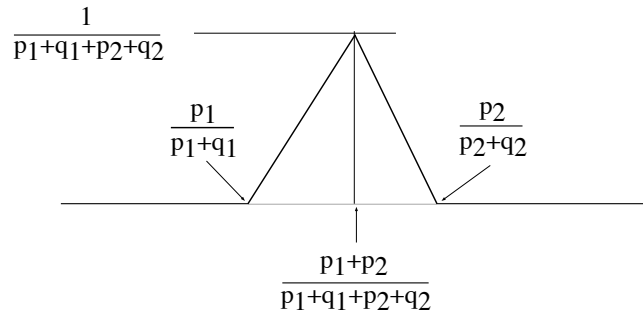


FIGURE 40. The function $\mathbf{e}_{p/q} \left(\binom{t}{1-t} \right)$.

This is clear from

$$\int_I \frac{d}{dt} \mathbf{e}_{p/q} \left(\binom{t}{1-t} \right) |dt| = \mathbf{e}_{p/q} \left(\binom{1}{0} \right) - \mathbf{e}_{p/q} \left(\binom{0}{1} \right) = 0 - 0 = 0. \tag{4.48}$$

Next, we need to show that

$$\int_I \left(\frac{d}{dt} \mathbf{e}_{p/q} \left(\binom{t}{1-t} \right) \right)^2 |dt| = 1. \tag{4.49}$$

Computing the integral of the square of the derivative of this function is elementary algebra, which we leave to the reader. We must use $q_1 p_2 - p_1 q_2 = 1$, of course, to make the integral come out to be 1.

This shows that Ψ is an isometric inclusion. Now we need to show that the image of S is dense, but that is fairly clear from Lemma 4.5.4: the derivatives of elements of S give all step-functions on I with rational discontinuities and integral 0. These are certainly dense in the hyperplane in $L^2(I)$ made up of functions with integral 0; for instance, they include the Haar basis. \square

There is quite a bit of freedom in this description, corresponding to the freedom in Proposition 4.1.2; we will need only the following special case.

Proposition 4.7.3. (a) *If you draw the vertical ray R_c defined by $p = c$, $q > 0$ in Q , and give it the measure $|dq|/c$, then the restriction of elements of S to R_c is an isometric inclusion of S into $\mathcal{H}^1(R_c)$, and the image of $S \otimes_{\mathbb{Z}} \mathbb{C}$ is dense in the subspace of $\mathcal{H}^1(R_c)$ of functions that vanish at the endpoint $\begin{pmatrix} 0 \\ c \end{pmatrix}$ of R_c .*

(b) *Consider the mapping $g : R_c \rightarrow R_c$ given by $g(q) = q/2$. Then*

$$(f_Q)_* \mathbf{f}|_{R_c} = g_* \mathbf{f}|_{R_c}. \quad (4.50)$$

Proof. We need to show that the restrictions of the $\mathbf{e}_{p/q}$ are an orthonormal family; this is a computation just like the one to justify Equation (4.49), and we omit it. The density follows, for instance, from the fact that S contains the Haar functions. \square

5

The compactification when $a = b = 0$

In this chapter, we will construct a compact space X_∞ on which the Newton map N of Chapter 3 is well defined. We will further analyze the topology of this space, showing that its behavior at infinity is very reminiscent of the KAM theorem [Ar].

5.1. The tower of blow-ups when $a = b = 0$

We will now define a projective system of pro-algebraic spaces X_0, X_1, \dots . The spaces X_0 and X_1 have already been defined: they are respectively $X_0 = \mathbb{P}^2$ and $X_1 = X_P$, the primitive space of Chapter 3. These spaces come with finite subsets $Z_i \subset X_i$, entirely made up of smooth points. We will enrich the structure for some of these points; i.e., for some of the $\mathbf{z} \in Z_i$, we will define two smooth germs of algebraic curves $C_1, C_2 \subset X_i$ intersecting transversally at \mathbf{z} . We may also need to *enhance* some other points. This will mean adding enough extra information at those points to define some finite sequence of blow-ups at these points.

This enhanced or enriched structure is denoted by \tilde{Z}_i , and we will define $X_{i+1} = \mathcal{F}(X_i, \tilde{Z}_i)$, where in this context \mathcal{F} denotes making ordinary blow-ups at ordinary points, Farey blow-ups at enriched points, and the appropriate finite sequence of blow-ups at the enhanced points.

Thus these spaces will naturally come with “blow-up maps” $\pi_{i+1} : X_{i+1} \rightarrow X_i$. We need to keep track of the *divisor at infinity* D_n , defined inductively by setting D_0 to be the line at infinity in \mathbb{P}^2 , and $D_{n+1} = \pi_{n+1}^{-1}(D_n)$. We will also need to keep track of the exceptional divisors corresponding to the ordinary and enhanced points or enriched points; we will do this by setting $E_1 = \{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ and $E_{n+1} = \pi_{n+1}^{-1}(E_n)$. (These points $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ are the points already seen in Equation (1.27); they are unrelated to the q ’s of Chapter 4.)

We will also construct maps $N_{i+1} : X_{i+1} \rightarrow X_i$ for $i \geq 1$, which will extend in an appropriate sense the Newton map N of Chapter 3.

Building X_1 and X_2 requires special constructions; after that everything has a unified definition.

Set $X_0 = \mathbb{P}^2$, with Z_0 consisting of the points \mathbf{p}_1 and \mathbf{p}_2 . The set \tilde{Z}_0 adds to this structure the enrichment of \mathbf{p}_1 by the x -axis and the line at infinity, and the enrichment of \mathbf{p}_2 by the y -axis and the line at infinity.

Now define $X_1 = \mathcal{F}(X_0, \tilde{Z}_0)$; this is exactly our primitive space X_P . The Newton map N extends to a rational map $N_P : X_1 \rightsquigarrow X_1$ with four points of indeterminacy at infinity, one each on the lines

$$L \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, L \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, L \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, L \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}. \quad (5.1)$$

Here we are denoting components of D_1 with vector indices, not fractions, because the primitive points of Q_P are not quite specified by their slopes.

We will meet these points again, so we give them names: A_1, B_1, B_2, A_2 .

- The polar hyperbola $4xy = 1$ and the component $L \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ will enrich A_1 , giving \tilde{A}_1 ;

- The parabola $x = -2y^2$ and the component $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ will enrich B_1 , giving \tilde{B}_1 ;
- The parabola $y = -2x^2$ and the component $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ will enrich B_2 , giving \tilde{B}_2 ;
- The polar hyperbola $4xy = 1$ and the component $L \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ will enrich A_2 , giving \tilde{A}_2 .

We set \tilde{Z}_1 to be the union of these four enriched points and the three ordinary points $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$, and define $X_2 = \mathcal{F}(X_1, \tilde{Z}_1)$. It is not quite so easy to define $N_2 : X_2 \rightarrow X_1$.

Proposition 5.1.1. *The Newton map N extends to a mapping $N_2 : X_2 \rightarrow X_1$.*

Proof. There is no difficulty on E_2 , the blow-ups of $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$; this was done in Lemma 1.2.4. The difficulty is to show that N extends to the Farey divisors corresponding to the enriched points $\tilde{A}_1, \tilde{A}_2, \tilde{B}_1, \tilde{B}_2$. In all cases, this follows from Theorem 4.2.2, but one must make some preliminary constructions to see how.

Begin by performing just three blow-ups on \mathbb{P}^2 : one at \mathbf{p}_1 , to create $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; a second at the intersection $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with the line at infinity (here called $L \begin{pmatrix} 1 \\ 1 \end{pmatrix}$), to create the divisor $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$; and finally, a third at the point of coordinate $m = -2$ on $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, to create a line we will temporarily call simply L . Call the surface constructed Y . We are now in a position to apply Theorem 4.2.2.

Indeed, the Newton map extends to a neighborhood of the union $L \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cup L \begin{pmatrix} 2 \\ 1 \end{pmatrix} \cup L$, mapping $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ to itself by the rational function given by Equation (3.25). The line L is critical, mapping to $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and finally $L \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ maps to the line at infinity.

In particular,

- the point $\tilde{\mathbf{x}}_1$ where L meets $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, enriched by L and $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, maps to the point $\tilde{\mathbf{y}}_1$ of coordinate 0 on $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, enriched by $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$;
- the point $\tilde{\mathbf{x}}_2$ where L meets the proper transform parabola of equation $x + 2y^2 = 0$, enriched by L and the proper transform of the parabola, maps to the point $\tilde{\mathbf{y}}_2$ of coordinate ∞ on $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, enriched by $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and the x -axis.

Thus N extends to a mapping

$$\mathcal{F}(Y, \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2\}) \rightarrow \mathcal{F}(Y, \{\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2\}). \quad (5.2)$$

But in the domain, L together with the two Farey divisors is precisely the Farey divisor corresponding to \tilde{B}_1 , whereas in the range, the Farey divisors corresponding to $\tilde{\mathbf{y}}_1$ and $\tilde{\mathbf{y}}_2$ are sub-Farey divisors of the Farey divisor corresponding to $\tilde{\mathbf{p}}_1$, so the mapping is well defined with values in X_1 .

Next, we need to extend the Newton mapping to the Farey divisor at \tilde{A}_1 ; this is easier. Perform in the domain \mathbb{P}^2 two blow-ups at \mathbf{p}_1 , to construct first $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and then $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. In the range \mathbb{P}^2 , perform two blow-ups at \mathbf{p}_2 , to construct $L \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Then the Newton map is a local isomorphism from the enriched point \tilde{A}_1 to $L \begin{pmatrix} 1 \\ 2 \end{pmatrix} \cap L \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, enriched by the lines themselves. Again, Theorem 4.2.2 now provides the extension; note that the Farey divisor in the range is a sub-Farey divisor of the divisor associated to $\tilde{\mathbf{p}}_2$, so the range is X_1 .

The constructions for \tilde{B}_2 and \tilde{A}_2 are completely parallel. \square

Here is a more computational approach to the extension at \tilde{B}_2 .

In the coordinates $u = \frac{x}{y}$ and $v = \frac{y}{x^2}$, the line $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ becomes the axis $u = 0$, and the parabola P_y of equation $y = -2x^2$ becomes the line $v = -2$. We know how to parametrize the components of the Farey blow-up in these coordinates: set $u = m^k s^p$, $v = m^l s^q$, where $kq - pl = 1$. This means to “parametrize” the components of the Farey blow-up by rational numbers $p/q \in \mathbb{Q}_+^*$, so that ∞ is on P_y and 0 is on $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ (see Figure 41).

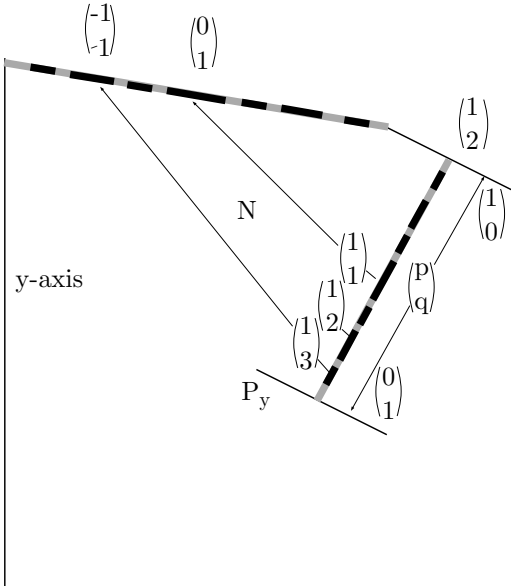


FIGURE 41. This represents how the Farey blow-up of \tilde{B}_2 maps to X_1 .

Then the point with these coordinates tends to the point of coordinate m of the $\begin{pmatrix} p \\ q \end{pmatrix}$ -component of the Farey blow-up as $s \rightarrow 0$. Set $s = 1/t$; changing back to the (x, y) coordinates, we see that the curve

$$x = \frac{t^{p+q}}{m^k(-2t^q + m^l)}, \quad y = \frac{t^{2p+q}}{m^{2k}(-2t^q + m^l)} \tag{5.3}$$

tends to the point of coordinate m of the $\begin{pmatrix} p \\ q \end{pmatrix}$ -component of the Farey blow-up.

Remark. These formulas represent x and y as Puiseux expansions in t with two terms. This is typical: as we blow up more and more times, paths leading to the deeper blow-ups need to be sneakier and sneakier, and require more and more terms. No finite Puiseux expansion gives paths that lead to the points of the projective limit; these require series, and the curves leading to them are necessarily transcendental. This sort of thing was explored earlier (for Hénon mappings) in [Bou], [BS].

The image of this curve is

$$\frac{m^{3k}(-2t^q + m^l)^2}{4t^{3p+2q} - m^{3k}(-2t^q + m^l)^2} \left(\frac{2t^{4p+3q}}{m^{4k}(-2t^q + m^l)^3} + \frac{t^{4p+2q}}{m^{4k}(-2t^q + m^l)^2} \right). \quad (5.4)$$

The characteristic of the curve tending to a point of indeterminacy is a cancellation of principal terms, specifically in the expression

$$\begin{aligned} \frac{2t^{4p+3q}}{m^{4k}(-2t^q + m^l)^3} + \frac{t^{4p+2q}}{m^{4k}(-2t^q + m^l)^2} &= \frac{t^{4p+2q}}{m^{4k}(-2t^q + m^l)^2} \left(\frac{2t^q}{-2t^q + m^l} + 1 \right) \\ &= \frac{m^l t^{4p+2q}}{m^{4k}(-2t^q + m^l)^3}. \end{aligned} \quad (5.5)$$

Other than this, we only have to find the principal terms of everything that remains, which gives

$$\frac{m^{3k}}{t^{3p}} \left(\begin{array}{c} -\frac{m^{l-4k}t^{4p-q}}{8} \\ -\frac{t^{5p}}{4m^{5k}} \end{array} \right) = \left(\begin{array}{c} -\frac{m^{l-k}}{8}t^{p-q} \\ -\frac{1}{4m^{2k}}t^{2p} \end{array} \right). \quad (5.6)$$

We see that the situation is different for $p - q$ odd and for $p - q$ even. If $p - q$ is odd, the $\binom{p}{q}$ -component of the Farey blow-up maps to $L\left(\binom{p-q}{2p}\right)$ by the map $m \mapsto -1/(2^{4p+2q}m^2)$, and the component is not critical.

If $p - q$ is even, then the $\binom{p}{q}$ -component of the Farey blow-up maps to $L\left(\binom{(p-q)/2}{p}\right)$ by an isomorphism, and the component is critical.

Further extensions

From here on, the extensions are easy. The inverse image $\tilde{Z}_2 = N_2^{-1}(\tilde{Z}_1)$ is Farey data on X_2 , and so forth. There is one irritant that must be mentioned: we started with the three points $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ in the finite plane; these are part of Z_1 . Then the inverse images of these points are part of Z_2 , the inverse images of these are parts of Z_3 , and so forth.

It may happen that at some stage, one of these points $\mathbf{z} \in Z_{k-1}$ may lie on a critical value parabola; the values of the parameters where this occurs form a countable union of sets of codimension 1, and may well be dense in the parameter space.

If that happens, we cannot simply blow up the inverse images $N_k^{-1}(\mathbf{z})$ in the critical cubic. As we saw in Example 4.2.1, the Newton map does not then extend to the exceptional divisor; instead, it blows it down to a point, except for one point of indeterminacy (corresponding to the kernel of the derivative $DN_k(N_k^{-1}(\mathbf{z}))$), which must be blown up again. We will deal with this by giving the points $N_k^{-1}(\mathbf{z})$ an enhanced structure that allows us to define this double blow-up; in this case the enhanced structure is a tangent direction at the point.

Once some point is enhanced, all its inverse images will be enhanced also. This is no problem so long as the inverse images are not critical, but if one of these enhanced points again lies on a critical value parabola, its inverse image will need to be further enhanced, by whatever structure it takes to define the finite sequence of blow-ups required for the Newton map to extend. Presumably, this sort of enhancement occurs at most twice, but we don't know this for sure.

By induction, define $\tilde{Z}_i = N_i^{-1}(\tilde{Z}_{i-1})$, (where the inverse image of an enhanced point is an enhanced point, perhaps further enhanced as above) and $X_{i+1} = \mathcal{F}(X_i, \tilde{Z}_i)$. This is possible because of Proposition 5.1.2.

Proposition 5.1.2. (a) For each $i \geq 2$, the set \tilde{Z}_i is Farey data, contained in the most recent Farey divisor.

(b) For every $i \geq 1$, the map N extends to a map $N_{i+1} : X_{i+1} \rightarrow X_i$.

Proof. We assume by induction that N_{i-1} maps $\mathcal{F}(\tilde{Z}_{i-2})$ to \tilde{Z}_{i-2} , and is a finite map on smooth points. Then \tilde{Z}_{i-1} is Farey data, hence X_i exists, and Theorem 4.2.2 guarantees that $(N_{i-1} : X_{i-1} - Z_{i-1}) \rightarrow (X_{i-2} - Z_{i-2})$ extends to $N_i : X_i \rightarrow X_{i-1}$, and is a finite mapping on the smooth points. \square

We now consider the projective limit $X_\infty = \varprojlim (X_i, \pi_i)$, which comes with the mappings $\pi_i : X_\infty \rightarrow X_i$, and a map $N_\infty : X_\infty \rightarrow X_\infty$. The space X_∞ contains a “divisor at infinity” $D_\infty = (\pi_0)^{-1}(l_\infty)$.

The space X_∞ is central to our way of understanding the Newton map, and we will spend the rest of this chapter trying to analyze it.

5.2. Sequence spaces

Spaces similar to X_∞ with infinitely many blow-ups have been constructed in the past, usually by the more straightforward device of considering sequence spaces [HPV], [FR2]. A sequence space can be constructed in this case also, but it is certainly not more attractive than our X_∞ .

The rational map $N_1 : X_1 \rightsquigarrow X_1$ has a graph Γ , and its closure $\bar{\Gamma} \subset X_1 \times X_1$ is a graph except above the seven points of indeterminacy. A natural space to consider is the subset $\Xi_\infty \subset X_1^{\mathbb{N}}$ formed of sequences (ξ_1, ξ_2, \dots) such that $(\xi_i, \xi_{i+1}) \in \bar{\Gamma}$.

There is a natural map $\Pi : X_\infty \rightarrow \Xi_\infty$ defined by

$$\Pi(x_1, x_2, \dots) = (\xi_1 = x_1, \xi_2 = N_2(x_2), \xi_3 = N_2(N_3(x_3)), \dots). \tag{5.7}$$

If we could, we would work with Ξ_∞ , as it is so much easier to write elements of it. This works for birational maps, but not in this case, because Π is not injective.

Proposition 5.2.1. The map Π is not injective.

Proof. Our computation to show this will depend on our computational approach to the extension at \tilde{B}_2 (see Figure 41), and will use the notation set up there. Consider the component $L_{1/4}$ in the Farey blow-up of the point \tilde{B}_2 , according to the parametrization above. This component maps to $L \begin{pmatrix} -3 \\ 2 \end{pmatrix}$, and this map is 2-to-1. Choose some point $\mathbf{y}_1 \in L \begin{pmatrix} -3 \\ 2 \end{pmatrix}$, and let its two inverse images in

L be \mathbf{y}'_0 and \mathbf{y}''_0 . These points can be thought of as belonging to X_∞ , since the component $L \begin{pmatrix} -3 \\ 2 \end{pmatrix}$ maps to the component $L \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ in zone 2, which itself maps to $L \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ in zone 1, where the Newton map is the identity. So no blow-ups will ever be performed on the points $\mathbf{y}'_0, \mathbf{y}''_0$.

Thus the two distinct points of X_∞

$$(B_2, \mathbf{y}'_0, \mathbf{y}'_0, \dots) \quad \text{and} \quad (B_2, \mathbf{y}''_0, \mathbf{y}''_0, \dots) \quad (5.8)$$

both map to the sequence in Ξ_∞

$$(B_2, \mathbf{y}_1, N(\mathbf{y}_1), N^{\circ 2}\mathbf{y}_1, N^{\circ 3}\mathbf{y}_1 = N^{\circ 2}\mathbf{y}_1, \dots). \quad (5.9)$$

This shows that the spaces X_∞ and Ξ_∞ are different. \square

It turns out that Π is a homeomorphism on the most interesting set of points. The most recurrent sequences in Ξ_∞ are the subset Υ_∞ of sequences that contain an infinity of B_1 's or B_2 's. Let us call $Y_\infty = \Pi^{-1}\Upsilon_\infty$.

Proposition 5.2.2. *The map $\Pi : Y_\infty \rightarrow \Upsilon_\infty$ is a homeomorphism.*

Proof. The proof is an unpleasant computation. First, observe that the components of D_1 that eventually map to $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ are

$$L \begin{pmatrix} 1 \\ 1 \end{pmatrix}, L \begin{pmatrix} -4 \\ 1 \end{pmatrix}, L \begin{pmatrix} 1 \\ -10 \end{pmatrix}, L \begin{pmatrix} -22 \\ 1 \end{pmatrix}, L \begin{pmatrix} 1 \\ -46 \end{pmatrix}, \quad (5.10)$$

where the indices n_i that are not 1 are obtained recursively by the rule $-n_{i+1} = -(2n_i + 2)$.

In particular, these indices are all even, and these components are all critical, except for $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

We will show that any component of a Farey exceptional divisor of X_n that maps to one of these by $N_1 \circ N_2 \circ \dots \circ N_n$ is a critical curve. This will prove the proposition.

Each Farey blow-up is “parametrized” by the positive reals in a natural way, for instance by the points of the line $p = 1$ in the associated quadrant Q .

These enriched points where we are making the blow-ups are either critical or not; if they are critical, then we will call the associated Farey blow-up critical. Only components of critical Farey blow-ups can be critical.

To begin with, we saw above that the critical components of the Farey blow-up of B_1 and B_2 are the $\begin{pmatrix} p \\ q \end{pmatrix}$ -components with $q - p$ even, and that these map to the components

$$\left(\begin{pmatrix} p \\ (p-q)/2 \end{pmatrix} \right) \quad \text{and} \quad \left(\begin{pmatrix} (p-q)/2 \\ p \end{pmatrix} \right) \quad (5.11)$$

respectively. So the inverse image of $L \begin{pmatrix} 1 \\ q \end{pmatrix}$ with $q < 2$ is the component $\begin{pmatrix} 1 \\ 1-2q \end{pmatrix}$ (in the blow-up of B_1), and exactly the same holds at the blow-up at B_2 . All these components are critical.

Now it is easy to proceed by inverse images. At every step, the inverse image of one of these components is a component labeled $\begin{pmatrix} 1 \\ q \end{pmatrix}$ for some q ; indeed, only such a component can map

to another whose first coordinate is 1, whether on a critical Farey blow-up or not. If it is on a non-critical Farey blow-up, it certainly maps injectively to its image (all the components of such a blow-up do), and if it is on a critical Farey blow-up, then it is necessarily on a critical component, and also maps bijectively to its image. \square

5.3. The real oriented blow-up of X_∞

The space X_∞ appears pretty frightful, but it really has the same complication as one of the most standard objects of dynamical systems: the family of invariant tori appearing in small perturbations of integrable Hamiltonian systems with two degrees of freedom.

At least, that is the structure which arises after having taken the real oriented blow-up of the Farey divisor at infinity. More precisely, the spaces

$$\mathcal{B}_{\mathbb{R}}^+(X_n, D_n) = \mathcal{B}^+ \mathcal{F}(X_{n-1}; \tilde{Z}_{n-1}, D_{n-1}) \quad \text{and} \quad \mathcal{B}_{\mathbb{R}}^+(D_n) = \mathcal{B}^+ \mathcal{F}(\tilde{Z}_{n-1}, D_{n-1}) \quad (5.12)$$

were defined in Chapter 4; we will consider the projective limits

$$\varprojlim \mathcal{B}_{\mathbb{R}}^+(D_n) \subset \varprojlim \mathcal{B}_{\mathbb{R}}^+(X_n, D_n), \quad (5.13)$$

which we will denote by $\mathcal{B}_{\mathbb{R}}^+ D_\infty \subset \mathcal{B}_{\mathbb{R}}^+(X_\infty, D_\infty)$.

Theorem 5.3.1. (a) *The subset $\mathcal{B}_{\mathbb{R}}^+(D_\infty) \subset \mathcal{B}_{\mathbb{R}}^+(X_\infty, D_\infty)$ is homeomorphic to a 3-sphere.*

(b) *The map $N_\infty : X_\infty \rightarrow X_\infty$ lifts to a map $\mathcal{B}_{\mathbb{R}}^+(D_\infty) - (S_1 \cup S_2) \rightarrow \mathcal{B}_{\mathbb{R}}^+(D_\infty)$, where S_1 and S_2 are the two circles corresponding to the points where the polar hyperbola meets $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $L \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. This map is a ramified covering map, ramified along the four circles where the critical cubics intersect $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}, L \begin{pmatrix} -1 \\ 1 \end{pmatrix}, L \begin{pmatrix} 1 \\ 2 \end{pmatrix}, L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.*

Proof. Part (a) follows immediately from Theorem 5.10 in [HPV].

Part (b) follows from Theorem 4.4.1. \square

This description is clean, but not very geometric. Figure 42 is an attempt to draw the first stage of the real oriented blow-up. Make the real oriented blow-up of the primitive space $X_P = X_1$ along the divisor D_1 at infinity. This is a 3-sphere with a Cantor set of concentric tori in it. The zones (or gaps) between tori correspond to the components of D_∞ , and hence are labeled by the primitive points, i.e., by pairs of coprime integers with at least one positive. These gaps are foliated by closed curves; the $\begin{pmatrix} p \\ q \end{pmatrix}$ -gap is foliated by $\begin{pmatrix} p \\ q \end{pmatrix}$ torus knots. Two of the tori are degenerate; they are round (ordinary) circles in the 3-sphere, and correspond to the adherence of the x and y axes.

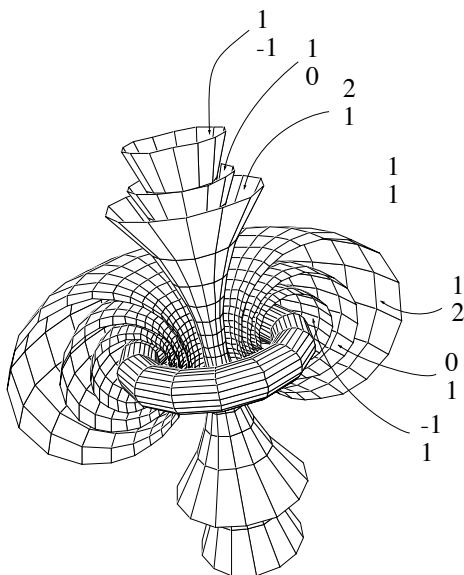


FIGURE 42. A sketch of the 3-sphere at infinity if we take a real oriented blow-up of X_1 along D_1 . We have not managed to actually draw a Cantor set of tori; one should imagine each of those drawn as thickened to give a Cantor-set section.

We next need to understand the Farey blow-up of the point of coordinate -2 in $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. That blow-up thickens a circle orbit in the zone corresponding to $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, and puts another Cantor set of tori inside the thickened circle orbit, now a solid torus. Figure 43 attempts to draw the first thickened circle, now torus (which looks flatter than we would like).

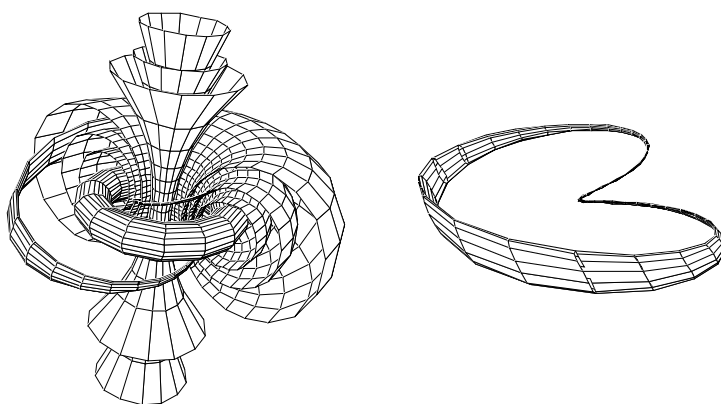


FIGURE 43. Left: one circle orbit in the zone $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ has been thickened to form a solid torus; one should imagine it filled with a Cantor set of concentric tori, with zones in between corresponding to the components of the Farey blow-up of the point of coordinate -2 in $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Right: the solid torus by itself, shown winding twice in one direction and once in the other, as its label $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ specifies.

The real oriented blow-up contains three other new tori, in the zones corresponding to $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, and $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$; it is beyond our software to draw these. At the next level, we will need to thicken more circles (12 of them), of which two will be inside the zones within the most recent Farey blow-up. We won't try to draw this in any detail, but Figure 44 attempts to sketch what we find in the zone $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

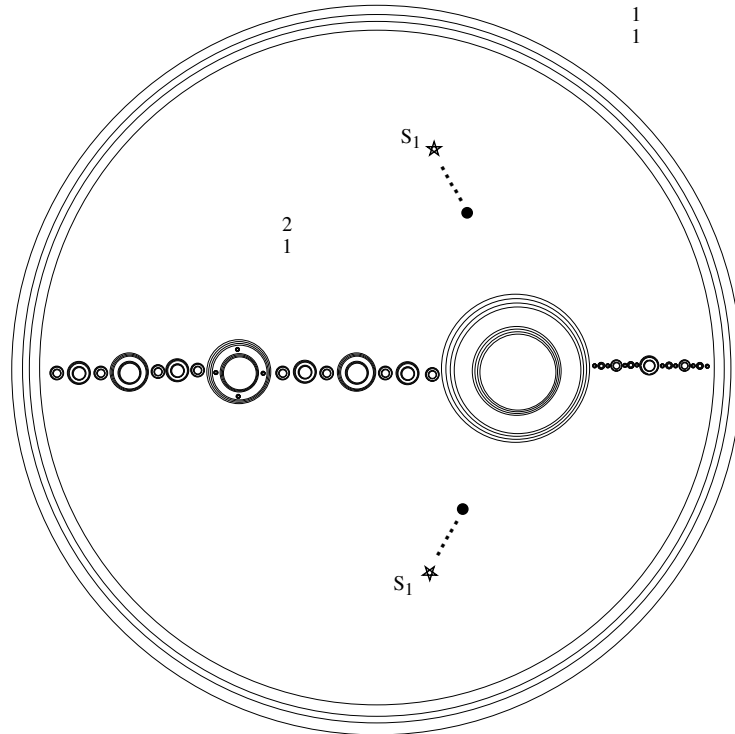


FIGURE 44. In a cross-section of zone $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, we see the region between two circles. Crossing from the inner one to the outer one, in two opposite places, we see the Julia set of the rational map (3.18). Originally, this Julia set is just a line segment (joining 0 to ∞), but the critical point (of coordinate -2) has been blown up to a solid torus, as have all its inverse images, which are dense in the interval and blown up, make a string of beads joining the interior to the exterior torus. In three dimensions, it is more like a ribbon. Each of these beads is itself a Cantor set of circles, with gaps, of which some themselves contain Cantor sets of circles, and so on ad infinitum. In addition, the circle S_1 is in this zone, and intersects our cross-section in two points, indicated by stars. This circle is one of the four circles of ramification of the map N_∞ . It is attracted to the circle corresponding to the fixed point at coordinate $2 - \sqrt{2}$. This circle also intersects the cross-section in two points, indicated as black dots.

What are the fibers of the projection $\mathcal{B}_\mathbb{R}^+(D_\infty) \rightarrow D_\infty$? Some are easy to understand, namely those that correspond to a point \mathbf{z} of some D_n at which no further Farey blow-ups have been performed. Then the fiber is either:

- A circle, if $\mathbf{z} \in X_n^*$.
- A torus, if \mathbf{z} is not a smooth point of D_∞ . This torus carries a foliation: an irrational foliation if \mathbf{z} is not in the closure of any component of D_n , and a rational foliation by closed curves if \mathbf{z}

is in the closure of some component of D_n . This applies in particular to the points \mathbf{z} of D_n that correspond to points of some earlier D_m where Farey blow-ups were performed.

The more complicated fibers occur above the points of Y_∞ ; they are solenoids. These solenoids are accumulations of more and more elaborate iterated torus knots, and come with a sequence $\left(\begin{smallmatrix} p_1 \\ q_1 \end{smallmatrix}\right), \left(\begin{smallmatrix} p_2 \\ q_2 \end{smallmatrix}\right), \dots$ that describes what sort of solenoid it is, and how it is embedded in S^3 . Objects of this sort are implicit in the standard picture of the KAM theorem, as the limits of solutions that live on islands off islands off islands \dots ; however, such solutions of Hamiltonian systems do not appear to have been analyzed in the literature.

There is much more to say about the dynamics of this real oriented blow-up, more particularly about the topology of the ‘‘Cantor set of solenoids,’’ and the invariant measures which it carries, which will be the topic of a future paper.

5.4. The homology of X_1^* .

Theorem 4.5.5 tells us what $H_2(X_\infty^*)$ is: adding a star to a space means removing its non-algebraic points. Theorem 4.6.1 tells us how to compute the action of the Newton map on homology. Moreover, Theorem 4.7.2 gives us a more conceptual way of thinking of this homology, as the Sobolev space \mathcal{H}^1 of a rather complicated infinite graph.

Let us start with X_1 . Since X_1 is obtained from $X_0 = \mathbb{P}^2$ by performing two Farey blow-ups, one at \mathbf{p}_1 and the other at \mathbf{p}_2 , Theorem 4.5.5 tells us that the homology of this space is

$$H_2(X_1^*) = \mathbb{Z} \oplus S_1 \oplus S_2, \quad (5.14)$$

where S_1 is a copy of S corresponding to \mathbf{p}_1 , and S_2 is a copy of S corresponding to \mathbf{p}_2 . The direct sum is orthogonal, and since the initial \mathbb{C} corresponds to $H_2(\mathbb{P}^2)$, represented by a line in the plane with self-intersection 1, the intersection quadratic form assigns norm 1 to the generator of the first factor.

There is an alternative way of describing this space, which we will prefer. As in Proposition 4.5.1, $H_2(X_1^*)$ can be thought of as a subgroup of $\mathbb{Z}^{\text{Irr } D_1}$ and as in Chapter 4, there is another way of thinking of $\text{Irr } D_1$. Let Q_P be subset of \mathbb{R}^2 with at least one coordinate ≥ 0 (i.e., the complement of the third quadrant). We have redrawn the ‘‘Farey plane’’ Q_P from Figure 31, on the left of Figure 45. The primitive points of Q_P correspond to the elements of $\text{Irr } D_1$.

Consider the group S_P of continuous functions on Q_P , piecewise linear and linear except on finitely many rays with rational slopes, which take integral values on $\mathbb{Z}^2 \cap Q_P$, and which are 0 at $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$. (This group S_P is just like the group S of Definition 4.5.3, except that the functions of S_P are defined on three quadrants, not one.) We can think of S_P as a subgroup of $\mathbb{Z}^{\text{Irr } D_1}$ by restricting a function on Q_P to the primitive points

Proposition 5.4.1. *The natural inclusion $H_2(X_1^*) \rightarrow \mathbb{Z}^{\text{Irr } D_1}$ is an isomorphism onto S_P .*

Proof. This is very similar to Theorem 4.5.5, or alternately follows from it with a bit of work. \square

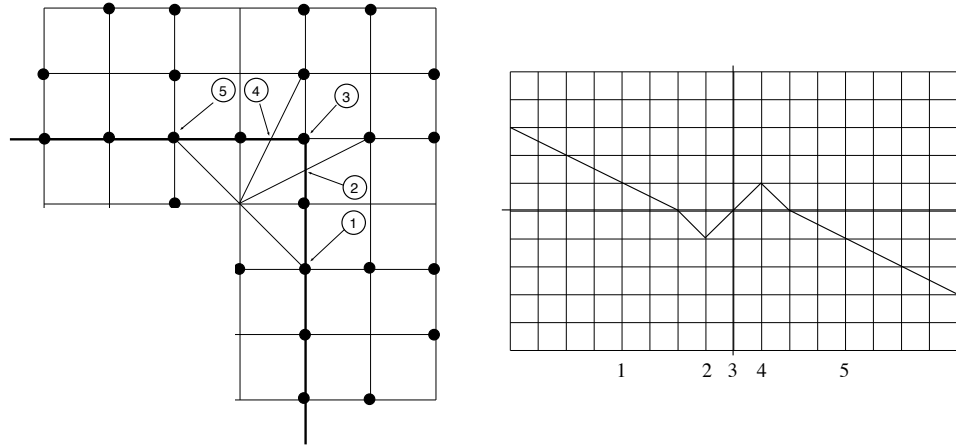


FIGURE 45. Left: the “Farey plane” of Figure 31, with the line segments joining $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ drawn in, and with seven distinguished points on the segments. Right: the graph of the function that encodes where each component of D_1 is mapped; this function will be denoted N_J .

For each primitive point $\begin{pmatrix} p \\ q \end{pmatrix} \in Q_P$, with parents $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$ and $\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}$ we will call $\mathbf{e}_{\begin{pmatrix} p \\ q \end{pmatrix}} \in S_P$ the tent function, analogous to that defined in Equation (4.29), that is linear except on the rays through $\begin{pmatrix} p \\ q \end{pmatrix}$, $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$, and $\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}$, has value 1 at $\begin{pmatrix} p \\ q \end{pmatrix}$ and value 0 at $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$ and $\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}$.

As for the components, we label the functions by vectors rather than fractions, since a ray in the primitive plane is not specified by its slope.

Proposition 5.4.2. *The elements $\mathbf{e}_{\begin{pmatrix} p \\ q \end{pmatrix}}$ of S_P form a basis.*

Proof. This is almost identical to Lemma 4.5.4; again we omit the proof. \square

Intersection on X_1^* induces a pseudo-inner product on $H_2(X_1^*)$, and our basis is well adapted to this product.

Proposition 5.4.3. *For the natural inner product, we have*

$$\langle \mathbf{e}_{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}, \mathbf{e}_{\begin{pmatrix} 1 \\ 1 \end{pmatrix}} \rangle = 1, \tag{5.15}$$

$$\langle \mathbf{e}_{\begin{pmatrix} p \\ q \end{pmatrix}}, \mathbf{e}_{\begin{pmatrix} p \\ q \end{pmatrix}} \rangle = -1 \quad \text{for all primitive vectors } \begin{pmatrix} p \\ q \end{pmatrix} \neq \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{5.16}$$

$$\langle \mathbf{e}_{\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}}, \mathbf{e}_{\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}} \rangle = 0 \quad \text{when } \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \neq \begin{pmatrix} p_2 \\ q_2 \end{pmatrix}. \tag{5.17}$$

Proof. This was implied by Proposition 5.4.1. Indeed, the homology class $\mathbf{e} \binom{p}{q}$ is represented by the divisor $L \binom{p}{q}$ at the level of blow-up at which it first appears; it is then deformed so as to avoid all the points where future blow-ups will take place. This applies to all such classes except $L \binom{1}{1}$, which is the original line at infinity, similarly deformed (for instance into the diagonal, or any line that avoids \mathbf{p}_1 and \mathbf{p}_2).

Thus clearly all these classes have self-intersection -1 , just as they have at the moment of creation (except $L \binom{1}{1}$, which wasn't created, and has self-intersection 1), and they are all orthogonal to each other. \square

As in Section 4.7, there is a way of thinking of $S_P \otimes_{\mathbb{Z}} \mathbb{C}$ as a dense subset of a Sobolev space.

In Figure 45, we have emphasized the half-lines $p = 1, q \leq 1$, and $q = 1, p \leq 1$; we call the union of these two rays J , and think of it as simply \mathbb{R} , with the first piece parametrized by $1 - q$ (hence corresponding to the positive reals) and the second by $p - 1$ (corresponding to the negative reals). Of course elements of S_P can be restricted to J as piecewise linear functions.

We will consider the space $\mathcal{H}^1(J)$ of continuous functions with distributional derivative in L^2 , and we will give it the pseudo-inner product

$$\langle f, g \rangle = - \int_J f'(t) \overline{g'(t)} dt + f(0) \overline{g(0)}. \quad (5.18)$$

Proposition 5.4.4. *The restriction mapping $S_P \rightarrow \mathcal{H}^1(J)$ is an isometric embedding, and the image of $S_P \otimes_{\mathbb{Z}} \mathbb{C}$ is dense in $\mathcal{H}^1(J)$.*

Proof. It is enough to see that the images of the $\mathbf{e} \binom{p}{q}$ are orthogonal and have the correct pseudo-length. The function $\mathbf{e} \binom{1}{1}$ restricts to the constant 1 and hence has the right pseudo-length. For the other basis vectors, the proof is identical to that of Theorem 4.7.2. For the density, consider for instance that the derivatives of our functions include the Haar functions. \square

The next step is to compute the map $(N_P)_* : H_2(X_1^*) \rightarrow H_2(X_1^*)$; Recall that N_P is defined in Proposition 3.2.1. The computation is very similar to the one in the proof of Theorem 4.6.1, and we won't go through all the details, especially as most of the work was done in Section 3.2. Just as there, we must deal separately with the classes $\mathbf{e} \binom{p}{q}$ when $\binom{p}{q}$ is in zones 1-5, and also look carefully at the special cases.

Let us begin with the hardest case, zone 1. We will compute the image of $\mathbf{e} \binom{p}{q}$ when the parents are $\binom{p_1}{q_1}$ and $\binom{p_2}{q_2}$ with q_1 even. If we perform two blow-ups in the domain and range, we get Figure 46, very analogous to Figure 38, where the labels describing what maps to what, and what is critical, were all computed in Section 3.2.

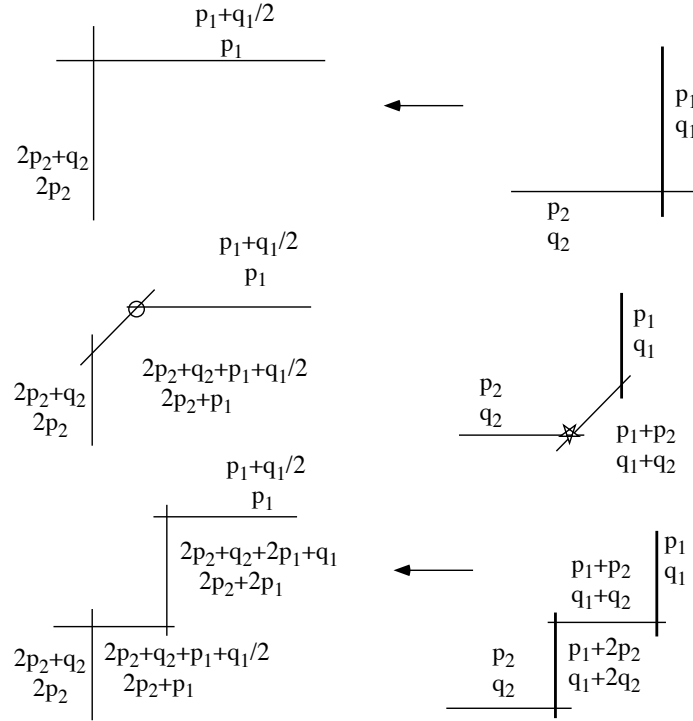


FIGURE 46. The Newton mapping, acting on a partial blow-up, on the way to constructing X_1 .

In the domain, the homology class $\mathbf{e} \begin{pmatrix} p \\ q \end{pmatrix}$ is represented by

$$L \begin{pmatrix} p \\ q \end{pmatrix} + L \begin{pmatrix} p_1+2p_2 \\ q_1+2q_2 \end{pmatrix}. \tag{5.19}$$

This is mapped (remembering what is critical and what isn't) to

$$L \begin{pmatrix} p_1+2p_2+q_2+q_1/2 \\ 2p_2+p_1 \end{pmatrix} + 2L \begin{pmatrix} 2p_1+2p_2+q_1+q_2 \\ 2p_2+2p_1 \end{pmatrix}. \tag{5.20}$$

Finally, this combination of divisor components, at this level of blow-up, represents the class

$$\mathbf{e} \begin{pmatrix} p_1+2p_2+q_2+q_1/2 \\ 2p_2+p_1 \end{pmatrix} + \mathbf{e} \begin{pmatrix} 2p_1+2p_2+q_1+q_2 \\ 2p_2+2p_1 \end{pmatrix}. \tag{5.21}$$

The same computation shows that $\mathbf{e} \begin{pmatrix} p_1+2p_2 \\ q_1+2q_2 \end{pmatrix}$ (for which both parents have second coordinate odd) maps to

$$\mathbf{e} \begin{pmatrix} p_1+2p_2+q_2+q_1/2 \\ 2p_2+p_1 \end{pmatrix} - \mathbf{e} \begin{pmatrix} 2p_1+2p_2+q_1+q_2 \\ 2p_2+2p_1 \end{pmatrix}. \tag{5.22}$$

This computes the action on homology in zone 1; zone 5 is identical, after exchanging the first and second coordinates.

In zone 2, the Newton map simply takes $\mathbf{e} \begin{pmatrix} p \\ q \end{pmatrix}$ to $\mathbf{e} \begin{pmatrix} p \\ p-q \end{pmatrix}$, and in zone 3, it acts as the identity.

This leaves the exceptional cases, which we simply list; the computations are all the same sort of thing: blow up just until you can see what is going on, and no more:

$$\begin{aligned}
\mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix} &\mapsto \mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
\mathbf{e} \begin{pmatrix} 2 \\ 1 \end{pmatrix} &\mapsto \mathbf{e} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \mathbf{e} \begin{pmatrix} 1 \\ 2 \end{pmatrix} &\mapsto \mathbf{e} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \mathbf{e} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\mathbf{e} \begin{pmatrix} 1 \\ -1 \end{pmatrix} &\mapsto \mathbf{e} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \mathbf{e} \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \mathbf{e} \begin{pmatrix} -1 \\ 1 \end{pmatrix} &\mapsto \mathbf{e} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
\end{aligned} \tag{5.23}$$

In some sense, this is a complete computation of the action of N on the homology of the primitive space. However, it is extremely unwieldy, and as it turns out, there is a remarkably simple way of writing all of it.

To this end, we notice that there is a map $N_J : J \rightarrow J$ that encodes where each component of D_1 is mapped; we saw the graph of this function in Figure 45, on the right.

Using this map, we can “define” a mapping

$$(N_J)_* : \mathcal{H}^1(J) \rightarrow \mathcal{H}^1(J) \quad \text{by} \quad (N_J)_*(\mathbf{f})(x) = \sum_{y \in N_J^{-1}(x)} \mathbf{f}(y). \tag{5.24}$$

There is something wrong with this mapping: the image of a continuous map may well fail to be continuous at the points $\pm 1/2$, corresponding to the points $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, where two inverse images collide and disappear. Formula (5.26) in Theorem 5.4.5 exactly compensates for this difficulty.

Theorem 5.4.5. *The mapping $(N_P)_* : H_2(X_P^*) \rightarrow H_2(X_P^*)$ extends uniquely to the mapping*

$$\widehat{(N_P)}_* : \mathcal{H}^1(J) \rightarrow \mathcal{H}^1(J) \tag{5.25}$$

given by

$$\widehat{(N_P)}_*(\mathbf{f}) = (N_J)_*(\mathbf{f}) + 2 \left(\mathbf{f} \left(-\frac{1}{2} \right) \chi_{[-1/2, -\infty)} + \mathbf{f} \left(\frac{1}{2} \right) \chi_{[1/2, \infty)} \right), \tag{5.26}$$

where χ_I is the characteristic function of I .

Proof. We will prove this by showing that it is true for each $\mathbf{e} \begin{pmatrix} p \\ q \end{pmatrix}$. The most interesting case is

where $\begin{pmatrix} p \\ q \end{pmatrix}$ is in zone 1 or 5, but in those cases the computation is identical to that in Theorem 4.6.1, and more specifically illustrated in Figure 39, and we omit it. For $\begin{pmatrix} p \\ q \end{pmatrix}$ in zones 2, 3, or 4, the formula is easy to check, but we need to look carefully at the exceptional cases.

The case of $\mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The function

$$(N_J)_* \left(\mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \tag{5.27}$$

is the constant 3 on $[-1/2, 1/2]$, and the constant function 1 elsewhere. The extra terms add on the constant 2 outside of $[-1/2], [1/2]$, giving the constant function 3.

The case of $\mathbf{e} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. In this case,

$$(N_J)_* \left(\mathbf{e} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right) \tag{5.28}$$

is the function $2x$ on $[0, 1/2]$, extended by 0. The extra term adds the constant 1 on $[1/2, \infty)$, and these two together are equal to $\mathbf{e} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

The case of $\mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The function

$$(N_J)_* \left(\mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) \tag{5.29}$$

is the constant 1 on $(-\infty, 1/2]$, and the constant function 0 elsewhere. The extra terms add on the constant 1 on $[1/2, \infty]$, giving the constant function 1, i.e., $\mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. \square

This proves the result, but the proof is very unsatisfactory. The statement of Theorem 5.4.5 is so clean that it seems impossible that the interpretation of homology in terms of Sobolev spaces is just a fluke. There must be a reason why this works other than several miraculous computations, but we were unable to find a conceptual proof.

5.5. The homology $H_2(X_\infty^*)$.

We will now compute the homology of X_∞^* , as the inductive limit $\varinjlim H_2(X_n^*)$ under the obvious inclusions. Exactly as in Proposition 4.5.2, this homology is naturally a subgroup of $\mathbb{Z}^{\text{Irr}(D_\infty \cup E_\infty)}$, which we will identify.

Each X_{i+1} is obtained from X_i by performing some finite set of Farey blow-ups on X_i , at points $\mathbf{z} \in D_i$. For each such Farey blow-up, take a new copy $Q_{\mathbf{z}}$ of the first quadrant in \mathbb{R}^2 . This copy comes with the commutative group $S_{\mathbf{z}}$ of continuous, piecewise-linear functions on $Q_{\mathbf{z}}$ that are linear except on finitely many rays with rational slopes, and take integral values at integral points. This group comes with a basis $\mathbf{e}_{p/q, \mathbf{z}}$ labeled by the points $\begin{pmatrix} p \\ q \end{pmatrix}$ with coprime integer components, as in Equation (4.29). When discussing these functions, which are defined on a single quadrant, we return to the notation of Chapter 4, and encode them by the slope p/q of the ray through $\begin{pmatrix} p \\ q \end{pmatrix}$.

We will need to add one element to this group: define $S_{\mathbf{z}}^+$ to be the same set of functions, except that we no longer insist that they vanish at $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. There is a basis vector $\mathbf{e}_{1/0}$ corresponding to this element, which is simply the function p , if we call (p, q) the coordinates on $Q_{\mathbf{z}}$.

Now we define a two-dimensional tree-like “complex,” starting with $\mathcal{Q}_1 = Q_P$, and making \mathcal{Q}_n by attaching a copy of $Q_{\mathbf{z}}$ to \mathcal{Q}_{n-1} for each point $\mathbf{z} \in D_{n-1}$ at which a Farey blow-up is performed.

This attaching is done by identifying the ray through $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ on $Q_{\mathbf{z}}$ to the ray through the point of Q_{n-1} corresponding to the component to which \mathbf{z} belongs, identifying the point $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to that point of Q_{n-1} (see Figure 48).

Set Q'_∞ to be the union of all the quadrants corresponding to Farey blow-ups at non-critical points, and Q''_∞ the union of the quadrants corresponding to Farey blow-ups at critical points. The attaching ray always belongs to the first quadrant where it occurs. By definition, within Q_1 , zones 1 and 5 both belong to Q''_∞ , and zones 2, 3, and 4 belong to Q'_∞ ; the Farey blow-ups at B_1 and B_2 belong to Q''_∞ , and the Farey blow-ups at A_1 and A_2 belong to Q'_∞ .

It should now be clear that the points of Q_n with coprime integral coordinates correspond to the components of D_n . We define $Q_\infty = \cup_n Q_n$. There is a natural group S_n of continuous functions on Q_n , piecewise linear and linear except for finitely many rays with rational slope, vanishing at all boundary rays, and taking integral values at all integral points. We can also consider the analogous group S_∞ of functions on Q_∞ with the same requirements. Note that the requirements imply that the function is linear except on finitely many of the $Q_{\mathbf{z}}$, i.e., a multiple of $\mathbf{e}_{1/0,\mathbf{z}}$ there.

There are obvious restriction mappings $S_n \rightarrow S_m$ when $n > m$ (in particular, when $n = \infty$), and there also are mappings $S_m \rightarrow S_n$ (still when $n > m$) that consist of extending functions linearly to all the new quadrants $Q_{\mathbf{z}}$. Using these extensions, it should be clear that

$$S_\infty = \varinjlim S_n. \quad (5.30)$$

To describe the homology of X_∞ we must also deal with the blow-ups of points in the finite plane, which we will do by simply defining $\mathcal{I}_n = \text{Irr } E_n$, and $\mathcal{I}_\infty = \text{Irr } E_\infty = \cup_n \mathcal{I}_n$.

Theorem 5.5.1. *The map*

$$S_\infty \oplus \mathbb{Z}^{(\mathcal{I}_\infty)} \rightarrow \mathbb{Z}^{\text{Irr } D_\infty \cup \text{Irr } E_\infty} \quad (5.31)$$

that on the first summand restricts functions to the integral points with coprime coordinates, and on the second is the obvious inclusion $\mathbb{Z}^{(\text{Irr } E_\infty)} \rightarrow \mathbb{Z}^{\text{Irr } E_\infty}$, is an isomorphism onto $H_2(X_\infty^) \subset \mathbb{Z}^{\text{Irr } D_\infty \cup \text{Irr } E_\infty}$.*

Proof. The homology is the inductive limit of the system

$$H_2(X_1^*) \rightarrow H_2(X_2^*) \rightarrow \dots \quad (5.32)$$

Thus we need to see two things: that

$$H_2(X_n^*) = S_n \oplus \mathbb{Z}^{\mathcal{I}_n} \quad (5.33)$$

and that the diagrams

$$\begin{array}{ccc} H_2(X_n^*) & \rightarrow & S_n \oplus \mathbb{Z}^{\mathcal{I}_n} \\ \downarrow & & \downarrow \\ H_2(X_{n+1}^*) & \rightarrow & S_{n+1} \oplus \mathbb{Z}^{\mathcal{I}_{n+1}} \end{array} \quad (5.34)$$

commute.

This is proved by induction, using Theorem 4.5.5 (and Theorem 4.6 of [HPV] for the points in the finite plane.) Let us spell it out a bit, though. To think of $H_2(X_\infty^*)$ as a subgroup of $\mathbb{Z}^{\text{Irr } D_\infty \cup \text{Irr } E_\infty}$, we must think of the projective system made up of all ways of making finitely many blow-ups (among the infinitely many that will eventually be performed to make X_∞). Then

for each of these spaces, the homology is generated by the fundamental classes of all the proper transforms of the exceptional divisors, plus the proper transform of the line at infinity.

At the time an exceptional divisor is created, we can also deform it so that it avoids any points where future blow-ups will take place. It then represents an element of $H_2(X_\infty^*)$, but that class is *not* the proper transform of the divisor in question after further blow-ups have been performed; the precise statement is given in Proposition 4.8 of [HPV], repeated here as Proposition 5.5.2.

Proposition 5.5.2. *The composition*

$$H_2(X) \rightarrow H_2(X - \{\mathbf{z}\}) \rightarrow H_2(\tilde{X}_{\mathbf{z}}) \tag{5.35}$$

maps the fundamental class $[C]$ of a curve C in X with m smooth branches through \mathbf{z} to $[C'] + m[E] \in H_2(\tilde{X}_{\mathbf{z}})$, where C' is the proper transform of C in $\tilde{X}_{\mathbf{z}}$, and E is the exceptional divisor.

Suppose now that a Farey blow-up is to be performed at a point $\mathbf{z} \in D_n$, and that \mathbf{z} is a point of some component D . In the system of finite blow-ups, let us put ourselves at a level at which D has just been created. When we make our first blow-up at \mathbf{z} , creating the component $L_{1/1,\mathbf{z}}$ of the eventual Farey blow-up, the class $[D]$ is represented by $[D'] + [L_{1/1,\mathbf{z}}]$. When we next blow up the point of intersection of these two, creating $L_{2/1,\mathbf{z}}$, the image of $[D]$ will be

$$[D'] + [L_{1/1,\mathbf{z}}] + 2[L_{2/1,\mathbf{z}}]. \tag{5.36}$$

By induction, it should be clear that when the whole Farey blow-up has been performed, the class $[D]$ will become $[D'] + \mathbf{e}_{1/0,\mathbf{z}}$. This shows that the diagrams (5.34) all commute, and proves Theorem 5.5.1. \square

The space \mathcal{Q}_∞ is a bit unwieldy, and we want to replace it by something more tangible, so that S_∞ will become a space of functions on a graph. We will draw this graph in \mathcal{Q}_∞ as follows, shown in Figure 47. Begin by setting $J_1 = J \subset \mathcal{Q}_1 = \mathcal{Q}_P$. Next, any time we attach a quadrant $Q_{\mathbf{z}}$ to \mathcal{Q}_{n-1} , the ray where the attaching takes place will intersect J_{n-1} in a unique point. Draw through that point the parallel to the free boundary ray of $Q_{\mathbf{z}}$. Note that the function $\mathbf{e}_{1/0,\mathbf{z}}$ restricts to a constant on this ray. The union of the rays obtained this way is our graph J_∞ .

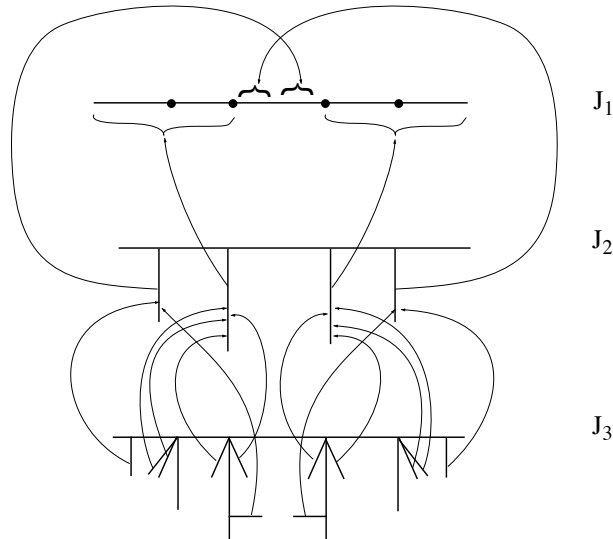


FIGURE 47. A sketch of J_1 , J_2 and J_3 , together with the way the map N_J transforms segments.

Give each ray I of J the unique multiple of Lebesgue measure so that the restrictions of the functions $\mathbf{e}_{p/q, \mathbf{z}}$ to the ray are orthonormal in $\mathcal{H}^1(I)$. This is possible by Theorem 4.7.3. This measure can be written $|dt|$, for a unique affine parametrization of the ray that is 0 at the point of attachment.

We will still denote by $|dt|$ the non-atomic measure on J that restricts to $|dt|$ on each ray. This puts a pseudo-inner product on S_∞ , by setting

$$\langle f, g \rangle = - \int_J f'(t)g'(t) |dt| + f(0)g(0) \quad (5.37)$$

In this formula, the point 0 is the point in the middle of J_1 , i.e., the point $\left(\frac{1}{1}\right)$ of Q_P . Note that this integral is finite since elements of S_∞ are constant except on finitely many rays, and there their derivatives have compact support.

Theorem 5.5.3. *The isomorphism $H_2(X_\infty^*) \rightarrow S_\infty$ of Theorem 5.5.1 transforms the intersection inner product of $H_2(X_\infty^*)$ into the inner product*

$$\langle f, g \rangle = - \int_J f'(t)\overline{g'(t)}|dt| + f(0)\overline{g(0)} \quad (5.38)$$

on S_∞ .

Proof. This follows from Theorem 4.7.3, together with Theorem 4.5.6. \square

Define $\hat{H}_2(X_\infty^*)$ to be the completion of $H_2(X_\infty^*)$ with respect to the inner product. Then Theorem 5.5.3 has the following corollary.

Corollary 5.5.4. *The isomorphism $H_2(X_\infty^*) \rightarrow S_\infty$ of Theorem 5.5.1 extends to an isometry $\hat{H}_2(X_\infty^*) \rightarrow \mathcal{H}^1(J)$.*

5.6. The action of N on homology

Corollary 5.5.4 makes it very tempting to try to understand the mapping N_* on the homology in terms of the Sobolev space interpretation. Since the part at infinity and the part in the finite plane behave independently, and since the part in the finite plane is simply a one-sided shift on the homology, we are going to ignore everything coming from the finite plane. For this Section, X_∞ means the space X_∞ where all the blow-ups in the finite plane have been blown back down.

The mapping N induces a map $N_Q : \mathcal{Q}_\infty \rightarrow \mathcal{Q}_\infty$, which maps integral points to integral points. In \mathcal{Q}_1 , this is the map computed in Section 3.2:

$$\left\{ \begin{array}{ll} \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} 2p+q \\ 2p \end{pmatrix} & \text{in zone 1} \\ \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} p \\ p-q \end{pmatrix} & \text{in zone 2} \\ \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} p \\ q \end{pmatrix} & \text{in zone 3} \\ \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} q-p \\ q \end{pmatrix} & \text{in zone 4} \\ \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} 2p \\ 2p+q \end{pmatrix} & \text{in zone 5.} \end{array} \right. \quad (5.39)$$

We also computed in Equation (5.6) what it was on the quadrant corresponding to the blow-up at B_2 ; this and a symmetric computation at B_1 give

$$\begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} 2q \\ q-p \end{pmatrix} \text{ at } B_1, \quad \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} p-q \\ 2p \end{pmatrix} \text{ at } B_2. \quad (5.40)$$

A similar computation can be performed at A_1 and A_2 , which gives

$$\begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} p+q \\ 2p+q \end{pmatrix} \text{ at } A_1, \quad \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} p+2q \\ p+q \end{pmatrix} \text{ at } A_2. \quad (5.41)$$

Any other Farey blow-up at some point $\mathbf{z} \in X_k$ will map to a previous Farey blow-up at $N_k(\mathbf{z})$, and there are natural coordinates (p, q) in the domain quadrant and in the range quadrant. The map N_Q is then written

$$\left\{ \begin{array}{ll} \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} p \\ q \end{pmatrix} & \text{if } \mathbf{z} \text{ is not critical} \\ \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} 2p \\ q \end{pmatrix} & \text{if } \mathbf{z} \text{ is critical} \end{array} \right. \quad (5.42)$$

To see that this is compatible with Theorem 4.6.1, we need to know that if the blow-up is at an enriched point $\tilde{\mathbf{z}}$ that is critical, then the critical curve through \mathbf{z} is the component of D_k , not the enriching curve in the finite plane. Thus the attaching point $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ corresponds to the end of the parametrized curve $\begin{pmatrix} t^1 \\ t^0 \end{pmatrix}$, which tends as $t \rightarrow 0$ to a point of the critical curve.

Theorem 5.6.1. *The mapping $H_2 : X_2^*(X_\infty) \rightarrow H_2(X_\infty^*)$, viewed as a mapping $S_\infty \rightarrow S_\infty$, is given by the formula*

$$N_*(\mathbf{f}) = (N_Q)_*(\mathbf{f}|_{\mathcal{Q}'_\infty}) + 2(N_Q)_*(\mathbf{f}|_{\mathcal{Q}''_\infty}). \quad (5.43)$$

Proof. We will check that the formula is true for each $\mathbf{e}_{p/q, \mathbf{z}}$; there are some cases where it is easy, and several where it is quite complicated.

First, let us see why the 2 in the formula is necessary. Figure 48 illustrates what is going on.

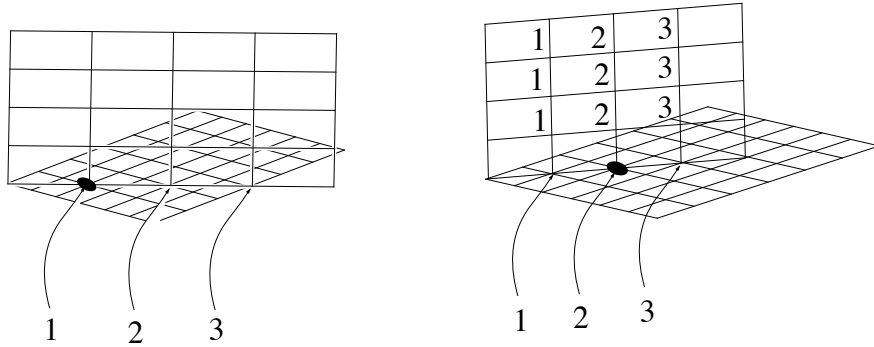


FIGURE 48. Left: a Farey quadrant Q_z attached to an earlier quadrant. Right: their images.

On the left we see a quadrant Q_z , attached to the quadrant Q of some earlier Farey blow-up. Note that since the component to which it is attached is critical, this earlier Farey blow-up itself was at a critical point. The marked point on the left, primitive on the ray to which the new quadrant is attached, is mapped to a non-primitive point in the quadrant $Q' = (N_Q)_*(Q)$, in fact twice a primitive point, as marked on the right.

On Q , we have a function (on the left) which is 1 at the primitive point (its values off the ray are irrelevant). It pushes forward, by $(N_Q)_*$ to the function which is 1 at the marked point on the right. But the image on the homology is twice that: it is the function which is 1 at the primitive point on the ray in the range. Thus to be continuous the function $e_{1/0,z}$ must map to twice the function $e_{1/0}$ of the image quadrant.

This argument proves a lot of Theorem 5.6.1, in fact, together with Theorem 4.6.1 and Theorem 5.4.5, it proves it for all $e_{p/q,z}$ with support only where N_Q is a local homeomorphism.

It is then clear that we only need to verify the formula for the exceptional cases:

$$e \begin{pmatrix} 1 \\ -1 \end{pmatrix}, e \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e \begin{pmatrix} 2 \\ 1 \end{pmatrix}, e \begin{pmatrix} 1 \\ 1 \end{pmatrix}, e \begin{pmatrix} 1 \\ 2 \end{pmatrix}, e \begin{pmatrix} 0 \\ 1 \end{pmatrix}, e \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{5.44}$$

This is quite tedious, and extremely prone to error. We will do it in detail only for $e \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $e \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and leave the others to the reader.

The case of $e \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Figure 49 illustrates what is happening in this case. We know from Equation (5.23) that

$$N_* \left(e \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) = e \begin{pmatrix} 1 \\ 2 \end{pmatrix} + e \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5.45}$$

From just the Q_1 , the formula

$$N_*(\mathbf{f}) = (N_Q)_*(\mathbf{f}|_{Q'_\infty}) + 2(N_Q)_*(\mathbf{f}|_{Q''_\infty}). \tag{5.46}$$

yields the discontinuous function on the right of the figure. The discontinuity comes from the fact that half the support of $e \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is in zone 1, but the other part (shaded) is in zone 2.

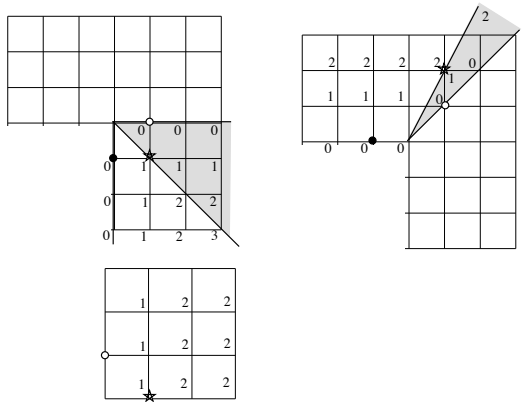


FIGURE 49. On the left, the primitive plane and the quadrant corresponding to the blow-up at A_1 . On the right, the primitive plane, viewed as the range.

However, the function $\mathbf{e} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ also has support in the quadrant corresponding to the blow-up of A_1 ; in fact it is the function $\mathbf{e}_{1/0}$ in that quadrant, as shown in the lower half of the picture. That quadrant also maps into the shaded region on the right, and the two together add to the correct value.

The case of $\mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Figure 50 illustrates what is happening in this case, which is quite a bit more complicated than the former one.

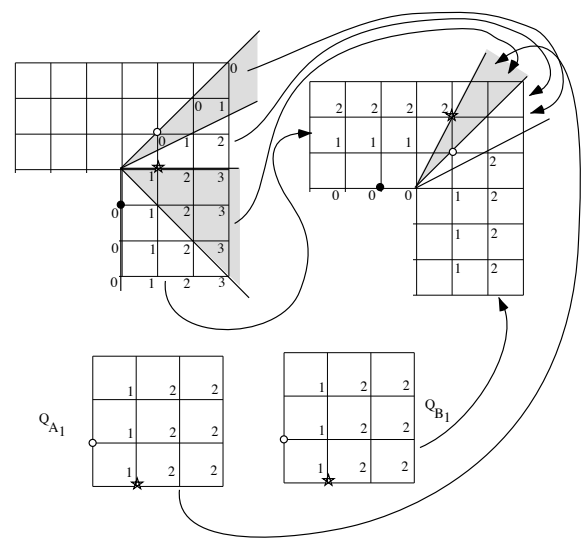


FIGURE 50. On the left, the primitive plane and the quadrant corresponding to the blow-ups at A_1 and B_1 . On the right, the primitive plane, viewed as the range.

Again, we know what we are supposed to find:

$$N_* \left(\mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) = \mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{5.47}$$

There are now six separate regions in the domain to think about. The part in zone 1 pushes forward to the restriction of $\mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in zones 4 and 5. The part in zone 2 with $q < 0$ pushes forward to half the restriction of $\mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to the part of zone 3 where $q > p$. The part in zone 2 where $q > 0$ and the part in zone 3 both push forward to the region in zone 3 where $q < p$, and there each contributes half to the restriction of $\mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to that region.

Since both A_1 and B_1 are in the support of $\mathbf{e} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, we need to think about what both of their quadrants contribute. The part in Q_{A_1} contributes the missing half in zone 3, $q > p$, and the part in Q_{B_1} contributes the restriction of $\mathbf{e} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to zones 1 and 2. \square

Since the graph J is embedded in \mathcal{Q}_∞ , and the restriction of a function $\mathbf{f} \in S_\infty$ to J determines the function by homogeneity, it is certainly possible to describe the mapping $N_* : H_2(X_\infty^*) \rightarrow H_2(X_\infty^*)$ as a mapping $\mathcal{H}^1(J_\infty) \rightarrow \mathcal{H}^1(J_\infty)$.

The fundamental difficulty in this program is that N_Q does not map J to J . In fact, it doesn't even map J_1 to J_1 . This introduces no difficulties in defining a map

$$N_J : J_\infty \rightarrow J_\infty, \quad N_J(p) = J \cap (\mathbb{R}_+ N_Q(p)), \tag{5.48}$$

which takes a point $p \in J$ and returns the intersection of the ray through the image of that point and J . Another way of saying this is to say that J_∞ is canonically homeomorphic to $\mathbb{P}^+(Q_\infty)$, the ‘‘projective space’’ of oriented rays in Q_∞ , and since N_Q is piecewise linear, it maps rays to rays, inducing the mapping

$$N_J = \mathbb{P}^+(N_Q) : \mathbb{P}^+(Q_\infty) \rightarrow \mathbb{P}^+(Q_\infty). \tag{5.49}$$

But the function space S_∞ is defined by restricting to J , and cannot be naturally defined on the quotient. This leads to problems with the push-forward of functions $(N_J)_*$. Let us illustrate this with an example.

Example 5.6.2. The line $L \begin{pmatrix} -4 \\ 1 \end{pmatrix}$ of X_1 maps to $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, and is critical. We have $N_Q \begin{pmatrix} -4 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$, and we want to understand what happens to the quadrants corresponding to the Farey blow-ups at \tilde{A}_1 and at its inverse image in $L \begin{pmatrix} -4 \\ 1 \end{pmatrix}$. Figure 51 illustrates the situation.

To repair this difficulty, we define a scaling function $\rho : J_\infty \rightarrow \mathbb{R}_+$ by the formula

$$\rho(p) = \frac{N_J(p)}{N_Q(p)}; \tag{5.50}$$

the division makes sense because the two points are on a common ray.

Now define a modified push-forward on functions on J by the formula

$$(N_J)_\times(\mathbf{f})(p) = \sum_{p' \in N_J^{-1}(p)} \rho(p') \mathbf{f}(p'). \tag{5.51}$$

This has evidently been cooked up to map S_∞ to S_∞ . Before we can describe the action of the Newton map in terms of $\mathcal{H}^1(J_\infty)$, we need one more bit of notation: Call

$$J'_\infty = J_\infty \cap Q'_\infty \quad \text{and} \quad J''_\infty = J_\infty \cap Q''_\infty. \tag{5.52}$$

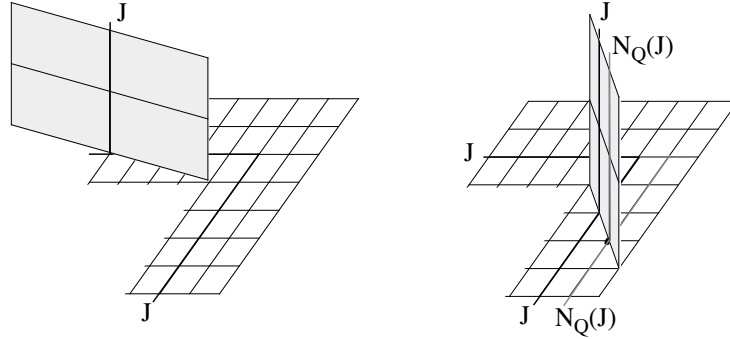


FIGURE 51. The mapping N_Q does not map J to J .

Theorem 5.6.3. *The mapping $N_* : H_2(X_\infty^*) \rightarrow H_2(X_\infty^*)$ is written, using the isomorphism $H_2(X_\infty^*)$ with S_∞ of Theorem 5.5.1, as*

$$N_*(\mathbf{f}) = (N_J)_\times(\mathbf{f}|_{J'}) + 2(N_J)_\times(\mathbf{f}|_{J''}). \tag{5.53}$$

Proof. This is precisely what Theorem 5.6.1 gives. \square

This has the following corollary.

Corollary 5.6.4. *The mapping $\hat{N}_* : \hat{H}_2(X_\infty^*) \rightarrow \hat{H}_2(X_\infty^*)$ is given on $\mathcal{H}^1(J_\infty)$ by the formula*

$$N_*(\mathbf{f}) = (N_J)_\times(\mathbf{f}|_{J'}) + 2(N_J)_\times(\mathbf{f}|_{J''}). \tag{5.54}$$

This allows us to understand \hat{N}_* as a bounded operator in a Hilbert space. The spectral analysis of this operator should be interesting: it should contain all the cohomological information about invariant $(1, 1)$ -currents. We intend to write another paper about this.

6

The case where a and b are arbitrary

In this chapter, we will see that essentially everything we have done in the case $a = b = 0$ in Chapters 3 and 5 is true for all a and b . The main difference is that whereas the x and y axes map to each other when $a = b = 0$, they no longer play any special dynamical role when a and b are arbitrary. The first step is to discover the appropriate curves to take the place of the x and y axes in the enriched points $\tilde{\mathbf{p}}_1$ and $\tilde{\mathbf{p}}_2$. It turns out that there are also curves of order two in this case, although the curves are transcendental, and we don't know much about their global behavior.

6.1. A curve of order two

We will now denote by N the mapping

$$N_F \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{4xy - 1} \begin{pmatrix} 2x^2y + y^2 - 2ay - b \\ 2xy^2 + x^2 - 2xb - a \end{pmatrix}. \quad (1.34), \text{ again}$$

We will assume that a and b are arbitrary.

Theorem 6.1.1. *There exist $R > 0$ and unique analytic mappings*

$$\gamma_1(t) = \begin{pmatrix} t + o(t) \\ \frac{a}{2t^2} + o\left(\frac{1}{t^2}\right) \end{pmatrix} \quad \text{and} \quad \gamma_2(t) = \begin{pmatrix} \frac{b}{2t^2} + o\left(\frac{1}{t^2}\right) \\ t + o(t) \end{pmatrix} \quad (6.1)$$

defined for $|t| > R$, such that

$$N(\gamma_1(t)) = \gamma_2(t^2) \quad \text{and} \quad N(\gamma_2(t)) = \gamma_1(t^2). \quad (6.2)$$

Proof. The following lemma will be the principal tool. Let $U_R = \{|z| \geq R\}$ and let E_R be the Banach space of bounded continuous functions on U_R , analytic on the interior. We will denote by $\|f\|_R$ the sup norm on this Banach space. We will denote by $\Gamma(f)$ and $\Gamma'(f)$ the curves of equation $y = f(x)$ and $x = f(y)$.

Lemma 6.1.2. *There exist $R > 0$ and constants $0 < K_1 < K_2$ (depending on a and b) such that for any function $f \in E_R$ with $\|f\| < K_2$, there exists a unique $\tilde{f} \in E_R$ with $\|\tilde{f}\|_R < K_1$ such that*

$$N^{-1}(\Gamma(f)) \cap \{|y| \geq R, |x| < K_2\} = \Gamma'(\tilde{f}). \quad (6.3)$$

Proof of Lemma 6.1.2. Looking at the dominant terms of the formula for N , we see that for all $M > 0$, there exist R and K such that if we set

$$A_{M,R} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid |y| \geq R, x \leq \frac{M}{|x|^2} \right\}, \quad (6.4)$$

then N maps $A_{M,R}$ to a neighborhood of $\mathbf{p}_1 = [1 : 0 : 0]$ by a mapping of the form

$$N : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto - \begin{pmatrix} y^2 + \alpha(x, y) \\ 2xy^2 - a + \beta(x, y) \end{pmatrix} \quad \text{with} \quad |\alpha(x, y)| \leq K|y|, |\beta(x, y)| \leq \frac{K}{|y|}. \quad (6.5)$$

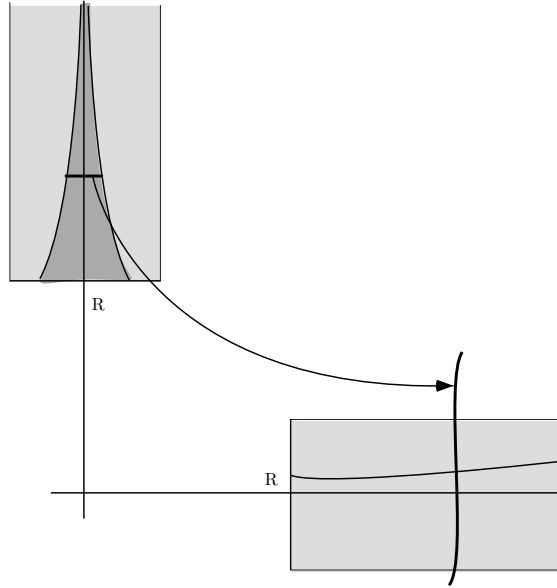


FIGURE 52. This shows how the regions in Equation (6.4) are situated.

It follows that any horizontal disc $y = c, |x| \leq M/|c|^2$ maps injectively to an almost vertical disc, with x -coordinate approximately c^2 , and with y -coordinate approximately filling the disc of radius $2M$ around a . Choose $M = |a| + 1$, so that this disc will contain the disc of radius M around 0 for R sufficiently large.

Then such a disc will intersect $\Gamma(f)$ in a single point if $\|f\|_R < M$, and we see the $N^{-1}(\Gamma(f))$ contains an analytic curve in the interior of $A_{M,R}$ that intersects every analytic disc in a single point. Such a curve is $\Gamma'(\tilde{f})$ for a well-defined \tilde{f} that depends analytically on f . To complete the proof of the lemma, we must show that there exists $M' < M$ such that $\|\tilde{f}\|_R$. This is clear from the definition of $A_{M,R}$: the maximum of $|x|$ over $A_{M,R}$ is M/R^2 , and this will be less than M if $R > 1$. \square

Proof of Theorem 6.1.1 from Lemma 6.1.2. Consider the map $(f, g) \mapsto (\tilde{g}, \tilde{f})$ from the product of the ball of radius C_2 with itself in $E_R \times E_R$, to itself. Since the image is in fact in the product of balls of radius C_1 , this is an analytic map of a ball in a Banach space strictly inside itself. As such, it is strongly contracting and has a unique fixed point $(F, G) \in E_R \times E_R$. Now $N^{\circ 2} : \Gamma(F) \rightarrow \Gamma(F)$ is a map with a super attractive fixed point: $\Gamma(F)$ is parametrized by x , and the map is of the form $x \mapsto x^4 + \dots$. Thus there is a Böttcher coordinate t and in fact a unique one if we impose that $t \sim x$.

The same argument applies to $\Gamma'(G)$, and the two Böttcher coordinates solve our problem. \square

6.2. The primitive space for arbitrary a and b .

Call C_i the curve that is the image of γ_i . Further, define the Farey data $Z_1 = (\mathbf{p}_1; l_\infty, C_1)$ and $Z_2 = (\mathbf{p}_2; l_\infty, C_2)$. Set X_1 to be the Farey blow-up

$$X_1 = \mathcal{F}(\mathbb{P}^2; Z_1, Z_2). \tag{6.6}$$

Theorem 6.2.1. *The Newton map, defined on \mathbb{P}^2 with the five points of indeterminacy removed, extends to a mapping $N_1 : X_1 \rightarrow X_1$, with seven points of indeterminacy.*

Proof. When $a = b = 0$, we proved this using explicit computations. We could have proved it using the naturality statement 4.2.2, and we will do so here, as any explicit computation would require knowing the equations of the curves C_1 and C_2 .

We must make a number of preliminary blow-ups first, to be in a position to apply Theorem 4.2.2.

Blow-up \mathbf{p}_1 to create the exceptional divisor $L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; then again at the intersection with the proper transform of C_1 , to create $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$; then one more time, to create $L \begin{pmatrix} 1 \\ -2 \end{pmatrix}$. In the range, we blow up once at \mathbf{p}_2 , to create $L \begin{pmatrix} 0 \\ 1 \end{pmatrix}$; and once at the intersection of $L \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $L \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then the Newton map is well defined on a neighborhood of $L \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, mapping it with normal degree 2 to $L \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The situation is drawn in Figure 53: our one-term asymptotic expansion (6.1) of the equation of C_1 guarantees that after these blow-ups, the proper transform of C_1 does intersect $L \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, and not at the point where $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ intersects it.

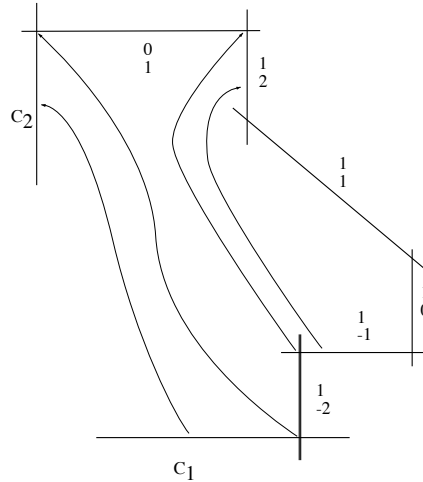


FIGURE 53. The configuration needed to apply Theorem 4.2.2.

We can now use Theorem 4.2.2 to extend the Newton mapping to Farey blow-ups at the points of intersection of $L \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ with C_1 and with $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, and going to the Farey blow-ups at the double points on $L \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Near \mathbf{p}_1 , we still need to extend the Newton map to the Farey blow-ups of the intersections $L \begin{pmatrix} 1 \\ -1 \end{pmatrix} \cap L \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $L \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cap L \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. At the first intersection, we can use Theorem 4.2.2 again, or claim that the computations of Chapter 3 go through without modification, as the terms contributed by a and b are never the dominant term in that region. The second strategy also works at $L \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cap L \begin{pmatrix} 1 \\ 1 \end{pmatrix}$;

if we wish to apply the first, we will need to do another preliminary blow-up, since one of the critical curves goes to this point, and is not part of the enriching structure.

A completely symmetric construction will extend the mapping near \mathbf{p}_2 . \square

6.3. The Russakovskii measure for arbitrary a and b

Our next task is to see that Corollary 3.3.3 is still true when a and b are arbitrary.

Theorem 6.3.1. *For all a, b , the Russakovskii-Shiffman measure for $N_{a,b}$ does not charge the points \mathbf{p}_1 and \mathbf{p}_2 .*

Proof. We will not repeat the proof in detail, as it is almost identical with the proof in Chapter 3. We will simply comment on that proof, showing where minor modifications need to be made.

The first step is to see that the “Markov” decomposition can still be defined, i.e., that Lemma 3.3.1 is still true, and there is no difficulty there. Theorem 2.3.1 was made without restrictions, and it asserts that all points of the lines $L \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ and $L \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ have unstable manifolds. Unions of these

manifolds and their inverse images were used to construct the boundaries M_k of the regions A_k .

One has to look more carefully to make sure that Lemmas 3.3.4, 3.3.5, 3.3.6, and 3.3.7 are still true. For the first three, there is no problem, since in the corresponding regions A_k , $-3 \leq k \leq 3$, the asymptotics of the particular map used there coincide with the asymptotics in the general case. As for Lemma 3.3.7, the only way the asymptotics of the mapping N were used was that in those regions (A_k , $|k| > 3$) either the x^2 in the second coordinate (near \mathbf{p}_1) or the y^2 in the first coordinate (near \mathbf{p}_2) was the dominant term of that coordinate function. That is certainly still true; in fact it is the main thing used in the proof of Theorem 6.1.1.

Thus the entire argument goes through without modifications, except to adjust the regions A_k to the dynamics at hand. \square

6.4. Building the space X_∞

To complete our program, we need to see that for any a, b we can build a compact space X_∞ to which $N_{a,b}$ extends, and to understand its topology. As we will see, the general case behaves just like the special case we studied.

We already have the space X_1 , with the four points of indeterminacy of $(N_{a,b})_1 : X_1 \rightsquigarrow X_1$. We will call these points as above A_1, B_1, B_2, A_2 . We will enrich them, forming $\tilde{A}_1, \tilde{B}_1, \tilde{B}_2, \tilde{A}_2$, in the more or less obvious way: A_1 and A_2 are enriched as before, A_1 by the lines $L \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and the polar curve, and A_2 by $L \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and the polar curve. The point B_1 is enriched by $L \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and the inverse image of C_1 that touches B_1 ; B_2 is enriched by $L \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and the inverse image of C_2 that touches B_2 .

Now the Farey data \tilde{Z}_1 is formed of $\tilde{A}_1, \tilde{B}_1, \tilde{B}_2, \tilde{A}_2$, together with $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$, and $X_2 = \mathcal{F}(X_1, \tilde{Z}_1)$.

The remainder of the construction of X_∞ is identical to that in Chapter 5.

6.5. The basins of the roots

A first statement concerns the basins of the roots. They have an obvious definition in X_∞ , simply as the set of points that are attracted to the roots under N_∞ . These basins did not have a nice structure in \mathbb{C}^2 , because of the points of indeterminacy.

Theorem 6.5.1. *In X_∞ , the basins of the roots are connected and holomorphically convex. If they contain no enhanced points, they are Stein.*

Proof. Let \mathbf{a} be a root, $U_{\mathbf{a}}$ be its basin, and $G_{\mathbf{a}} : U_{\mathbf{a}} \rightarrow \mathbb{R} \cup \{-\infty\}$ be the potential function of the root, as defined in [HP]:

$$G_{\mathbf{a}}(\mathbf{x}) = \lim_{n \rightarrow \infty} \frac{1}{2^n} \log \|N_\infty^{on}(\mathbf{x}) - \mathbf{a}\|. \quad (6.7)$$

Clearly the basins of the roots are entirely made up of smooth points of X_∞ , since all such points are mapped by an iterate of N_∞ to a small neighborhood of the roots that is an open subset of \mathbb{C}^2 . The key point is that $G_{\mathbf{a}}$ is proper pluri-subharmonic function, i.e., that for all $k > 0$, the set where $G_{\mathbf{a}}(\mathbf{x}) \leq -k$ is compact (this is not true in \mathbb{C}^2).

First, the function $G_{\mathbf{a}}$ is continuous on $U_{\mathbf{a}}$, except at the root itself, because the root has no inverse images other than itself. Moreover, there exists a compact neighborhood V of \mathbf{a} and a constant $K > 0$ such that $G_{\mathbf{a}}(\mathbf{x}) > -K$ if $\mathbf{x} \notin V$.

Choose $k > 0$, and a sequence \mathbf{x}_n satisfying $G_{\mathbf{a}}(\mathbf{x}_n) \leq -k$. We can certainly extract a convergent subsequence in X_∞ , converging to some point \mathbf{y} . If \mathbf{y} is in the basin of \mathbf{a} , we are done by the continuity of $G_{\mathbf{a}}$. Otherwise, \mathbf{y} must be in the boundary of the basin. But then, for any $m > 0$ we can find $n > 0$ such that $N_\infty^{om}(\mathbf{x}_n) \notin V$, since

$$\lim_{m \rightarrow \infty} N_\infty^{om}(\mathbf{x}_n) = N_\infty^{om}(\mathbf{y}) \quad (6.8)$$

is in the boundary of the basin. That implies that $G_{\mathbf{a}}(\mathbf{x}_n) > -K/2^m$, and contradicts $G_{\mathbf{a}}(\mathbf{x}_n) \leq -k$ if $2^m k > K$.

This proves that the basins are holomorphically convex; to see that they are Stein, we must show that the global functions separate points. There is no problem for pairs of points that do not belong to the same component of some E_n : the projection to \mathbb{C}^2 will then separate the points. Neither is there a problem if both points belong to a single exceptional divisor: just so long as that divisor is mapped to one of the double tangents by some iterate of N , then $\pi_\infty \circ N^k$ will separate the points. But if some exceptional divisor is collapsed to a point in the interior of a basin, which may happen as far as we know, then the global functions will of course not separate points on that exceptional divisor. \square

Remark. In \mathbb{P}^2 , sequences in a basin can approach points of indeterminacy, without the potential there tending to 0. \triangle

This gives some information about the complex analytic structure of the basins in \mathbb{C}^2 .

Theorem 6.5.2. *The basins of the roots in \mathbb{C}^2 are Stein manifolds.*

Proof. Let us first consider the case where there are no enhanced points in the basin $U_{\mathbf{a}}$ of a root \mathbf{a} . Then the basin in \mathbb{C}^2 is the complement of $\cup_n E_n$ in $U_{\mathbf{a}}$, and the complement of a closed hypersurface in a Stein manifold is Stein.

Actually, this argument also works if there are enhanced points. We must first blow down all the projective lines that are blown down by some iterate of N . We then obtain a Stein space, which has singularities since even in the simplest case the lines being blown down have self-intersection -2 ; but this space is holomorphically convex, and the global functions do separate points.

Again in this space, the basin in \mathbb{C}^2 is the complement of $\cup_n E_n$, which is a closed hypersurface. \square

6.6. Real oriented blow-ups and homology

All the statements about the real oriented blow-up of X_∞ along D_∞ , and all the statements about the homology of X_∞^* and the action of the Newton mapping on the homology, are valid in the case of a, b arbitrary with no changes.

Bibliography

- [Ar] V.I. Arnold, *Small divisor problems in classical and celestial mechanics*, Usp. Mat. Nauk 18, no. 6 (114), 91–192 (1963).
- [B] E. Bedford, *Iteration of polynomial automorphisms of \mathbb{C}^2* , Proc. Int. Cong. Math. Kyoto, (1990), 847–858.
- [BS] E. Bedford and J. Smillie, *Fatou-Bieberbach domains arising from polynomial automorphisms*, Indiana Math. Jour., 40 (1991), 789–792.
- [BS1] E. Bedford and J. Smillie, *Polynomial diffeomorphisms of \mathbb{C}^2 I: Currents, equilibrium measure and hyperbolicity*, Invent. Math., 87 (1990), 69–99.
- [BS2] E. Bedford and J. Smillie, *Polynomial diffeomorphisms of \mathbb{C}^2 II: Stable manifolds and recurrence*, J Amer. Math. Soc., 4 (1991), 657–679.
- [BS3] E. Bedford and J. Smillie, *Polynomial diffeomorphisms of \mathbb{C}^2 III: Ergodicity, exponents and entropy of the equilibrium measure*, Math. Ann., 294 (1992), 395–420.
- [BS5] E. Bedford and J. Smillie, *Polynomial automorphisms of \mathbb{C}^2 V: Critical points and Lyapunov exponents*, Journal of Geometric Analysis 8,3 (1998).
- [BS6] E. Bedford and J. Smillie, *Polynomial diffeomorphisms of \mathbb{C}^2 VI: Connectivity of J* , Ann. Math., 148(1998), 695–735.
- [BS7] E. Bedford and J. Smillie, *Polynomial diffeomorphisms of \mathbb{C}^2 VII: Hyperbolicity and external rays*. Annales Scient. E.N.S. Paris 32(1999), 455–497.
- [BS8] E. Bedford and J. Smillie, *External rays in the dynamics of polynomial diffeomorphisms of \mathbb{C}^2* , Contemporary math. 222(1999), 41–79.
- [BCSS] L. Blum, F. Cucker, M. Shub, S. Smale, *Complexity and real computation*, Springer, New York (1998).
- [Bou] T. Bousch, *Quelques problèmes de dynamique holomorphe*, Thesis, University of Paris at Orsay, 1992
- [Br] J.Y. Briend, *Exposants de Liapounof et points périodiques d'endomorphismes holomorphes de $\mathbb{C}P^k$* , Thesis, Toulouse (1997).
- [De] P. Deligne, *Intersections sur les surfaces régulières*, SGA 7, Exposé 10, IHES (1969).
- [Dil] J. Diller, *Dynamics of birational maps of \mathbb{P}^2* , Ind. J. Math. 45.3 (1996), 721–772.

- [Dou] R. and A. Douady, *Algèbre et Théorie Galoisienne, I and II*, Cedic/Fernand Nathan, Paris 1977, 1979.
- [FS1] J. Fornæss and N. Sibony, *Complex Dynamics in Higher Dimension I*, Asterisque 222 (1994), 201-231.
- [FS2] J. Fornæss and N. Sibony, *Complex Dynamics in Higher Dimension II*, Ann. Math. Studies 137 (1995), 134-182.
- [FS3] J. Fornæss and N. Sibony, *Complex Dynamics in Higher Dimension*, in Complex Potential Theory (1995), Kluwer Academic Publishers, 131-186.
- [Fo] J. Fornæss, *Dynamics in several complex variables*, CBMS 87
- [Fr1] S. Friedland, *Entropy of polynomial and rational maps*, Ann. Math, 133 (1991), 359-368.
- [Fr2] S. Friedland, *Entropy of algebraic maps*, J. Fourier Anal. and Appl.(1995), 215-228
- [GH] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, Wiley, N.Y., 1978.
- [Har] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag.
- [Hat1] A. Hatcher, *Topology*, Cambridge University Press (2000).
- [HMS] M. Hirsch, J. Marsden, M. Shup, *From topology to computation: Proceedings of the Smale-fest*, Springer (1993).
- [Hirz] F. Hirzebruch, *Hilbert Modular Surfaces*, l'Ens. Math. 19 (1973), 183-281.
- [HNK] F. Hirzebruch, W.D. Neumann and S.S. Koh, *Differentiable manifolds and quadratic forms*, Marcel Dekker, N.Y. 1971.
- [HBH] J. H. Hubbard and B.B. Hubbard *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*, Prentice Hall, Upper Saddle River, N.J., 1999.
- [Hu] J.H. Hubbard, *The Hénon mapping in the complex domain*, in *Chaotic Dynamics and Fractals*, Barnsley and Demko, eds., Academic Press, New York (1986), 101-111.
- [HO1] J.H. Hubbard and R.W. Oberste-Vorth, *Hénon mappings in the complex domain I: The global topology of dynamical space*, Pub. Math. IHES 79 (1994), 5-46.
- [HO2] J.H. Hubbard and R.W. Oberste-Vorth, *Hénon mappings in the complex domain II: Projective and inductive limits of polynomials*, in *Real and Complex Dynamical Systems*, Branner and Hjorth, eds., Kluwer Academic Publishers (1995), 89-132.
- [HP] J.H. Hubbard and P. Papadopol, *Superattractive fixed points in \mathbb{C}^n* , Ind. J. Math. 43.1 (1994), 321-365.
- [HPV1] J.H. Hubbard, P. Papadopol and V. Veselov, *A Compactification of Hénon Mappings in \mathbb{C}^2 as Dynamical Systems*, SUNY StonyBrook Institute for Mathematical Sciences, Preprint 1997/11 (97 pages).
- [HPV] J.H. Hubbard, P. Papadopol, and V. Veselov, *A Compactification of Hénon Mappings in \mathbb{C}^2 as Dynamical Systems*, Acta Math., to appear.

- [HSS] J.H. Hubbard, D. Schleicher, and S. Sutherland, *How to really find roots of polynomials by Newton's method*, preprint, submitted.
- [Jo] M. Jonsson, *Dynamics of polynomial skew products on \mathbb{C}^2* , Math. Ann. 314 (1999), no. 3, 403–447.
- [KH] A. Katok, B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge U. Press (1995).
- [MSS] R. Mane, P. Sad, D. Sullivan, *On the dynamics of rational maps*, Ann. Scient. E.N.S. Paris, 16 (1983), 193–217.
- [Ma] Y. Manin, *Cubic forms*, North Holland (1974).
- [Mi1] J. Milnor, *Singular points of complex hypersurfaces*, Ann. Math. Stud. 61, Princeton University Press (1968).
- [Mi2] J. Milnor, *Dynamics in one complex variable*, Springer (1999).
- [Pe] Y. Pesin, *Characteristic Liapounov exponents and smooth ergodic theory*, Russ. Math. Surveys, 32 (1977), 55–114.
- [PS] C. Pugh and M. Shub, *Ergodic attractors*, TAMS 312 (1989), 1–54.
- [RS] Russakovskii A., and Shiffman, B., *Value distribution for sequences of rational mappings and complex dynamics*, Ind., Jour. Math. 46 (1997), 897–932.
- [Se] O. Sester, *Dynamique des polynômes fibrés*, Thesis, University of Paris at Orsay, 1996.
- [S1] I.R. Shafarevic, *Algebraic surfaces*, Proc. Stek. Inst. Math., 75 (1965), AMS Providence (1967).
- [S2] I.R. Shafarevic, *Basic Algebraic Geometry, I&II*, Springer Verlag, NY 1994.
- [Sh1] M. Shub, *Global stability of dynamical systems*, Springer (1987).
- [ShSm] M. Shub and S. Smale, *Computational complexity: On the geometry of polynomials and a theory of cost I.*, Ann. Sci. Ecole Norm. Sup. 18 (1985), 107–142.
- [Si] N. Sibony, *Dynamique des applications rationnelles de \mathbf{P}^k* , preprint 1998.
- [Sm1] S. Smale, *A convergent process of price adjustment and global Newton methods*, J. Math. Economy 3 (1976), 107–120.
- [Sm2] S. Smale, *The fundamental theorem of algebra and complexity theory*, Bulletin of the Amer. Math. Soc. 4 (1981), 1–36.
- [Sm3] S. Smale, *Algorithms for solving equations*, Proceedings of the International Congress of Mathematicians (1986), pp. 172–195, American Mathematical Society.
- [Sm4] S. Smale, *Newton's method estimates from data at one point*, in R. Ewing, K. Gross, and C. Martin (eds.), *The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics* (1986). Springer-Verlag.
- [Spa] E. Spanier, *Algebraic topology*, McGraw Hill, N.Y., 1966.

- [Or] P. Orlik, *Seifert manifolds*, LNM, Nr. 291, Springer, 1972.
- [Ueda] Ueda, T., *Fatou sets in complex dynamics in projective spaces*, J. Math. Soc. Japan 46 (1994), 545-555.

JOHN H. HUBBARD
Department of Mathematics
Malott Hall
Cornell University
Ithaca, NY 14850

and

LATP
39 Rue Joliot Curie
Université de Provence
Marseille CEDEX 13453

PETER PAPADOPOL
Dynamical Systems Laboratory
Grand Canyon University
Phoenix, AZ 85047

and

Department of Mathematics
Malott Hall
Cornell University
Ithaca, NY 14850