

A Real-Analytic Jordan Curve Cannot Bound Infinitely Many Relative Minima of Area

Michael Beeson
Department of Mathematics
San Jose State University
San Jose, California 95192
USA
email: beesonpublic@gmail.com

May 20, 2015

Abstract

Let Γ be a real-analytic Jordan curve in R^3 . Then Γ cannot bound infinitely many disk-type minimal surfaces that provide relative minima of area.

1 Introduction

Minimal surfaces are mathematical objects that are intimately related to the physical surfaces formed by thin soap films. Aside from the idealization to zero thickness, there are some other differences between physical soap films and minimal surfaces. First, physical soap films are *stable*, in the sense that if they are disturbed slightly, they regain their shape. Secondly, physical soap films sometimes have internal edges where several films meet, or other kinds of more complicated topology. Here we are concerned with the classical *problem of Plateau*, which requires, given a Jordan curve Γ , to find (or at least prove the existence of) a surface of the topological type of the disk bounded by Γ and minimizing area among such surfaces. Surfaces forming a *relative minimum* of area correspond to soap films stable in the physical sense just described.

Plateau's problem was solved independently by Douglas and Rado in the early 1930s. Their solution methods produced minimal surfaces that might have certain singularities known as *branch points*. It was not until the seventies that the regularity (lack of branch points) of solutions providing an absolute minimum of area with given real-analytic boundary was proved. Readers not already intimately familiar with branch points may want to view the animated pictures posted to the Web at [7] for a visual introduction to branch points. One picture of a boundary branch point is given here in Figure 1.

The “finiteness problem” addressed in this paper is to prove that for a given Jordan curve Γ , only finitely many disk-type soap films are bounded by Γ . The reason that minimal surfaces are called “minimal” is that a soap film furnishes a relative minimum of area; that is, perturbing the surfaces slightly will not decrease the area. Thus the mathematical form of the finiteness problem is to prove that a given Jordan curve Γ cannot bound infinitely many relative minima of area. We prove that theorem in this paper (for real-analytic boundaries and disk-type surfaces).

Under those same assumptions, Tomi proved [34] that there cannot exist infinitely many *absolute minima* of area. But generally there will exist many relative minima of area that are not absolute minima, so Tomi’s theorem does not rule out infinitely many physical soap films. Tromba proved [36] *generic finiteness*; that is, the set of boundary curves for which finiteness holds is open and dense in a suitable topology. Here we prove a theorem about *every* real-analytic Jordan boundary. The finiteness problem for relative minima that we solve here is a long-standing open problem.¹

Branch points of minimal surfaces not furnishing a relative minimum of area are also of considerable interest, because they are intimately connected with the finiteness problem. It follows (as shown in [2], p. 121) from the work of Böhme [9] that if a real-analytic Jordan curve Γ bounds infinitely many disk-type relative minima of area, then it bounds an analytic one-parameter family of such minimal surfaces, all with the same Dirichlet’s integral, which can be continued until it either loops or terminates in a branched minimal surface. In [34], it is proved that a loop of disk-type minimal surfaces contains one which is not a relative minimum of area. Therefore, to prove that Γ does not bound infinitely many relative minima of area, we only have to rule out the possibility that Γ bounds a one-parameter family of minimal surfaces $u(t, z)$, terminating in a branched minimal surface when the parameter t is zero, and furnishing a relative minimum of area for each positive t . Tomi solved the finiteness problem for absolute minima of area and real-analytic boundaries [34] in this way, since the limit of absolute minima of area is again an absolute minimum, and hence has no branch points. But the problem for relative minima has remained open. (A discussion of other open problems in this area is given at the end of this paper.)

Suppose we have a boundary branch point; orient the boundary so it is tangent to the X -axis with the branch point at origin. Consider the unit normal N restricted to the boundary. As the parameter t varies, we do not know *a priori* how N behaves, but at least it must remain perpendicular to the boundary. Close to the branch point it must almost lie in the YZ plane. Some of our paper contains calculations concerning the possible behavior of N as the parameter t goes to zero.

The finiteness problem for relative minima of area was first attacked in [2], where interior branch points arising as the limit of one-parameter families of

¹See e.g. [29], pages 252 and 379, for discussion and many references. Although the date of that book is 1989, there has been no progress on finiteness since then. See also [10] for a somewhat more detailed discussion.

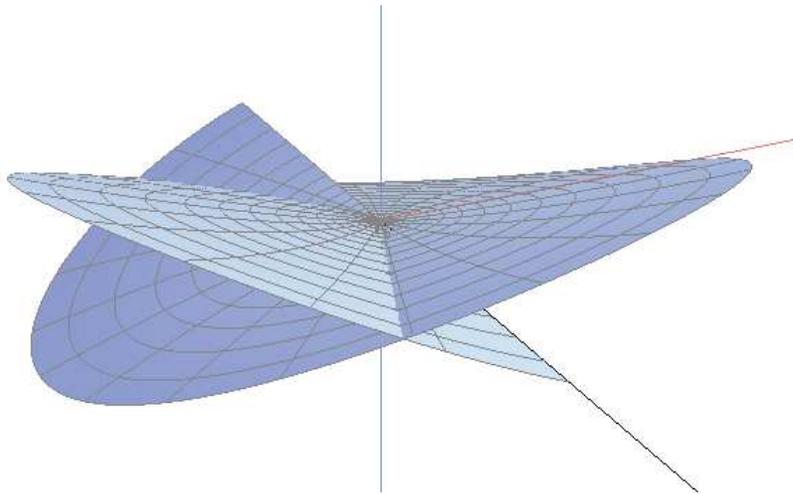


Figure 1: This is a boundary branch point of order 2 and index 1. A boundary branch point of order $2m$ goes “around” $(2m + 1)/2$ times while going “up and down” $(2m + k + 1)/2$ times. Could a one-parameter family of surfaces, each of which is a relative minima of area, and all with the same Jordan boundary, end in a surface like this?

relative minima were shown not to exist, and in [4], such boundary branch points were eliminated under special conditions on Γ , but not in general; so the finiteness problem for relative minima remained open until now, except for special boundaries.

It is well known that there is a natural eigenvalue problem associated to the second variation of the area functional. That problem, and its first eigenvalue and the associated eigenfunction, play an important role in our work. The main line of attack in [2] is an eigenvalue argument, that in the case of an interior branch point, the Gaussian image of u for small positive t must include more than a hemisphere, and hence u cannot be a relative minimum of area. The reason why it must include more than a hemisphere is the Gauss-Bonnet-Sasaki-Nitsche formula, which expresses the total boundary curvature as a sum of contributions from the Gaussian area and the branch points. This formula implies that as soon as $t > 0$, the contribution of the branch point term to the total curvature when $t = 0$ must be made up by “extra” Gaussian area totaling $2m\pi$, where $2m$ is the order of the boundary branch point. For boundary branch points, this argument implies that for small positive t , there cannot be a neighborhood in the parameter domain on which the Gauss map covers more than the upper hemisphere. Near the boundary the normal is confined to within $O(t)$ of the YZ plane, and it seems plausible that the extra Gaussian area is contributed in the form of m hemispheres. However, that is difficult to prove. The behavior of N could, *a priori*, be quite complicated.

The main line of attack in this paper is the calculation of the eigenfunction

corresponding to the first eigenvalue. The starting point of this calculation is the relation between the second variations of area and of Dirichlet's integral; namely, the condition for ϕ to belong to the kernel of the second variation of area is given by $\phi = k \cdot N$, where k belongs to the kernel of the second variation of area.² The "tangent vector" $k = u_t$ can be shown to be a "forced Jacobi direction" associated with the branch point; these are purely tangential, so as a function of the parameter t , the eigenfunction goes to zero. We calculate how the eigenfunction behaves on the w -plane, that is, on a region near the origin that shrinks to zero as t^γ . The leading term in t of this function is a harmonic function in the w -plane which inherits from the eigenfunction the property that it must have only one sign in the upper half plane. Using this property, we are able to derive additional information about the behavior of N near the origin.

Let $g(z)$ be the stereographic projection of the unit normal N to u^t . Then the poles of g , say $a_i(t)$, are the places where $N(z)$ is the "north pole" $(0, 0, 1)$. We analyze the way in which the $a_i(t)$ approach zero as t goes to zero. These go to zero asymptotically in this way: $a_i(t) = \alpha_i t^\gamma$, for some γ . Possibly some of the a_i go to zero faster than others; we group the a_i into "rings", each ring containing the roots that go to zero with the same power of t . In the vicinity of (some of) these a_i the Gauss map of u^t for $t > 0$ must contribute enough Gaussian area to make about $2m\pi$ "extra" Gaussian area, which will disappear in the limit to account for the branch point term in the Gauss-Bonnet-Sasaki-Nitsche formula. Consider the behavior of the normal N on the boundary near the point where the branch point is when $t = 0$. The normal must be perpendicular to the boundary, so it is confined to the vicinity of the YZ -plane if we take Γ to be tangent to the X -axis. But its behavior there might be quite uncontrollable, and moreover, several of the a_i can be associated with the same α_i .

In this argument there is another difficulty: the terms calculated for the eigenfunction in the w -plane might all cancel out, leaving the true eigenfunction hidden in the error terms. This is the case we call " \mathbb{H} constant", since the eigenfunction is a power of t times the imaginary part of a complex-valued function we call \mathbb{H} . One of the main difficulties is to rule out this case. We can pick any "ring" of roots that go to zero as t^{γ_i} , and look at the w -plane where $z = t^{\gamma_i} w$. We call the " N -condition" the condition that the unit normal should converge to the south pole $(0, 0, -1)$ uniformly away from the values of w where the normal takes on the north pole. That is true on the outer ring. One key to our argument is to show that if \mathbb{H} is constant on the outer ring, then the N -condition holds on the next ring, and moreover, there *is* a next ring if \mathbb{H} is constant—not all the roots can go to zero as t^{γ_i} if \mathbb{H} is constant on the t^{γ_i} ring. Once we have shown that, then on the innermost ring, \mathbb{H} cannot be constant, since there is no next ring. To prove this, let the β_i be the places in the w -plane where the normal takes on the south pole, and the α_i the places where it takes

²Note that ϕ is a scalar, because with area it is enough to consider normal variations given by $u + \phi N$. On the other hand k is a vector since the second variation of Dirichlet's integral is defined as a Frechet derivative in the space of (vectors defining) surfaces u . The theorem that $\phi = k \cdot N$ is proved in [2].

on the north pole. We show that the N -condition on all rings follows if on each ring, there are more β_i than α_i . The explicit formula for \mathbb{H} that applies when \mathbb{H} is constant shows first of all that there *is* another ring of roots going to zero faster, and second, it allows us to propagate the N -condition inwards to the next ring.

Examining the formula for \mathbb{H} near the origin of the w -plane shows that, if \mathbb{H} is not constant on the t^γ ring, then there are no roots going to 0 faster than t^γ , so this is the innermost ring. Now we know that the N -condition holds on every ring of roots, so the formula for the eigenfunction in the w -plane holds in each of the different w -planes (one for each different γ_i). At that point the Gauss-Bonnet theorem can be used: There are altogether m roots a_i where the normal takes on the north pole. We show (using the formula for the eigenfunction in the w -plane), applied to *each ring* of roots, that each a_i contributes at most one hemisphere of extra Gaussian area (i.e., the Gaussian area for t positive is $2m\pi$ more than when $t = 0$). But then, there must really be m of the a_i , and not some smaller number of a_i . From that it follows that the branch points s_i do not actually occur.

At that point, we know that on the innermost ring, \mathbb{H} is not constant; the rest of the proof works only with that innermost ring. In order to finish the proof, we have to analyze the behavior of \mathbb{H} at infinity. Since its imaginary part $\text{Im } H$ is the limit of the first eigenfunction, it has only one sign in the upper half-plane. That means that many terms in the eigenfunction formula must cancel out. These cancellations are very difficult to analyze directly. In the last part of the proof, we exhibit a differential equation that is satisfied by the function stereographic projection of the unit normal, involving the eigenfunction \mathbb{H} . This equation is the key to the analysis of \mathbb{H} near infinity (in the w -plane).

There is an appendix giving a “Dictionary of Notation”, in which we list the symbols that are used in more than one section of the paper, and give a reminder of their definitions, as an aid to the reader.

We note in passing that as of yet, nobody has produced an example of a real-analytic Jordan curve bounding a minimal surface with a boundary branch point, although Gulliver has exhibited in [20] a surface with a C^∞ boundary and a boundary branch point. The experts I have asked all believe that they can exist, but an example is still missing. Also, it is not known whether Gulliver’s example is area-minimizing or not.³

I am grateful to Fritz Tomi for his careful reading of several early versions of this paper.

³Wienholtz has used the solution of Björling’s problem ([16], page 124) with boundary values $\Gamma(t^3)$ where Γ is a regular parametrization of a real-analytic arc, and a suitably specified normal, to produce examples of a minimal surface partially bounded by a real-analytic arc which is not a straight line segment, with a branch point on the arc.

2 Outline of the Proof

In this section, we list the main steps of the proof for reference. The work of Böhme and Tomi reduces the task to ruling out the existence of a one-parameter family of relative minima of area bounded by a real-analytic Jordan curve Γ , running into a surface with a branch point. We therefore begin with the assumption that there is a one-parameter family $u(t)$, sometimes written ${}^t u$, of minimal surfaces bounded by such that $u(0)$ has a branch point and for $t > 0$, $u(t)$ has least eigenvalue 2 and is immersed (both of which follow if $u(t)$ is a relative minimum for $t > 0$). The proof ends when this assumption is shown to lead to a contradiction. The case of an interior branch point was taken care of in [2], so here we assume that u has a boundary branch point when $t = 0$. The parameter domain is the upper half plane and the branch point is at the origin. We assume Γ is tangent to the X -axis at the origin and the unit normal N (which extends to the branch point) is $(0, 0, -1)$ when t and z are both 0.

(1) The Weierstrass representation has the form

$$u(t, z) = \operatorname{Re} \begin{bmatrix} \frac{1}{2} \int A^2 S + B^2 S dz \\ \frac{i}{2} \int A^2 S - B^2 S dz \\ \int ABS dz \end{bmatrix}$$

where the zeroes a_i , b_i , and s_i of A , B , and S go to zero as t_i^γ , specifically as $\alpha_i t_i^\gamma$, $\beta_i t_i^\gamma$, and $\zeta_i t_i^\gamma$. The α_i , β_i , and ζ_i are nonzero. The order of the branch point is $2m$, its index is k . The s_i are branch points in the lower-half plane converging to 0 as t goes to zero. (Halfway through the proof, we show they don't really exist.)

(2) Fix one of the γ_i and call it γ . Define w by $z = t^\gamma w$. Then in the w -plane, $z - t_j^\gamma$ becomes $t^\gamma(w - \alpha_j)$ if $\gamma_j = \gamma$, or it becomes $t^\gamma w$ if $\gamma_j > \gamma$ (a “fast root”), or it becomes $t_j^\gamma \alpha_j$, if $\gamma_j < \gamma$ (a “slow root”). Slow roots give powers of t , fast roots give powers of w .

(3) In the w -plane, A , B , and S become \mathbb{A} , \mathbb{B} , and \mathbb{S} times a power t^n (plus terms with higher powers of t). \mathbb{A} , \mathbb{B} , and \mathbb{S} are polynomials in w . The degree of \mathbb{A} is the number of α_i that go to zero at the rate t^γ . There is a different w -plane for each distinct choice of γ among the γ_i . We refer to the α_i , β_i , and γ_i that go to zero as t^γ as a “ring of roots.”

(4) B/A , which is g in the Weierstrass representation, is the stereographic projection of the unit normal N . When we pass to the w -plane, there is a number \mathcal{K} such that

$$\frac{B}{A} = t^\mathcal{K} \frac{\mathbb{B}}{\mathbb{A}} + O(t^{\mathcal{K}+1})$$

\mathcal{K} depends on the ring, i.e. there may be a different \mathcal{K} for each ring. If (and only if) $\mathcal{K} > 0$ then we have the “ N -condition”, which means that as t goes to zero, N converges to the south pole $(0, 0, -1)$, uniformly on compact subsets of the w -plane away from the α_i . On the slowest (outermost) ring, we have $\mathcal{K} = k\gamma$, where k is the index of the branch point, so the N -condition holds there.

(5) There is a number \mathcal{M} (for each ring) such that

$$\int A^2 S dz = t^{\mathcal{M}} \int \mathbb{A}^2 \mathbb{S} dw + O(t^{\mathcal{M}+1}).$$

On the outermost ring, $\mathcal{M} = (2m + 1)\gamma$, so $\mathcal{K}/\mathcal{M} = k/(2m + 1)$.

(6) The eigenfunction for the least eigenvalue is given by $\phi = u_t \cdot N$. The eigenfunction has only one sign. Our plan is to calculate the eigenfunction and derive more and more information, and eventually a contradiction, from the fact that the eigenfunction has only one sign.

(7) If the N -condition is satisfied, then \mathbb{A} , \mathbb{B} , and \mathbb{S} are real on the real axis (Lemma 9).

(8) If the N -condition is satisfied, then the eigenfunction is given (Lemma 10) by

$$\frac{-1}{t^{\mathcal{M}+\mathcal{K}-1}} \phi = \text{Im } \mathbb{H} + O(t)$$

where

$$\mathbb{H} = \frac{\mathbb{B}}{\mathbb{A}} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw \right) - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw$$

Here \mathbb{A} and \mathbb{B} are polynomials in w , so \mathbb{H} is a rational function. The principal difficulty in the proof is that \mathbb{H} might be constant.

(9) If \mathbb{H} is constant and the N -condition holds, then by setting the derivative \mathbb{H}_w equal to zero we find (Lemma 17)

$$\frac{\mathbb{B}}{\mathbb{A}} = C \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\mathcal{K}/\mathcal{M}}$$

Then \mathbb{B}/\mathbb{A} is a polynomial, vanishing at the origin, so there are some fast roots, hence this is not the innermost ring.

(10) If \mathbb{H} is not constant and the N -condition holds, then there are no fast roots, since if there were, then $\text{Im } \mathbb{H}$ would take two signs in the upper half plane near the origin. Hence such a ring must be the innermost ring.

(11) The N -condition propagates from a ring where \mathbb{H} is constant and the N -condition holds, to the next ring down (Lemma 20). The reason for this is that, since \mathbb{B}/\mathbb{A} is a polynomial, there are at least as many β_i as α_i , on a ring where \mathbb{H} is constant. Passing to a lower ring, we pick up powers of t from each β_i and α_i , but at least as many enter the numerator of B/A as enter the denominator, so the power of t in B/A is still positive on the next ring down. Hence the N -condition holds on all rings, and \mathbb{H} is constant on all but the last ring, and \mathbb{H} is not constant on the last ring.

(12) The value of \mathcal{K}/\mathcal{M} is the same on all the rings, namely $k/(2m + 1)$ (Lemma 21.) The reason for this is that, by the equation for \mathbb{H} , \mathcal{K}/\mathcal{M} is the

ratio of the degrees of \mathbb{B}/\mathbb{A} and $\int \mathbb{A}^2 \mathbb{S} dw$, and looking at the behavior near the origin on one plane gives us the degrees of those polynomials on the next ring down.

(13) Each a_i contributes one hemisphere of Gaussian area (not a whole sphere). Since the normal is confined within $O(t)$ of the YZ -plane, it suffices (Lemma 13) to show that near α_i in the w -plane, when t is positive, the normal cannot take on both the “west pole” $(-1, 0, 0)$ and the “east pole” $(1, 0, 0)$ (Lemma 11.) The reason for this is that the eigenfunction is given by $u_t \cdot N$, and the first component of u_t near α_i is approximately (a power of t times) $\mathcal{M} \int_0^{\alpha_i} \mathbb{A}^2 \mathbb{S} dw$, which is positive or negative according as α_i is positive or negative; hence if both the east and west poles were taken on, ϕ would take two signs.

(14) The total extra Gaussian area is therefore 2π times the total number of a_i . But by the Gauss-Bonnet theorem, it must be $2\pi m$. Therefore the total number of a_i is m . But the degree of $\mathbb{A}^2 \mathbb{S}$ is $2m$; hence $S = 1$ and the branch points s_i do not exist.

(15) The rest of the proof focusses only on the innermost ring, where \mathbb{H} is not constant. Recall $\sigma = \int_0^w \mathbb{A}^2 dw$. We have (Lemma 25)

$$\frac{\mathbb{B}}{\mathbb{A}} = c \sigma^{\mathcal{K}/\mathcal{M}} \exp \left(\int \frac{\mathbb{H}_w}{\mathcal{M} \sigma \mathbb{B}/\mathbb{A}} dw \right)$$

The integral is a complex path integral. This generalizes the equation given above for the case of constant \mathbb{H} , since in that case $\mathbb{H}_w = 0$ and the exponential factor is 1.

(16) \mathcal{K}/\mathcal{M} (which is equal to $k/(2m + 1)$) is an integer (Lemma 26). The reason why this result is important is that it makes $\sigma^{\mathcal{K}/\mathcal{M}}$ a polynomial, so \mathbb{B}/\mathbb{A} is a polynomial times the exponential factor.

The idea of the proof is that, as w passes through the origin on the real w -axis, \mathbb{B}/\mathbb{A} remains real, and does not change signs (since there are no fast roots). Hence the right side must also remain real. Calculation shows $\mathbb{H}_w \neq 0$ at the origin, so the integrand has a simple pole. We calculate the residue and find it to be $-\mathcal{K}/\mathcal{M}$. This must be an integer in order that the right hand side remain real on both sides of the origin.

(17) The non-real zeroes of σ are also zeroes of \mathbb{B} , and with the same multiplicity as zeroes of \mathbb{B} and zeroes of $\sigma^{\mathcal{K}/\mathcal{M}}$ (Lemma 28).

(18) Let R be the degree of \mathbb{B} and Q the degree of \mathbb{A} . We consider the Laurent expansion

$$\sigma^{-k/\mathcal{M}} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) = 1 + \mu w^{-J} + O(w^{-J-1}).$$

On the left side we have a rational function, since \mathcal{K}/\mathcal{M} is an integer. If we know the degrees of the numerator and denominator of a rational function, we can bound the exponent of the first term in its Laurent expansion (Lemma 29.) On

the face of it the numerator is \mathbb{B} and the denominator is $\mathbb{A}\sigma^{\mathcal{K}/\mathcal{M}}$. But by (17), there is cancellation: all the non-real zeroes of \mathbb{B} cancel out of both numerator and denominator. After that cancellation the general bound in Lemma 29 gives us $J \leq R - 2Q$ (Lemma 30).

(19) Now we come to the final steps, given in the proof of Theorem 8. First we examine the integrand in

$$\int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} dw$$

Using the fact that H_w is asymptotic to a constant or to w^{-2} , we find that the integrand is asymptotic to $w^{-2+Q+R\pm 1}$. Recall that Q and R are the degrees of \mathbb{A} and \mathbb{B} , so $Q + R \geq 1$, which means the exponent is negative with magnitude at least 2. Hence the integral has a finite limit, say L , as w goes to infinity.

(20) So we have

$$\frac{\mathbb{B}}{\mathbb{A}\sigma^{\mathcal{K}/\mathcal{M}}} = c \exp\left(\int \frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}}\right).$$

On the left side, the first term w^{-J} in the Laurent expansion has $J \leq R - 2Q$. But on the right, because H_w is asymptotically constant or w^{-2} , the integrand is asymptotically $w^{-(2+Q+R\pm 1)}$. If we differentiate, we get on the left a term in $w^{-(J+1)}$ and on the right a term in $w^{-(2+Q+R\pm 1)}$. There is a tension here: on the left, the first term in the Laurent series is not too far out, because \mathcal{K}/\mathcal{M} is an integer. On the right, it has to be at least so far out, because \mathbb{H}_w is constant or w^{-2} . Calculation shows that these results cannot both be true unless $Q = 0$, i.e. there are no α_i ; but that is also impossible (Lemma 24). That is the final contradiction.

3 Preliminaries

3.1 Analyticity at the boundary

One of the reasons we need to work with a real-analytic boundary is that we need to know that the minimal surface can be analytically extended across the boundary, so that it is defined in some neighborhood of each boundary point and given by a power series there. This well-known result is due to Lewy, and a proof can be found in [17], p. 107, Theorem 3. It will be taken for granted in the rest of this paper and not cited explicitly.

3.2 Boundary parametrization

By a real-analytic arc, or real-analytic Jordan curve, we mean an arc or curve that can be parametrized as a real-analytic function of arc length. We will consider surfaces bounded by a real-analytic Jordan curve Γ . We suppose that Γ passes through the origin tangent to the X -axis, and that u takes the portion

of the real axis near origin onto Γ , with $u(0) = 0$. We still are free to orient the Y and Z axes. We do this in such a way that the normal at the branch point (which is well-defined) points in the positive Z -direction. With Γ oriented in this way, there will be two positive integers p and q such that Γ has a parametrization in the form

$$\Gamma(\tau) = \begin{bmatrix} \tau + O(\tau^2) \\ \frac{C_1 \tau^{q+1}}{q+1} + O(\tau^{q+2}) \\ \frac{C_2 \tau^{p+1}}{p+1} + O(\tau^{p+2}) \end{bmatrix}$$

for some nonzero real constants C_1 and C_2 . But it is possible to choose τ more carefully so that we do not have the $O(\tau^2)$ term in the first coordinate. Let $\tau(z) = X(z) = \operatorname{Re} \frac{1}{2} \int f - fg^2 dz$. Then on the boundary we have

$$u(z) = \Gamma(\tau(z))$$

This parametrization and function $\tau(z)$ were inspired by Lewy's equation (see [17], p. 108).

We therefore have

$$\Gamma'(\tau) = \begin{bmatrix} 1 \\ C_1 \tau^q + O(\tau^{q+1}) \\ C_2 \tau^p + O(\tau^{p+1}) \end{bmatrix} \quad (1)$$

3.3 Order and index of a branch point

We write $u(z) = (X(z), Y(z), Z(z))$. We make use of the Enneper-Weierstrass representation of u (see e.g. [16], p. 112)

$$u(z) = \operatorname{Re} \begin{bmatrix} \frac{1}{2} \int f - fg^2 dz \\ \frac{i}{2} \int f + fg^2 dz \\ \int fg dz \end{bmatrix}$$

where f is analytic and g is meromorphic in the upper half-disk.

Definition 1 *The order of the branch point is the order of the zero of f . The index of the branch point is the order of the zero of g .*

Remark. We follow [29], p. 315 in this definition of "index". Tromba [17] defines "index" to mean the sum of the orders of f and g , i.e. Tromba's index is Nitsche's index plus the order.

The order of a boundary branch point of a solution of Plateau's problem must be even (since the boundary is taken on monotonically). It is customary to write it as $2m$, and to use the letter k for the index. Thus $f(z) = z^{2m} + O(z^{2m+1})$ and $g(z) = cz^k + O(z^{k+1})$ for some constant c .

3.4 Gauss-Bonnet theorem for branched minimal surfaces

The “geodesic curvature” κ_g of a boundary curve Γ bounding a surface u is the component of the curvature vector of Γ in the plane of u . It is therefore bounded by the magnitude of the curvature vector of Γ , which is defined independently of any surface.

The Gauss-Bonnet formula says that for regular surfaces (minimal or not)

$$\int_{\Gamma} \kappa_g = 2\pi - \int KW \, dx \, dy$$

Note that for minimal surfaces, KW is negative, so both terms on the right are positive. For minimal surfaces with branch points, there is another term in the Gauss-Bonnet formula:

$$\int_{\Gamma} \kappa_g = 2\pi - \int KW \, dx \, dy + 2\pi M$$

where M is the sum of the orders of the interior branch points and half the orders of the boundary branch points. See [17], p. 195, or [29], p. 331, equation (155).

The “total curvature” of a minimal surface is $-\int KW \, dx \, dy$. This is the “Gaussian area”, or the area of the image on the Riemann sphere of the unit normal N to the surface. The expression $|KW|$ is the Jacobian of N . One way to remember the Gauss-Bonnet formula is that each interior branch of order $2m$ point contributes as much m spheres of Gaussian area, and each boundary branch point of order $2m$ contributes as much as m hemispheres of Gaussian area. This way of looking at the formula will be helpful when we consider a family of minimal surfaces without branch points for positive values of the parameter t , but having a branch point when t is zero. So for positive values of t , there must be compensating “bubbles” of Gaussian area.

Somewhat confusingly, the phrase “total curvature” is applied both to surfaces and boundary curves. Applied to a boundary curve, it means the integral of the magnitude of the curvature vector over the whole curve. The total curvature of Γ is thus greater than the left side of the Gauss-Bonnet theorem above, stated using the geodesic curvature.

3.5 On the zero set of real-analytic functions

A subset U of R^n is called an *analytic set* if locally it is the set of simultaneous zeroes of a finite number of real-analytic functions from R^n to some R^m . There is a classical theorem about the structure of analytic sets:

Lemma 1 *Analytic sets are analytically triangulable. That means that each analytic set is the union of a finite number of real-analytic homeomorphic images of closed simplexes, meeting only at their boundaries.*

Six references for this theorem are given in [2], where the theorem is discussed on page 117. The theorem refers to *closed* simplexes, which means that the

homeomorphisms in the conclusion are analytic even at the endpoints of intervals or boundary points of 2-simplexes.

Several times in this paper we have reason to consider the zero set of a function that is analytic in two variables (one real and one complex). When we refer to a function of a real variable t and a complex variable z as “analytic”, we mean it is given by a power series in z and t , so in particular it is real-analytic in t for fixed z and complex-analytic in z for fixed t . The following lemma describes the zero sets of such functions.

Corollary 1 *Let A be an open subset of $\mathbf{R} \times \mathbf{C}$ containing the point $(0, p)$. Let f be a complex-valued analytic function defined on A . Suppose $f(0, p) = 0$ and that for each sufficiently small $t > 0$, $f(t, \cdot)$ is not constant. Then there exists a neighborhood V of $(0, p)$ and finitely many functions c_i such that $f(t, c_i(t)) = 0$, every zero of f in V for $t > 0$ has the form $(t, c_i(t))$, and the $c_i(t)$ are analytic in some rational power of t , with $c_i(0) = p$.*

Proof. If f does not depend on t at all, the theorem is trivial. Otherwise, if f is divisible by some power of t , we can divide that power out without changing the zero set of f for $t > 0$, so we can assume that $f(t, \cdot)$ is not constant for any sufficiently small t , including $t = 0$. Let S be the zero set of f . Then the dimension of S is one, and by Lemma 1, S is composed of finitely many analytic 1-simplexes, i.e. analytic paths given by $t = d_i(\tau)$, $w = c_i(\tau)$, where τ is the parameter in which c_i is analytic. If any of these simplexes do not pass through the $(0, p)$, decrease V to exclude them. The ones that do pass through the $(0, p)$ should be divided into two simplexes, each ending at $(0, p)$. Then we can assume that $\tau = 0$ corresponds to $(0, p)$, i.e. $d_i(0) = 0$ and $c_i(0) = p$. We have

$$f(d_i(\tau), c_i(\tau)) = 0.$$

Since d_i is analytic and not constant, it has only finitely many critical points. Decrease the size of V if necessary to exclude all nonzero critical points of d_i . Then either d_i is increasing in some interval $[0, b]$ or decreasing in some interval $[0, b]$. In either case it has an inverse function $\phi = d_i^{-1}$, so $d_i(\phi(t)) = t$. If the leading term of d_i is τ^n , then ϕ is real-analytic in $t^{1/n}$. We can parametrize the zero set of f by $\tilde{c}_i(t) = c_i(\phi(t))$, which is real-analytic in $t^{1/n}$. That completes the proof.

Remark: In applications it will usually be possible to replace the original parameter t by $t^{1/n}$, enabling us to assume that the zero set is analytic in t .

3.6 Regularity results for relative minima of area

We need the following theorem:

Theorem 1 *Let u be a minimal surface bounded by a real-analytic Jordan curve Γ and furnishing a relative minimum of area in the C^0 metric on the closed disk. Then u has neither interior nor boundary branch points.*

It is an open problem to replace C^0 by C^n in this theorem. We now summarize the relevant results from the literature.

In [17], pp. 554-560, there is an extensive *Scholia* reviewing the various results and proofs of regularity for interior and boundary branch points as matters stood in 2010. This *Scholia* is reprinted and updated to 2012 in [37], pp. 169–175. We are here concerned only with the case of real-analytic boundary, but we need the regularity of *relative* minimizers (of area or Dirichlet energy), not just absolute minimizers, and both interior and boundary regularity.

In the theory of branch points, one distinguishes between true branch points and false branch points (see p. 58 of [17] for the definition). False branch points cannot exist for any solution of Plateau’s problem, minimizing or not; see the discussion in the *Scholia* (*op. cit.*) for references. In the rest of this discussion, we consider true branch points.

First we take up interior regularity. It seems that [1] was the first to claim interior regularity for relative minima, as opposed to absolute minima, but Osserman’s proof also yields such a result. Osserman’s proof works for C^0 relative minima, but not for C^1 relative minima, because the proof involves “smoothing off” a wedge. The argument in [1] is (also) a local argument, but does not involve cut-and-paste: it shows how to decrease the area (in a C^n smooth way) in a neighborhood of a branch point. To obtain a global result one must supplement that argument by a smoothing argument to remove the “crease” introduced by decreasing area locally. Wienholtz has correctly pointed out that smoothing argument given in [1] works only for C^1 , although C^n is claimed. Nevertheless the result is correct for C^1 , as confirmed in the *Scholia* just mentioned. So as far as interior branch points go, we could replace C^0 by C^1 in the theorem.

In [37], Tromba made direct calculations attempting to decrease the Dirichlet integral (globally) for a minimal surface with an interior branch point. While his method was successful in most cases, in certain cases he was still forced to make a local argument, so that C^1 can still not be replaced with C^n in the theorem; but Tromba did provide an independent proof of the interior regularity for C^1 relative minima.

Now we consider boundary regularity. In the case of real-analytic boundary curves, area and energy can be locally decreased near a true boundary branch point, as proved independently in [21] and [38]. Both proofs rely on an Osserman-style cut-and-paste. Hence they work for C^0 relative minima, although the papers claim the theorem only for absolute minima, and the *Scholia* cited above also does not state that the theorem works for C^0 relative minima.

However, we do not have to rely on my claim that these two proofs work for C^0 relative minima. Instead we can rely on the work of Tromba [37], who also made calculations for a minimal surface with a boundary branch point. While serious difficulties remain in the case of boundaries that are only C^n , Tromba’s method does work in the case of a real-analytic boundary curve. But, as mentioned above, Tromba also needed a local argument in some cases, so his proof also works only for C^0 relative minima. For Tromba’s explicit statement that his method works in the real-analytic boundary case, see p. 168 of [37]. It is the last sentence in the book, not counting the *Scholia*.

3.7 Dirichlet integral and area

Let D be a plane domain; for our purposes, we may suppose D is a disk or a half-plane, so we will not worry about the conditions on the boundary ∂D of D . The Dirichlet integral, or “energy”, of a surface u defined in D is defined by

$$E(u) = \frac{1}{2} \int_D u_x^2 + u_y^2 dx dy$$

The area functional is defined by

$$A(u) = \int_D |u_x \times u_y| dx dy$$

These functionals are defined on various spaces of functions on D , for example $C^{m,\alpha}$ or the Sobolev spaces $W^{k,p}$. Since we will be restricting attention to surfaces bounded by a fixed real-analytic Jordan curve Γ , we are interested in a space of functions defining surfaces bounded by Γ . These function spaces can be considered as Hilbert manifolds, and the tangent space at a surface u is a vector space of functions defined on D whose values are tangent to Γ on the boundary of D . These functionals have Frechet derivatives, which we denote by $DE(u)$ and $DA(u)$. These are linear mappings on the tangent space.

Minimal surfaces are defined as critical points of E , but it is well known that they are also critical points of A . In connection with A , it is common to consider only “normal variations”, i.e. to restrict $DA(u)$ to the subspace of the tangent space consisting of those k such that $k(x, y)$ is normal to u . It can be shown that if $DA[u]$ vanishes on this subspace then u is minimal. (Details can be found in [29], p. 94). In connection with E , it is common to consider only harmonic surfaces; or equivalently, spaces of functions defined on the boundary of the parameter domain and mapping it to Γ in a way homotopic to the identity (since such functions have a unique harmonic extension to the interior).

3.8 The second variation of Dirichlet’s integral and the forced Jacobi fields

At a minimal surface u , we can consider the second variation $D^2E(u)$. This is a bilinear mapping on the tangent space at u . The members of this tangent space are “tangent vectors” k , i.e. functions defined on ∂D such that $k(\xi)$ is tangent to $u(\xi)$ for ξ on ∂D . The kernel of $D^2E(u)$ consists of tangent vectors k such that $D^2E(u)[k, j] = 0$ for all tangent vectors k . We write $D^2E(u)[k]$ for $D^2E(u)[k, k]$. By diagonalizing the bilinear form, one can prove that the kernel of $D^2E(u)$ consists of those k for which $D^2E(u)[k] = 0$.

Theorem 2 (Tromba) *The second variation of Dirichlet’s integral is given by*

$$D^2E[u](h, k) = \int k(h_r - \tilde{h}_\theta) d\theta$$

where $k = \lambda u_\theta$ and $h = \eta u_\theta$ and $\tilde{h} = \eta u_r$ and $\tilde{k} = \lambda u_r$. The tangent vector k to the minimal surface u belongs to $\text{Ker } D^2E[u]$ if and only if

$$u_\theta(k_r - \tilde{k}_\theta) = 0$$

or equivalently

$$k_z u_z = 0$$

Proof. The first formula is equivalent to the one given in [36] (bottom of p. 53, with h and k interchanged); for a stand-alone one-page proof of it by direct calculation, and the easy derivation of the second two formulas from the first, see [5].

Consider the kernel equation $u_\theta(k_r - \tilde{k}_\theta) = 0$. One way in which this could be satisfied is if $k_r - \tilde{k}_\theta = 0$; vectors k satisfying this condition and not induced by the conformal group are called “forced Jacobi fields” or “forced Jacobi directions”. Tromba proved that they do not occur in the absence of branch points, and that in the presence of branch points there are two for each interior branch point (counting multiplicities) and one for each boundary branch point, so that the space of forced Jacobi fields is finite dimensional. (There can be at most finitely many branch points, even if the boundary is not real-analytic, as long as the total curvature of the surface is finite, thanks to the Gauss-Bonnet formula for branched minimal surfaces.) The forced Jacobi directions are just the directions k such that the function $K = k + i\tilde{k}$ is complex analytic, i.e. such that \tilde{k} is the conjugate harmonic function of k .

Another important characterization of the forced Jacobi fields is this: they are exactly the tangent vectors of the form

$$k = \text{Re}(i\omega z u_z)$$

where $i\omega z$ is a function meromorphic in the parameter domain, and having a pole of order at most m at each branch point of order m . Any function ω with suitable behavior on the boundary, and poles of the right orders at the branch points, will produce a tangent vector by this equation. The reason for writing the equation with ωz instead of with ω is that in case the parameter domain is the unit disk, the appropriate boundary condition is that ω be real on S^1 . In case the parameter domain is the upper half plane, the condition is that $i\omega z$ be real on the x -axis. The Appendix of [13] contains Tromba’s treatment of the forced Jacobi fields.

Lemma 2 (Tromba [36]) *Suppose u is a minimal surface, and k is a tangent vector belong to $\text{Ker } D^2E[u]$ whose harmonic extension is everywhere tangent to u . Then k is a forced Jacobi direction or a direction induced by the conformal group.*

3.9 The second variation of area

Lemma 3 (Laplacian of the Gauss map) *Let u be a minimal surface. Then the Laplacian of its unit normal is given by $\Delta N = 2KWN$.*

Proof. To prove this elegantly, we make use of the general fact that the Laplace-Beltrami operator of any surface S , applied to the position vector of S , is exactly twice the mean curvature of S . (For a proof of that fact, see p. 45 of [16], formula (30), with f the position vector of X so that H_f in formula (27) is the first fundamental form of X .) Apply this fact to the Riemann sphere, whose position vector $h(w)$ coincides with its unit normal. Thus $\Delta h = -2h$. Next, note that the map from the disk D to the Riemann sphere induced by the Gauss map of u is a conformal map with Jacobian $-KW$. Under a conformal map, the Laplace-Beltrami operator changes to the Laplace-Beltrami operator on the range surface, multiplied by the Jacobian of the mapping. Hence $\Delta N = 2KW N$, and the lemma is proved.

One can define the second variation of area, $D^2 A(u)$, on the original tangent space, or on the subspace of the tangent space corresponding to normal variations. Nitsche tells us ([29], p. 95) that we get the same result, but that the calculation is too long to put in his 562 page book. It is customary to consider the second variation of area on normal variations only. The second variation $D^2 A(u)\phi$ is a bilinear functional on normal variations; technically normal variations are “tangent vectors” to a Hilbert manifold of surfaces, but of course they are not tangent to the surfaces themselves. Abusing notation slightly, we write $D^2 A(u)[\phi, \psi]$ to stand for $D^2 A(u)[k, j]$ where k is the tangent vector $\phi \cdot N$ and $j = \psi \cdot N$ (and N is the unit normal to u). Thus $D^2 A(u)[\phi, \psi]$ is, in classical terms,

$$\frac{\partial^2}{\partial s \partial t} A(u + s\phi + t\psi)$$

evaluated at $t = 0$ and $s = 0$. This derivative is only used when ϕ and ψ are in the kernel of the first variation DA (which of course holds when u is minimal). In case $\phi = \psi$ we recover the classical second variation of area:

$$D^2 A(u)[\phi, \phi] = \frac{\partial^2}{\partial t^2} A(u + t\phi),$$

which by a further abuse of notation is usually written

$$D^2 A(u)[\phi].$$

One can calculate that

$$D^2 A(u)[\phi, \psi] = \int_D \psi(-\Delta\phi + 2KW\phi) dx dy, \quad (2)$$

where K is the Gauss curvature and $W = |u_x \times u_y|$. See for example [29], p. 96, where the case $\psi = \phi$ is calculated, but the calculation easily adapts to the bilinear second variation.

Lemma 4 *ϕ belongs to the kernel of $D^2 A(u)$ (for normal variations) just in case ϕ is zero on the boundary of the parameter domain and satisfies*

$$\Delta\phi = 2KW\phi$$

in the interior.

Proof. Apply the fundamental lemma of the calculus of variations to the preceding formula.

3.10 Connections between D^2A and D^2E

Theorem 3 *Let u be a minimal surface in R^3 with C^n boundary, and unit normal N . Let k be in $KerD^2E[u]$. Then $\phi = k \cdot N$ belongs to $KerD^2A[u]$.*

Corollary 2 *If $KerD^2A[u]$ has no kernel among normal variations, $KerD^2E[u]$ contains only the conformal and forced Jacobi directions.*

Proof. The Corollary follows immediately from the theorem and Tromba's lemma (Lemma 2). We now prove the theorem. Suppose k is in $KerD^2E[u]$. By Lemma 4, it will suffice to show that $\phi = k \cdot N$ satisfies $\Delta\phi - 2KW\phi = 0$. We have

$$\Delta\phi = \Delta(k \cdot N) = (\Delta k) \cdot N + 2\nabla k \nabla N + k \Delta N.$$

The first term vanishes because k is harmonic. We claim the second term vanishes also. To prove this, fix a point z_0 in the unit disk, and choose coordinates a and b in a neighborhood of z_0 that diagonalize the first fundamental form at z_0 , so that $N_a = \kappa_1 u_a$ and $N_b = \kappa_2 u_b$, where κ_1 and κ_2 are the principal curvatures of u at z_0 . If these equations hold in a whole neighborhood, then a and b are called "local curvature coordinates"; it costs some trouble to prove they exist, and we do not need them; we need the first fundamental form to be diagonalized at one point z_0 only. To do this, we take a and b to be a certain linear combination of x and y . Let ν be the angle between the positive x -direction and the positive a -direction (so ν is a function of z_0 but not of z). Then we define $w = e^{i\nu}(z - z_0)$, and define a and b by $w = a + ib$, so a and b are coordinates rotated by ν from (x, y) .

Because u is a minimal surface, we have $\kappa_1 = -\kappa_2$. Then at z_0 we have

$$\begin{aligned} \nabla k \cdot \nabla N &= k_a N_a + k_b N_b \\ &= k_a \kappa_1 u_a + k_b \kappa_2 u_b \\ &= \kappa_1 (k_a u_a - k_b u_b) \end{aligned}$$

We have $u_w = u_a - i u_b$ and $k_w = k_a - i k_b$, so

$$\begin{aligned} \nabla k \cdot \nabla N &= \kappa_1 \operatorname{Re}(k_w \cdot u_w) \\ &= \kappa_1 \operatorname{Re}((k_z z_w) \cdot (u_z z_w)) \\ &= \kappa_1 \operatorname{Re}(z_w (k_z \cdot u_z)) \\ &= \kappa_1 \operatorname{Re}(e^{-i\nu} (k_z \cdot u_z)) \end{aligned}$$

since $z_w = e^{-i\nu}$. Since k is assumed to be in $KerD^2E[u]$, we have $k_z u_z = 0$. Hence the term $\nabla k \nabla N$ vanishes at z_0 . But z_0 was arbitrary; hence $\nabla k \nabla N$ vanishes everywhere, and we have proved $\Delta\phi = k \cdot N$.

The proof of the theorem is thus reduced to proving $\Delta N = 2KW N$. But this is Lemma 3. That completes the proof.

The converse of this theorem is also true (but more difficult): every ϕ in the kernel of $D^2A(u)$ arises as $k \cdot N$ for some k in the kernel of $D^2E(u)$. We do not need this result in this paper, but the curious can find a proof in [5] or [3].

3.11 Stereographic projection

We take the “Riemann sphere” to be the unit sphere $\{(x, y, z) : x^2 + y^2 + z^2 = 1\}$. Stereographic projection \mathbf{St} is defined by

$$\mathbf{St}((x_1, x_2, x_3)) = \frac{x_1 + ix_2}{1 - x_3}$$

and its inverse is given by

$$\mathbf{St}^{-1}(z) = \frac{1}{1 + |z|^2} \begin{bmatrix} 2 \operatorname{Re} z \\ 2 \operatorname{Im} z \\ |z|^2 - 1 \end{bmatrix}.$$

Thus the equator projects onto the unit circle, and the “north pole” $(0, 0, 1)$ projects onto ∞ , while the southern hemisphere projects onto the unit disk, with the south pole $(0, 0, -1)$ going to origin. The picture then has the plane passing through the equator of the sphere. We follow [30], p. 46, in these details, and we mention them because some other authors use a sphere of radius $1/2$, with the picture having the sphere entirely above the plane, tangent to the plane where the south pole of the sphere touches the origin of the plane.

3.12 The eigenvalue problem associated with the second variation of area

Associated with a conformal map N defined on a region Ω in the plane and taking values in the Riemann sphere, there is a natural eigenvalue problem:

$$\begin{aligned} \Delta\phi - \frac{1}{2}\lambda|\nabla N|^2\phi &= 0 & \text{in } \Omega \\ \phi &= 0 & \text{on } \partial\Omega \end{aligned}$$

The Jacobian of N is $\frac{1}{2}|\nabla N|^2$. In case N is the Gauss map of a minimal surface, the Jacobian is also $-KW$, so $\frac{1}{2}|\nabla N|^2 = -KW$ and the eigenvalue equation becomes

$$\Delta\phi + \lambda KW\phi = 0$$

As proved above in Lemma 4, if ϕ is in the kernel of $D^2A(u)$, then ϕ is an eigenfunction of this equation for $\lambda = 2$. But for purposes of this section, N can be any map from the disk to the Riemann sphere.⁴

⁴Some readers may be familiar with another form of the eigenvalue equation for which the critical eigenvalue is zero rather than 2, or with this form but with a factor of 2 inserted so that the critical eigenvalue is 1 instead of 2; both forms are discussed in [29], p. 103, cf. equations (62) and (62').

We denote the least eigenvalue λ of this problem by $\lambda_{N,\Omega}$, or simply by λ_Ω when N is clear from the context. Sometimes we use the notation λ_{\min} . Sometimes, for a minimal surface u , we speak of the “least eigenvalue of u ” rather than the “least eigenvalue of the second variation of u ” or “the least eigenvalue of the eigenvalue problem associated with the second variation of u .”

The least eigenvalue is well-known to be equal to the infimum of the Rayleigh quotient

$$R[\phi] = \frac{\int \int_\Omega |\nabla \phi|^2 dx dy}{\int \int_\Omega \frac{1}{2} |\nabla N|^2 \phi^2 dx dy}.$$

When we speak of the least eigenvalue λ_Ω of a region Ω on the Riemann sphere, we mean the following: Let Δ be the stereographic projection of Ω and N the inverse of stereographic projection. Then $\lambda_\Omega := \lambda_{N,\Delta}$. The eigenvalue problem $\Delta \phi - \frac{1}{2} \lambda |\nabla N|^2 \phi$ on Δ is equivalent to the problem $\Delta \phi = \lambda \phi$ on Ω , where now Δ is the Laplace-Beltrami operator on the sphere. If Ω contains the north pole, we should use stereographic projection from some point not contained in Ω . We do not need to discuss the case when Ω is the entire sphere.

Example. We compute the least eigenvalue when $N(\Omega)$ is a hemisphere. In this case the eigenfunction in the lower hemisphere is minus the Z -component of N . For example with Ω equal to the unit disk and $g(z) = z$, we have $N(z)$ the inverse of stereographic projection. With $|z| = r$ and $z = x + iy$ we have

$$N(z) = \frac{1}{1+r^2} \begin{bmatrix} 2x \\ 2y \\ r^2 - 1 \end{bmatrix}.$$

The eigenfunction ϕ is given, with $|z| = r$, by

$$\phi(z) = \frac{1-r^2}{1+r^2}.$$

A few lines of elementary computations (or a couple of commands to a computer algebra program) show that

$$\begin{aligned} \Delta \phi &= \phi_{rr} + \frac{1}{r} \phi_r \\ &= \frac{8(r^2 - 1)}{(1+r^2)^3} \end{aligned}$$

and

$$\begin{aligned} |\nabla N|^2 &= N_x^2 + N_y^2 = \frac{8r}{(1+r^2)^2} \\ |\nabla N|^2 \phi &= \frac{8(1-r^2)}{(1+r^2)^3} \end{aligned}$$

Hence, $\Delta \phi - |\nabla N|^2 \phi = 0$, which means the eigenvalue of a hemisphere is 2.

Lemma 5 *[Passing to the Riemann sphere does not decrease the eigenvalue]* Let Ω be a connected open set on the sphere with least eigenvalue λ_Ω . Suppose Δ is a region (open set) in the plane and $\Omega \subseteq N(\Delta)$ and $N(\partial\Delta) \cap \Omega = \phi$. Suppose that boundary of Ω is C^2 and the boundary of Δ is piecewise C^2 with the pieces meeting at positive angles. Then $\lambda_{N,\Delta} \leq \lambda_\Omega$, and strict inequality holds if $N^{-1}\partial\Omega$ contains an interior point of Δ .

Remark. Regarding the assumptions on the boundaries, the proof requires that the least eigenvalue be the minimum of the Rayleigh quotient, and that the gradient of the least eigenfunction of Ω not vanish at any boundary point. The hypotheses given imply these conditions but still allow Δ to be a half-disk. See [19].

Proof. Let ϕ be the least eigenfunction of Ω and define ψ on Δ by setting $\psi(z) = \phi(N(z))$ if $N(z) \in \Omega$ else $\psi(z) = 0$. Then ψ is admissible in the Rayleigh quotient for Δ , since ϕ is zero outside Ω and $N(\partial\Delta) \cap \Omega = \phi$. The Rayleigh quotient in question is

$$\frac{\int \int_{\Delta} |\nabla\psi|^2 dx dy}{\int \int_{\Delta} \frac{1}{2} |\nabla N|^2 \psi^2}$$

(The factor 1/2 was explained above.) Since $\Omega \subseteq N(\Delta)$, on the support of ψ , N is a covering map (i.e., locally a homeomorphism), except at the points of ramification of N , which are isolated. Since Ω is connected and $N(\partial\Delta) \cap \Omega = \phi$, the number of sheets over (cardinality of the pre-image of) N of each non-ramification point is the same. Hence each of the two integrals in the (numerator and denominator of the) Rayleigh quotient is the number of sheets times the corresponding integral on the Riemann sphere, with ϕ in place of ψ . That is, the Rayleigh quotient for ψ on Δ equals the Rayleigh quotient for ϕ on Ω , which is λ_Ω . Since λ_Δ is the minimum of such Rayleigh quotients, $\lambda_\Delta \leq \lambda_\Omega$. Now suppose there is an interior point p of Δ in $N^{-1}(\partial\Omega)$. By Hopf's lemma, $\nabla\phi$ is never zero at a point on $\partial\Omega$ where the boundary is C^2 . By analyticity, ∇N is zero only at isolated points; so there is a point q near p which is still on $N^{-1}(\partial\Omega)$ at which ∇N is not zero and $\nabla\psi$ is not zero. Hence near q , the set $N^{-1}(\partial\Omega)$ is a smooth arc, and $\nabla\psi$ is zero on one side of it and bounded away from zero on the other side. Hence we can "smooth out the edge" near q to obtain a function ψ' which is admissible for the Rayleigh quotient and has smaller Rayleigh quotient than ψ . Hence $\lambda_\Delta < \lambda_\Omega$. This completes the proof of the lemma.

Corollary 3 *If the Gaussian image of a minimal surface defined in (the open set) Δ contains a hemisphere, and at least one boundary point of the hemisphere, then the eigenvalue $\lambda_{N,\Delta}$ is less than 2.*

Proof. The eigenvalue of a hemisphere is 2.

3.13 Dependence of the Gauss map on a parameter

In this section we present a result which will be applied to the Gauss map of a one-parameter family of minimal surfaces. However, we present it here as a

result about conformal mappings from a plane domain to the Riemann sphere.

Let D^+ be the upper half of the unit disk. We will consider a one-parameter family of analytic mappings from D^+ to the Riemann sphere. (These arise in our work as the Gauss maps of a one-parameter family of minimal surfaces.) We write t for the parameter and $N(t, z)$ for the value of the mapping. For brevity we often omit the explicit t -dependence and write $N(z)$, or $N^t(z)$. Thus N^t is the map $N(t, \cdot)$ from D^+ to the Riemann sphere, for a fixed t . We define $g = \mathbf{St} \circ N$. The map N is thus conformal except at the zeros of g' and poles of g of order more than one. (These are sometimes known as *ramification points* of g .)

We will suppose that for sufficiently small t , the map g is a quotient of functions B/A , where B and A are analytic jointly in t and z (in the sense explained in section 3.5), and that when $t = 0$, A has a zero of order m , and B has a zero of order $m + k$, for some positive integers m and k . (In our work, these arise from a boundary branch point of order $2m$ and index k , but here it is not necessary to be so specific.)

By Corollary 1, near the origin there exist paths $a_i(t)$ and $b_i(t)$ describing the poles and zeroes of g , respectively, so that each a_i and b_i is analytic in a rational power of t . Let s be the greatest common divisor of the denominators of these rational powers, and replace t by $t^{1/s}$. Then a_i and b_i will be analytic in t . Thus we may suppose that the parameter t has been chosen such that a_i and b_i are analytic in t . We also may suppose without loss of generality that for all sufficiently small positive t , we have $a_i(t) \neq b_j(t)$ for all i, j , since otherwise by analyticity we would have, for some i and j , $a_i(t) = b_j(t)$ for all sufficiently small t , and then the factor $z - a_i(t)$ could be cancelled out of both A and B . The partial derivatives of a quotient of real analytic functions are again quotients of real-analytic functions, so the structure of the set of critical points of the real and imaginary parts of such a function is also known: it is the union of a finite set of real-analytic arcs.

3.14 Dependence of eigenvalues on a parameter

Let D be the unit disk and let $0 < \alpha < 1$. Let $F : [0, 1] \times D \rightarrow \mathbf{R}$, and suppose that $F(t, x, y) \geq 0$ for all t, x , and y . For each fixed t let F^t be the function of x, y given by $F^t(x, y) = F(t, x, y)$. Then the eigenvalue problem $\phi = \lambda F^t \phi$ (with $\phi = 0$ on the boundary) has eigenvalues $\lambda_1^t, \lambda_2^t, \dots, \lambda_n^t$, given in non-decreasing order. For n such that the eigenspace of λ_n has dimension greater than 1, that leaves some ambiguity about the indexing of the eigenvalues. Suppose that F depends in some smooth way on t , perhaps even real-analytically. Then what can be said about the dependence of λ_n on t ? At least this much:

Lemma 6 (Continuous dependence of eigenvalues) *Consider the eigenvalue problem $\Delta \phi = \lambda F_t \phi$, with $\phi = 0$ on the boundary. For each fixed value t_0 of t , there are continuous functions $\lambda_n(t)$ defined in a neighborhood of t_0 giving the eigenvalues.*

Proof. The proof can be found on page 419 of [15].

In fact more is true. If F depends real-analytically on t , then it is possible to number the eigenvalues in such a way that they are given by real-analytic functions of t , as shown in [23], p. 370, p. 387; in general this numbering will not be in order of size, since two eigenvalues, for example λ_2 and λ_3 , might “cross” at a certain value of t . Of course no other eigenvalue can cross or even touch λ_1 since its eigenspace is always one-dimensional. For our purposes simple continuity of λ_1 and λ_2 suffices.

3.15 Tomi’s no-immersed-loops theorem

We state and sketch the proof of a theorem due to Tomi [33], or maybe it is due to Tomi and Böhme [12]. It is difficult to give an exact reference for this theorem as the paper where it is stated [33] contains deeper theorems about the structure of the solution set of Plateau’s problem, and the calculation needed for this proof is referenced to [12] where, at the crucial point, the paper says “Durch eine elementare aber etwas mühsame Rechnung finder man. . .” (by an elementary but somewhat tiresome computation one finds), and does not give the computation. Also, they required the boundary to be $C^{4,\alpha}$, but that was for other reasons in their paper; C^2 is enough for the theorem stated here (although in our paper, the boundary is always real-analytic anyway).

Theorem 4 (Tomi) *Let Γ be a C^2 Jordan curve and suppose $u = u(t)$ is a periodic one-parameter family of minimal surfaces, C^2 as a function of z and C^1 in t , bounded by Γ for each t , and satisfying a three-point condition. Suppose that u_t is not identically zero as a function of z for any t , and that each $u(t)$ has $\lambda_{\min} = 2$. Then some $u(t)$ has a branch point, either in the interior or on the boundary.*

Remarks. By a “periodic family”, we mean that $u(t + 2\pi) = u(t)$; the exact period is not relevant. The condition $\lambda_{\min} = 2$ will be fulfilled if $u(t)$ is a relative minimum of area, but it is a more general condition. The condition that the $u(t)$ satisfy a three point condition is only needed to guarantee that u_t is not a conformal direction.

Proof. Suppose, for proof by contradiction, that $u(t)$ has no branch point. Since each $u(t)$ is a minimal surface, the first variation of Dirichlet’s integral E is zero, so $E(t) = E[u(t)]$ is constant. Hence the second derivative $\partial^2 E / \partial t^2 = 0$. Hence the second variation of E is zero in the direction u_t . That is,

$$D^2 E[u](u_t) = 0$$

Define

$$\phi := u_t \cdot N$$

where N is the unit normal to $u(t)$. (We suppress the t -dependence in our notation, writing u instead of $u(t)$ and not indicating the t -dependence of ϕ and u_t .) Because there is no branch point, there are no forced Jacobi directions. Because of the three-point condition, u_t is not a conformal direction. Then

$D^2[A](\phi) = 0$ as shown in the previous section. Since u has no branch points, ϕ is not identically zero. By Theorem 3, ϕ is an eigenfunction of

$$\Delta\phi = 2KW\phi$$

over the parameter domain D , with $\phi = 0$ on the boundary ∂D . We define the “volume integral” to be

$$V(t) := \int_D u \cdot (u_x \times u_y) \, dx dy = \int_D u \cdot NW \, dx dy$$

where N is the unit normal to u . The key to the proof is the “mühsame Rechnung” that

$$\frac{\partial V}{\partial T} = V_t = \int \phi W \, dx dy$$

That computation, not found in either [33] or [12], is written out in detail in [5]. Since by hypothesis $\lambda_{\min} = 2$, for each t the function $\phi = \phi(t)$ has only one sign in the interior of the parameter domain. Since ϕ is continuous in t (because u is C^1 in t), that sign is the same for all t in $[0, 2\pi]$. Now choose t_0 at which $V(t)$ has its minimum value. Then we have

$$\begin{aligned} 0 &= V_t(t_0) \\ &= \int u_t \cdot NW \, dx dy \\ &= \int \phi W \, dx dy \end{aligned}$$

But since ϕ is not identically zero and has one sign, this is a contradiction.

3.16 Böhme and Tomi’s structure theorem

Since the 1930s it has been known that, if Γ bounds infinitely many minimal surfaces, there is a sequence u_n of them converging to a minimal surface u bounded by Γ . We need an “analytic compactness” result that will give us a one-parameter family of minimal surfaces instead of just a sequence. Böhme and Tomi [12, 9] and Tromba [36] applied nonlinear global analysis to the theory of minimal surfaces. We will state and use the structure theorem of Böhme and Tomi. This theorem is Satz 2.4, page 15 in [12]. That theorem is stated more generally, to cover the case of constant mean curvature as well as the minimal-surface case; here we state it only for minimal surfaces.

Theorem 5 (Böhme-Tomi structure theorem) *Let u be a minimal surface bounded by a real-analytic Jordan curve in \mathbf{R}^3 . Let $0 < \alpha < 1$. Let \mathcal{F} be the space of $C^{2+\alpha}$ maps from D to \mathbf{R}^3 satisfying a three-point condition and whose restrictions to the boundary map S^1 onto the range of Γ and are homotopic to Γ . Let $\mathcal{M}'(\Gamma)$ be the set of minimal surfaces in \mathcal{F} . Then there is a neighborhood W of u in the space $C^{2,\alpha}$, such that $W \cap \mathcal{M}'(\Gamma)$ is an analytic subset of a finite-dimensional analytic submanifold of \mathcal{F} .*

Warning. The surfaces in W are *not* required to map the boundary of the parameter domain onto Γ monotonically. That is why we do not simply say that surfaces in \mathcal{F} are “bounded by Γ .” The notation \mathcal{M}' instead of \mathcal{M} means that monotonicity is not required.

Remarks. The proof goes by introducing a map \mathcal{L} from W to $C^{2+\alpha}(D) \times C^{2+\alpha}S^1$, defined by

$$\mathcal{L}(u) = (\Delta u, u_r \cdot u_\theta)$$

Since a harmonic function conformal on the boundary is also conformal in the interior, the zeroes of \mathcal{L} are exactly the minimal surfaces. Böhme and Tomi show that \mathcal{L} is a Fredholm map, which means that the Frechet derivative $D\mathcal{L}(u)$ has finite-dimensional kernel and cokernel. The finite-dimensional manifold in the theorem is parametrized by the kernel of \mathcal{L} . The theorem essentially boils down to an application of the implicit function theorem; if $D\mathcal{L}$ is zero in the n directions $t = (t_1, \dots, t_n)$, and nonzero in other directions, then the non-isolated zeroes of \mathcal{L} are locally analytic functions of t . That remark, of course, is not a proof—it is intended only to convey the idea of the proof in [12].

The kernel of $D\mathcal{L}$ is essentially the kernel of D^2E , the second variation of Dirichlet’s integral. We now know that kernel consists of the forced Jacobi directions (in case u has branch points), plus directions with non-zero normal components.

In [9], Satz 6 and Satz 11, Böhme spells out the consequences of the structure theorem more explicitly. With Γ , u , and \mathcal{F} be as in the previous theorem, let $\mathcal{M}'(\Gamma)$ be the set of minimal surfaces in \mathcal{F} . The notation $\mathcal{M}'(\Gamma)$ is Böhme’s. It contrasts with $\mathcal{M}(\Gamma)$, in that members of $\mathcal{M}(\Gamma)$ must take on Γ monotonically (that is, topologically); both sets impose a three-point condition. Let \mathcal{L} be as defined above, so $\mathcal{L}(u)$ is zero just when u is a minimal surface in \mathcal{F} . Then the kernel of $D\mathcal{L}(u)$ is a finite-dimensional vector space V . Böhme’s Satz 11 uses terminology from his Satz 6; putting in that terminology, Satz 11 becomes:

Theorem 6 (Böhme analytic structure theorem) *There is a neighborhood W of u and an analytic map j defined on the unit ball U in R^n , where n is the dimension of the kernel of D^2E , such that $\mathcal{A} := j^{-1}\mathcal{M}'(\Gamma) \cap W$ is an analytic subset of U .*

Böhme does not define “analytic subset”, and we wish to make sure there is no ambiguity about it. It means, the set of simultaneous zeroes of a finite number of real-analytic functions from U to some R^m . Explicitly, we have

$$\mathcal{M}(\Gamma) \cap W = \{v \in j(U) : \mathcal{L}(v) = 0\}$$

Then $j(t)$ is a function of z , and $j(t)(x, y)$, which we can write as $j(t, x, y)$, is real-analytic in all three variables, as Böhme points out. We have

$$\mathcal{A} = \{t \in V : (\mathcal{L} \circ j)(t) = 0\}$$

This “analytic set” is the set of zeroes of a real-analytic function from a finite-dimensional space to an infinite-dimensional space, but since U is finite-dimensional and $\mathcal{L} \circ j$ is analytic, $\mathcal{L} \circ j(U)$ is also a finite-dimensional manifold.

The structure of the set of minimal surfaces $\mathcal{M}'(\Gamma)$ bounded by Γ in a neighborhood of u is thus reduced to the study of the local structure of analytic subsets of R^n . If we want to also require monotonicity on the boundary, the set $\mathcal{M}(\Gamma)$ becomes *semianalytic*; that is, defined by a Boolean combination of analytic sets. For our purposes, it is enough to consider $\mathcal{M}'(\Gamma)$.

Corollary 4 *Let Γ be a real-analytic Jordan curve in \mathbf{R}^3 . Then the set of minimal surfaces bounded by Γ is locally analytically triangulable. That is, if u is a minimal surface bounded by Γ , and \mathcal{F} is the space of minimal surfaces bounded by Γ and satisfying a three-point condition, then there is a neighborhood W of u in \mathcal{F} such that the set of minimal surfaces in W is a union of analytic images of simplexes in R^n .*

Proof. By Böhme’s structure theorem and Lemma 1.

Böhme’s theorem leaves open the possibility that the analytic simplexes have their vertices at u . If u is a relative minimum of area, more can be said: the space of minimal surfaces is one-dimensional and a 1-simplex passes right through u , not just up to it:

Lemma 7 (Tomi) *Let u be a relative minimum of area bounded by a real-analytic Jordan curve Γ . If u is not isolated, then the set of minimal surfaces bounded by Γ in some neighborhood of u is exactly a one-parameter family u^t , with $u^0 = u$, defined for t in some open interval containing 0.*

Proof. See [34].

4 Basic setup and Weierstrass representation

4.1 A one-parameter family terminating in a branched surface

The starting point of our work is the following theorem. The statement and proof can be sketched as follows: if Γ bounds infinitely many relative minima of area, then there is a one-parameter family of minimal surfaces u^t bounded by Γ . If they are all without branch points, then they all have $\lambda_{\min} = 2$, and by compactness the family must loop, but Tomi’s theorem prevents that, so they must run into a branched minimal surface. To make this “run into” precise we need Böhme’s structure theorem; and to guarantee that they remain relative minima until they run into the branched surface, we need the continuity of the first two eigenvalues. Here are the details:

Theorem 7 *Let Γ be a real-analytic Jordan curve bounding infinitely many minimal surfaces without branch points (satisfying a three-point condition) and with $\lambda_{\min} \geq 2$. Then there exists a one-parameter family of minimal surfaces u^t , bounded by Γ , defined for t in some closed interval containing 0, such that for $t > 0$, u^t is a minimal surface without branch points and with $\lambda_{\min} \geq 2$, and u^0 has an (interior or boundary) branch point.*

Proof. Suppose that Γ bounds infinitely many minimal surfaces with $\lambda_{\min} \geq 2$ and without branch points. Then by compactness, there is a surface u that is a limit of a sequence u_n of such surfaces, satisfying the same three-point condition as the u_n . Since the limit of minimal surfaces is minimal, this surface is also minimal. Let \mathcal{F} be the set of minimal surfaces satisfying the same three-point condition and taking S^1 to Γ with winding number 1, but not necessarily monotonically. By (the corollary to) Böhme's structure theorem, the set of minimal surfaces in \mathcal{F} near u is analytically triangulable. Pick n large enough that u_n lies on the triangulation, i.e., u_n is the image under some analytic map j defined on a k -simplex S in R^n taking values in \mathcal{F} . Since u_n has no branch points, it takes the boundary monotonically. By Lemma 7, the set of minimal surfaces bounded by Γ near u_n is one-dimensional, so the simplex S is a 1-simplex, and there is a one-parameter family of minimal surfaces u^t defined for t in $[0, 1]$, connecting $u^0 = u$ to $u^1 = u_n$.⁵ Let q be the infimum of the set of $t \leq 1$ such that u^t has no boundary or interior branch points. This set is an open set, since it is the set where $|\nabla u|^2 > 0$ on the closed unit disk; hence $q < 1$. Now, we just forget about the u^t for $t < q$. That is, replacing t by $t + q(1 - t)$, we can assume without loss of generality that $q = 0$; that is, for $t > 0$ the surface u^t has no branch points and, in view of the fact that there are no boundary branch points, each u^t takes the boundary monotonically. Let $\phi = u_t \cdot N$. Since for $t > 0$, u^t has no branch points, Theorem 3 implies that for $t > 0$, ϕ is an eigenfunction of $\Delta\phi + 2KW\phi$. Hence 2 is an eigenvalue of $\Delta\phi + \lambda KW\phi$ for each $t > 0$. For $t = 1$ we have $\lambda_1 = 2$ and $\lambda_2 > \lambda_1$, since the least eigenvalue has a one-dimensional eigenspace. By Lemma 6, λ_1 and λ_2 depend continuously on t for $t > 0$, since for $t > 0$ we have $KW > 0$. Because the eigenspace of λ_1 is one-dimensional, we cannot have $\lambda_2 = \lambda_1$ for any value of t ; hence the two continuous functions $\lambda_1(t)$ and $\lambda_2(t)$ do not cross (or even touch), and we have $\lambda_1 = 2$ for all t in $(0, 1]$. That completes the proof.

Corollary 5 *Let Γ be a real-analytic Jordan curve bounding infinitely many relative minima of area in the C^0 metric. relative minima of area. There there exists a one-parameter family of minimal surfaces u^t , bounded by Γ , defined for t in some closed interval containing 0, such that for $t > 0$, u^t is a minimal surface without branch points and with $\lambda_{\min} \geq 2$, and u^0 has an (interior or boundary) branch point.*

Proof. This would be true with C^n in place of C^0 , if we knew that C^n relative minima have no branch points. See the discussion in Section 3.6 above.

Remark. The conclusion of the corollary can be strengthened to claim that each u^t for $t > 0$ not only has $\lambda_{\min} = 2$ but is also a relative minimum of area. We do not need this result, and it is more complicated to prove than the lemma above; in [2] it is proved by reference to a normal form theorem for the Dirichlet integral due to Tromba.

⁵Actually, Böhme's structure theorem suffices for this proof, since according to that theorem there is *some* simplex containing u_n , possibly of dimension more than 1, that extends analytically to u^0 . In that simplex we can find some analytic path.

4.2 Weierstrass Representation

We suppose we are in the following situation: a real-analytic Jordan curve, not lying in a plane, bounds a one-parameter family of minimal surfaces $u(t, z)$. The surfaces are parametrized by z in the upper half plane, so that on the real line each $u(t, \cdot)$ is a reparametrization of Γ . The surface $u(0, \cdot)$ has a boundary branch point of order $2m$ and index k at the origin. The surfaces $u(t, \cdot)$ for $t > 0$ have no branch points, and they have $\lambda_{\min} = 2$. All our arguments will be local; we shall only be concerned with what happens near the origin.⁶

We suppose that Γ passes through the origin tangent to the X -axis. Then since Γ is a Jordan curve, it is not contained in a line. We still are free to orient the Y and Z axes. We do this in such a way that the normal to $u(0, \cdot)$ at the branch point (which is well-defined) points in the positive Z -direction.

Finally, we can assume that for all t we have $u(0, 0) = 0$. If this is not already the case, then we can make it so by applying (for each $t > 0$) a conformal transformation that moves $u(0, 0)$ to the origin. If we do this while using the upper half-plane for a parameter domain, then the transformation in question is simply the map $z \mapsto z - \xi(t)$, where $u(\xi(t)) = 0$. $\xi(t)$ can be found analytically in a rational power of t , and we can replace t by a new parameter in which the family will be analytic. This assumption will be used (for example) to know that u can be found on the boundary by integrating u_x starting from 0.

The surfaces $u(t, z)$ have an Enneper-Weierstrass representation

$$u(t, z) = \operatorname{Re} \begin{bmatrix} \frac{1}{2} \int f - fg^2 dz \\ \frac{i}{2} \int f + fg^2 dz \\ \int fg dz \end{bmatrix}$$

The branch points of $u(t, \cdot)$ are the places where f and fg^2 vanish simultaneously. In [2] we were concerned with interior branch points only. In that case we can argue that since there are no branch points for $t > 0$, the zeroes of f are of even order and coincide with the poles of g , so that f and g have the forms stated in Lemma 3.3 of [2], namely:

$$\begin{aligned} f(z) &= f_0(z) \prod_{i=1}^m (z - a_i)^2 \\ g(z) &= \frac{h(z) \prod_{i=1}^{m+k} (z - b_i)}{\prod_{i=1}^m (z - a_i)} \end{aligned}$$

In this paper, boundary branch points are considered. Hence, we must consider the possibility that f and fg^2 may have some common zeroes $s_i(t)$,

⁶We do not need to assume that the branch point is a true branch point. We can rule out the existence of a one-parameter family of the type considered without that assumption. This does not, however, provide a new proof of the non-existence of false branch points which are not the terminus of a one-parameter family of relative minima of area.

corresponding to a branch point (or branch points) in the lower half plane that converge to the origin as t goes to zero. These common zeroes of f and fg^2 could occur with different multiplicities in f and fg^2 . Say that s_i occurs with multiplicity m_i in f and n_i in fg^2 . We must have $\text{Im}(s_i(t)) < 0$ for $t > 0$, since there will be branch points at the s_i , and $s_i(t)$, like $a_i(t)$ and $b_i(t)$, depends analytically on some power of t called t^γ and converging to 0 as t approaches 0.

We define

$$S(z) = \prod_{i=1} (z - s_i)^{\min(n_i, m_i)}$$

Then all zeroes of S are zeroes of f and fg^2 with at least the multiplicity they have in S . If s_i occurs with a greater multiplicity in fg^2 than in f , the extra occurrences are zeroes of g^2 , and hence of g , so $n_i - m_i$ is even. Similarly, if s_i occurs with a greater multiplicity in f than in fg^2 , the extra occurrences are poles of g^2 , and hence poles of g , so $m_i - n_i$ is even. We have

$$\frac{fg^2}{f} = g^2$$

so the remaining zeroes of f and fg^2 are double. We can therefore define functions A and B , real-analytic in t and complex-analytic in z , such that

$$\begin{aligned} f(z) &= A^2 S \\ f(z)g^2(z) &= -B^2 S \\ g(z) &= \frac{fg^2}{f} = \frac{iB}{A} \\ f(z)g(z) &= iABS \end{aligned}$$

We already defined the $a_i(t)$ and $b_i(t)$ to be the poles and zeroes of g , respectively. Let $2N$ be the number of zeroes of S . (It must be even since the number of zeroes of f is $2m$ and all the zeroes of f besides those of S are double.) Then we can find functions A_0 and B_0 , real-analytic in t and complex-analytic in z , not vanishing at the origin, such that

$$\begin{aligned} A(t, z) &= A_0(t, z) \prod_{i=1}^{m-N} (z - a_i) \\ B(t, z) &= i\lambda B_0(t, z) \prod_{i=1}^{m+k-N} (z - b_i) \end{aligned}$$

and $A_0(0, 0) = 1$, $B_0(0, 0)$ is real, and $|\lambda| = 1$ with $\lambda \neq -1$. Note that the common zeroes s_i do not enter into the formula for g . We can assume $A_0(0, 0) = 1$, since we can always change z to a constant times z . The leading term of $B(z)$, namely $i\lambda B_0(0, 0)$ is some nonzero complex constant, but because of the presence of λ in the formula, we can assume $B_0(0, 0)$ is real.

Note that with this definition of A and B , some of the a_i and/or some of the b_i may have been defined to be equal to some of the s_i , in case s_i occurs

with a greater multiplicity in f than in fg^2 or vice-versa. This is possible since these extra occurrences are double, since they occur as roots or poles of g . Each branch point s_i will either occur as an a_i , or as a b_i , or neither one (in case its multiplicities in f and fg^2 are equal); so it will still be true that $a_i(t)$ is not identically equal to $b_j(t)$ as t varies, even if the s_i occur and are treated this way.

As we have already remarked, it follows from Corollary 1 that the a_i , b_i , and s_i depend analytically on t . Although they all approach zero as t goes to zero, some may go to zero faster than others. The roots a_i , b_i , and s_i occur in “rings”, going to zero at different speeds. There is a (finite) sequence of powers $\gamma_1 < \gamma_2 < \dots$ such that every root goes to zero as t^{γ_n} for some n ; we say those roots belong to the n -th ring. Specifically, the roots in the n -th ring are those for which there exist nonzero complex constants α_i , β_i , and ζ_i such that

$$\begin{aligned} a_i &= \alpha_i t^{\gamma_n} + O(t^{\gamma_n+1}) \\ b_i &= \beta_i t^{\gamma_n} + O(t^{\gamma_n+1}) \\ s_i &= \zeta_i t^{\gamma_n} + O(t^{\gamma_n+1}) \end{aligned}$$

5 The Gauss map and the w -plane

Suppose that γ is one of the numbers γ_n ; that is, γ is a number such that some of the roots a_i , b_i , or s_i go to zero as t^γ , but not necessarily the least such number. Roots that go to zero as some higher power of t are called “fast roots”, and ones that go to zero as a lower power of t are called “slow roots”.

$$Q \text{ is the number of } a_i \text{ that go to zero as } t^\gamma \text{ or faster.} \quad (3)$$

$$S \text{ is the number of } s_i \text{ that go to zero as } t^\gamma \text{ or faster.} \quad (4)$$

$$R \text{ is the number of } b_i \text{ that go to zero as } t^\gamma \text{ or faster.} \quad (5)$$

$$P \text{ is the total number of fast roots.} \quad (6)$$

Thus if $\gamma = \gamma_1$, we have $Q = m$, $S = 2N$, and $R = m + k$. Recall that $z = a_i(t)$ are the zeroes of A and $z = b_i(t)$ are the zeroes of B . We will generally be considering only one “ring” of roots at a time, and when we speak of α_i , β_i , and ζ_i , we have in mind those roots that go to zero as t^γ :

$$\begin{aligned} a_i &= \alpha_i t^\gamma (1 + O(t)) \\ b_i &= \beta_i t^\gamma (1 + O(t)) \\ a_i &= \zeta_i t^\gamma (1 + O(t)) \end{aligned}$$

We have

$$A = A_0(t, z) \prod_{i=1}^{m-N} (z - a_i(t))$$

We introduce a new variable w by $z = t^\gamma w$. Then a factor $z - \alpha_i t^\gamma$ becomes $t^\gamma(w - \alpha_i)$, and a factor $z - \alpha_i t^{\gamma_i}$ for $\gamma_i < \gamma$ becomes

$$t^{\gamma_i}(zt^{\gamma-\gamma_i} - \alpha_i) = t^{\gamma_i}(-\alpha_i + O(t)).$$

Thus the slow roots introduce constants and powers of t , and the roots that go to zero as t^γ introduce factors containing $w - \alpha_i$, and the roots that go to zero faster than t^γ introduce powers of w times $(1 + O(t))$. All products without explicit indices will be understood to be over those roots which are $O(t^\gamma)$, and in such products (but not elsewhere) we will assume that the α_i corresponding to the fast roots are zero. Thus $\prod(w - \alpha_i)$ contains a factor of w for each fast root, and the total number of factors is the number of a_i that go to zero as t^γ or faster. Define, for the roots that go to zero as t^γ or faster,

$$\begin{aligned} \alpha &:= \text{the product of all the } -\alpha_i \text{ over the slow } a_i \\ \beta &:= \text{the product of the } -\beta_i \text{ over the slow } b_i \text{ times } B_0(0,0) \\ \zeta &:= \text{the product of the } -\zeta_i \text{ over the slow } \zeta_i \\ \mathbf{a}_i &:= a_i/t^\gamma \\ \mathbf{b}_i &:= b_i/t^\gamma \\ \mathbf{s}_i &:= s_i/t^\gamma \\ \tilde{\mathbb{A}} &:= \alpha \prod (w - \mathbf{a}_i) \quad (\text{the product is over } i \text{ such that } a_i = O(t^\gamma)) \\ \tilde{\mathbb{B}} &:= i\lambda B_0 \beta \prod (w - \mathbf{b}_i) \\ \tilde{\mathbb{S}} &:= \zeta \prod (w - \mathbf{s}_i) \end{aligned}$$

Thus \mathbf{a}_i is zero for the fast roots a_i , but it could also be zero for some roots a_i that go to zero as t^γ , and \mathbf{a}_i is not defined for the slow roots. When $t = 0$, \mathbf{a}_i becomes α_i . Over \mathbb{A} , \mathbb{B} , or \mathbb{S} , a tilde indicates t -dependence. When $t = 0$ we drop the tilde:

$$\begin{aligned} \mathbb{A} &= \alpha \prod (w - \alpha_i) \\ \mathbb{B} &= i\lambda\beta \prod (w - \beta_i) \\ \mathbb{S} &= \zeta \prod (w - \zeta_i) \end{aligned}$$

Now we consider the relation between $A(t, z)$ and $\tilde{\mathbb{A}}(t, w)$. As discussed above, the roots that go to zero as t^γ or faster contribute factors of $t^\gamma w$ or $t^\gamma(w - \alpha_i)$ to $\tilde{\mathbb{A}}$, and the slow roots contribute factors $-\alpha_i t^{\gamma_i}$, with $\gamma_i < \gamma$. The product of the factors contributed by all the slow roots is thus a constant α (defined above) times a power t^d of t , where d is the sum of the γ_i over all the slow roots a_i . Since the constant α is part of the definition of \mathbb{A} , it does not show explicitly in the relation between A and \mathbb{A} :

$$A = t^d t^{Q\gamma} \mathbb{A}(1 + O(t)) \tag{7}$$

Similarly, we have numbers e and f such that

$$B = t^e t^{R\gamma} \mathbb{B} (1 + O(t)) \quad (8)$$

$$S = t^f t^{S\gamma} \mathbb{S} (1 + O(t)) \quad (9)$$

No ambiguity will result from using f for this power of t since in this meaning, it will always occur in an exponent, and the Weierstrass function f never occurs in an exponent. Similarly the letter “ S ” has to do double duty, in exponents as the number of roots s_i that go to zero as t^γ or faster, and not in exponents as a function of z , as can be seen in (9). (We are short of letters in the alphabet, and the function S will eventually disappear.) We define

$$\mathcal{K} := (R - Q)\gamma + (e - d). \quad (10)$$

The importance of \mathcal{K} stems from the following equation, which results from dividing (8) by (7):

$$\frac{B}{A} = t^{\mathcal{K}} \frac{\mathbb{B}}{\mathbb{A}} + O(t^{\mathcal{K}+1}). \quad (11)$$

This equation holds on any compact subset of the w -plane away from the zeroes of \mathbb{A} . When $\gamma = \gamma_1$, so we are working with the first-ring roots, then $\mathcal{K} = k\gamma_1$, since no roots go to zero slower than γ_1 , so $e = d = 0$, and $R - Q = k$.

Our next aim is transform our formulas for the Weierstrass representation from their expressions in z and t to express them in terms of w and t . Starting with the expressions in terms of z and t , we have

$$\begin{aligned} u &= \operatorname{Re} \begin{bmatrix} \frac{1}{2} \int f - fg^2 dz \\ \frac{i}{2} \int f + fg^2 dz \\ \int fg dz \end{bmatrix} \\ &= \operatorname{Re} \begin{bmatrix} \frac{1}{2} \int A^2 S + B^2 S dz \\ \frac{i}{2} \int A^2 S - B^2 S dz \\ i \int ABS dz \end{bmatrix} \end{aligned}$$

Now we express A , B , and S as appropriate powers of t times $\tilde{\mathbb{A}}$, $\tilde{\mathbb{B}}$, and $\tilde{\mathbb{S}}$, as given above; and we note that since $z = t^\gamma w$, we have $dz = t^\gamma dw$, so when we change the variable of integration to w , we pick up an extra power of t^γ . Replacing $\tilde{\mathbb{A}}$ by $(1 + O(t))\mathbb{A}$, and similarly for $\tilde{\mathbb{B}}$ and $\tilde{\mathbb{S}}$, we have by (7), (8), and (9)

$$u = (1 + O(t)) \operatorname{Re} \begin{bmatrix} \frac{1}{2} \int t^{2d+f} t^{(2Q+2S+1)\gamma} \mathbb{A}^2 \mathbb{S} + t^{2e+f} t^{(2R+2S+1)\gamma} \mathbb{B}^2 \mathbb{S} dw \\ \frac{i}{2} \int t^{2d+f} t^{(2Q+2S+1)\gamma} \mathbb{A}^2 \mathbb{S} - t^{2e+f} t^{(2R+2S+1)\gamma} \mathbb{B}^2 \mathbb{S} dw \\ \int t^{d+e+lf} t^{(Q+R+S+1)\gamma} \mathbb{A} \mathbb{B} \mathbb{S} dw \end{bmatrix}$$

Since the expressions for these exponents are a bit unwieldy, we shorten them by introducing

$$\mathcal{M} := 2d + f + (2Q + S + 1)\gamma \quad (12)$$

With this definition we have

$$A^2 S = t^{\mathcal{M}-1} \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} = \mathbb{A}^2 (1 + O(t)) \quad (13)$$

$$B^2 S = t^{\mathcal{M}-1+2\mathcal{K}} \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}} = \mathbb{B}^2 (1 + O(t)) \quad (14)$$

Note that $dz = t^\gamma dw$, so when we integrate we get an extra factor of t^γ . Using \mathcal{M} and \mathcal{K} , the previous equation becomes

$$u = (1 + O(t)) \operatorname{Re} \left[\begin{array}{l} \frac{1}{2} \int t^{\mathcal{M}} \mathbb{A}^2 \mathbb{S} + t^{\mathcal{M}+2\mathcal{K}} \mathbb{B}^2 \mathbb{S} dw \\ \frac{1}{2} \int t^{\mathcal{M}} \mathbb{A}^2 \mathbb{S} - t^{\mathcal{M}+2\mathcal{K}} \mathbb{B}^2 \mathbb{S} dw \\ \int t^{\mathcal{M}+\mathcal{K}} \mathbb{A} \mathbb{B} \mathbb{S} dw \end{array} \right] \quad (15)$$

To get an intuitive grasp of these formulas, consider the special case in which we are working with the first-ring roots. Then $2Q + S + 1 = 2m + 1$, and if there are also no slow roots then $e = d = f = 0$, so $\mathcal{M} = (2m + 1)\gamma$. If there are some fast roots they contribute extra powers of t , namely t^{e+d+f} . In general, we will find that \mathcal{M} plays the role that $(2m + 1)\gamma$ would play in the simplest case.

We next review the formula for the unit normal. The function g in the Weierstrass representation, which is the stereographic projection of the unit normal N , is given by $g(z) = iB/A$, and the normal itself is given by

$$N = \frac{1}{|g|^2 + 1} \begin{bmatrix} 2\operatorname{Re}(g) \\ 2\operatorname{Im}(g) \\ |g|^2 - 1 \end{bmatrix}. \quad (16)$$

We write \bar{A} for the complex conjugate of A . Substituting $g = iB/A$, we have

$$N = \frac{1}{1 + |B/A|^2} \begin{bmatrix} -2\operatorname{Im}(B/A) \\ 2\operatorname{Re}(B/A) \\ -1 + |B/A|^2 \end{bmatrix}$$

Expressing this in terms of $\tilde{\mathbb{B}}/\tilde{\mathbb{A}}$, we have $B/A = t^{\mathcal{K}} \tilde{\mathbb{B}}/\tilde{\mathbb{A}}$, and hence

$$N = \frac{1}{1 + t^{2\mathcal{K}} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2} \begin{bmatrix} -2t^{\mathcal{K}} \operatorname{Im}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) \\ 2t^{\mathcal{K}} \operatorname{Re}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) \\ -1 + t^{2\mathcal{K}} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2 \end{bmatrix} \quad (17)$$

We remind the reader that 1N is the first component of N , and 2N the second component. We have

$$\frac{{}^1N}{{}^2N} = -\frac{\operatorname{Im}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}})}{\operatorname{Re}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}})} \quad (18)$$

We record in a lemma the geometric significance of \mathcal{K} :

Lemma 8 $\mathcal{K} > 0$ if and only if the unit normal N converges to $(0, 0, -1)$ on compact subsets of the w -plane away from the zeroes of \mathbb{A} , as t goes to zero.

Proof. The stereographic projection of the unit normal N is B/A , which by (11) is given by $t^{\mathcal{K}} \tilde{\mathbb{B}}/\tilde{\mathbb{A}}$. This goes to zero on compact subsets of the w plane if and only if $\mathcal{K} > 0$; but 0 is the stereographic projection of $(0, 0, -1)$. Alternately the lemma can be proved by direct examination of (17).

Lemma 9 *Suppose that on compact subsets of the w -plane away from the zeroes of \mathbb{A} , N converges to $(0, 0, -1)$ as t goes to zero. Then*

(i) \mathbb{A} , \mathbb{S} , and \mathbb{B} are real when w is real;

(ii) the α_i and ζ_i are real and the β_i are real and/or occur in complex-conjugate pairs.

Remark. The hypothesis is certainly true for the first ring of a_i , the slowest roots. We will later show inductively that it is true for all the rings of a_i , i.e. when each γ_n is used to define w .

Proof. First we prove that α_i does not have positive imaginary part. Suppose, for proof by contradiction, that it does. Then consider a circle of radius ρ about α_i , where ρ small enough that the entire circle lies in the upper half plane and each other α_i is either inside or outside the circle. Then by hypothesis, on the boundary of the circle, N goes to $(0, 0, -1)$ as t goes to zero. Hence the Gauss map covers the upper hemisphere inside the circle, so the least eigenvalue of u^t is less than 2, for $t > 0$ sufficiently small; but that is a contradiction. That contradiction proves that α_i does not have positive imaginary part.

Let $Y(w) := {}^2u(t^\gamma w)$, where again, 2u is the second component of u . Then for real w

$$Y(w) = \frac{C_2}{q+1} \tau^{q+1} (1 + O(s))$$

where τ , q , and C_2 are as in (1), and $q > 0$.

$$\begin{aligned} \tau &= X(w) = \int_0^z A^2 S + B^2 S dz \\ &= t^{\mathcal{M}} \int_0^w \mathbb{A}^2 \mathbb{S} + t^{2\mathcal{K}} \mathbb{B}^2 dw (1 + O(t)) \end{aligned}$$

by (13) and (10). The hypothesis that N converges to $(0, 0, -1)$ away from the zeroes of \mathbb{A} implies $\mathcal{K} > 0$, by Lemma 8. Hence the term in $t^{2\mathcal{K}}$ can be absorbed into the error term, and we have for real w

$$\tau = t^{\mathcal{M}} \int_0^w \mathbb{A}^2 \mathbb{S} dw (1 + O(t))$$

and hence

$$Y(w) = \frac{C_2}{q+1} t^{(q+1)\mathcal{M}} \int_0^w \mathbb{A}^2 \mathbb{S} dw (1 + O(t))$$

On the other hand from the Weierstrass representation (15), we have

$$Y(w) = t^{\mathcal{M}} \operatorname{Im} \int_0^w \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} dw + O(t^{2d+2Q\gamma+f+S\gamma+2})$$

Since $q > 0$, the power of t shown explicitly in this equation is less than the power of t in the preceding equation. It follows that

$$\operatorname{Im} \int_0^w \mathbb{A}^2 \mathbb{S} dw = 0$$

since it is independent of t but from the two previous equations must be $O(t)$. Therefore the polynomial $\int_0^w \mathbb{A}^2 \mathbb{S} dw$ is real on the real axis. It follows that its coefficients are real. Hence the coefficients of its derivative $\mathbb{A}^2 \mathbb{S}$ are real. Hence the roots of $\mathbb{A}^2 \mathbb{S}$, which are the α_i and ζ_i , come in complex-conjugate pairs. But we proved above that no α_i has positive imaginary part. Hence they are all real when $t = 0$. Similarly, the ζ_i do not have positive imaginary part, since u^t has no branch points for $t > 0$. Hence the ζ_i are real when $t = 0$.

Similarly, with $Z(w) = {}^3u(t^\gamma)w$ in place of Y , we find

$$\operatorname{Im} \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw = 0$$

That is, $\int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw = 0$ is real on the real axis. Therefore also its derivative $\mathbb{A} \mathbb{B} \mathbb{S}$ is real on the real axis. But $\mathbb{A} \mathbb{B} \mathbb{S}$ is a polynomial; since it is real on the real axis, its coefficients are real. We have already proved that $\mathbb{A} \mathbb{S}$ is real on the real axis; hence \mathbb{B} is also real on the real axis. Hence its roots, which are the $\hat{\mathbf{b}}_i$ when $t = 0$, are real or occur in complex-conjugate pairs. That completes the proof of the lemma.

6 The eigenfunction

In this section, we compute the eigenfunction of the variational problem associated with the second variation of area. This is shown in [2] to be $\phi = u_t \cdot N$, where $u_t = du/dt$ is the ‘‘tangent vector’’ to the one-parameter family, and N is the unit normal. Only one assumption will be made for our eigenfunction computation: we assume that $\mathcal{K} > 0$. (See (7), (8), (9), and (10) for the definition of \mathcal{K} and the quantities used to define \mathcal{K} .) This assumption ensures that the unit normal, whose stereographic projection is B/A , is equal to t to a positive power times \mathbb{B}/\mathbb{A} , so that on compact subsets of the w -plane away from zeroes of \mathbb{A} , the unit normal tends to $(-1, 0, 0)$. Specifically we have

$$\begin{aligned} \frac{B}{A} &= t^{(R-Q)\gamma+(e-d)} \frac{\tilde{\mathbb{B}}}{\tilde{\mathbb{A}}} \\ &= t^{\mathcal{K}\gamma} \frac{\tilde{\mathbb{B}}}{\tilde{\mathbb{A}}} \quad \text{by (10)} \end{aligned}$$

and $\mathcal{K} > 0$, by our assumptions. Differentiating the Weierstrass representation for u , we have

$$2u_t = \frac{d}{dt} \operatorname{Re} \begin{bmatrix} \int_0^z (A^2 S + B^2 S) dz \\ i \int_0^z (A^2 S - B^2 S) dz \\ 2i \int_0^z A B S dz \end{bmatrix}$$

Expressing this in terms of $\tilde{\mathbb{A}}$ and $\tilde{\mathbb{B}}$ we have by (13) and (14)

$$2u_t = \frac{d}{dt} \operatorname{Re} \begin{bmatrix} t^{\mathcal{M}} \int_0^w (\tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} + t^{2\mathcal{K}} \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}}) dw \\ i t^{\mathcal{M}} \int_0^w (\tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} - t^{2\mathcal{K}} \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}}) dw \\ 2i t^{\mathcal{M}+\mathcal{K}} \int_0^w \tilde{\mathbb{A}} \tilde{\mathbb{B}} \tilde{\mathbb{S}} dw \end{bmatrix} \quad (19)$$

The rule for differentiating with respect to t in this situation is found by applying the chain rule as follows, where d/dt means the partial derivative with respect to t , holding z fixed, and $\partial/\partial t$ means the partial derivative with respect to t , holding w fixed.

$$\begin{aligned}
dH/dt &= \frac{\partial H}{\partial t} + \frac{dH}{dw} \frac{dw}{dt} \\
&= \frac{\partial H}{\partial t} + \frac{dt^{-\gamma} z}{dt} \frac{dH}{dw} \\
&= \frac{\partial H}{\partial t} - \gamma t^{-\gamma-1} z \frac{dH}{dw} \\
&= \frac{\partial H}{\partial t} - \gamma t^{-\gamma-1} t^\gamma w \frac{dH}{dw}
\end{aligned}$$

The final result for differentiating with respect to t is then

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} - t^{-1} \gamma w \frac{dH}{dw} \quad (20)$$

Applying this rule on the right side of (19), we find

$$\begin{aligned}
2u_t = & \\
& \operatorname{Re} \left[\begin{aligned} & t^{\mathcal{M}-1} \mathcal{M} \int_0^w \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} dw + t^{\mathcal{M}+2\mathcal{K}-1} (\mathcal{M} + 2\mathcal{K}) \int_0^w \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}} dw \\ & it^{\mathcal{M}-1} \mathcal{M} \int_0^w \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} dw + it^{\mathcal{M}+2\mathcal{K}-1} (\mathcal{M} + 2\mathcal{K}) \int_0^w \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}} dw \\ & 2it^{\mathcal{M}+\mathcal{K}-1} (\mathcal{M} + \mathcal{K}) \int_0^w \tilde{\mathbb{A}} \tilde{\mathbb{B}} \tilde{\mathbb{S}} dw \end{aligned} \right] \\
& - \operatorname{Re} \left[\begin{aligned} & t^{\mathcal{M}-1} \gamma \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} w + t^{\mathcal{M}+2\mathcal{K}-1} \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}} w \\ & it^{\mathcal{M}-1} \gamma \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} w + it^{\mathcal{M}+2\mathcal{K}-1} \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}} w \\ & 2it^{\mathcal{M}+\mathcal{K}-1} \gamma \tilde{\mathbb{A}} \tilde{\mathbb{B}} \tilde{\mathbb{S}} w \end{aligned} \right] \\
& + \operatorname{Re} \left[\begin{aligned} & t^{\mathcal{M}} \mathcal{M} \frac{\partial}{\partial t} \int_0^w \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} dw + t^{\mathcal{M}+2\mathcal{K}} (\mathcal{M} + 2\mathcal{K}) \frac{\partial}{\partial t} \int_0^w \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}} w \\ & it^{\mathcal{M}} \mathcal{M} \frac{\partial}{\partial t} \int_0^w \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} dw + it^{\mathcal{M}+2\mathcal{K}} (\mathcal{M} + 2\mathcal{K}) \frac{\partial}{\partial t} \int_0^w \tilde{\mathbb{B}}^2 \tilde{\mathbb{S}} dw \\ & 2it^{\mathcal{M}+\mathcal{K}} (\mathcal{M} + \mathcal{K}) \frac{\partial}{\partial t} \int_0^w \tilde{\mathbb{A}} \tilde{\mathbb{B}} \tilde{\mathbb{S}} dw \end{aligned} \right]
\end{aligned}$$

The terms in the third vector, involving $\partial/\partial t$, have one higher power of t and are $O(t^{\mathcal{M}})$. These swamp the $t^{\mathcal{M}+2\mathcal{K}-1}$ terms. Eliminating the swamped terms we have

$$\begin{aligned}
2u_t = & \\
& \operatorname{Re} \left[\begin{aligned} & t^{\mathcal{M}-1} \mathcal{M} \int_0^w \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} dw - t^{\mathcal{M}-1} \gamma \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} w + O(t^{\mathcal{M}}) \\ & it^{\mathcal{M}-1} \mathcal{M} \int_0^w \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} dw - it^{\mathcal{M}-1} \gamma \tilde{\mathbb{A}}^2 \tilde{\mathbb{S}} w + O(t^{\mathcal{M}}) \\ & 2it^{\mathcal{M}+\mathcal{K}-1} (\mathcal{M} + \mathcal{K}) \int_0^w \tilde{\mathbb{A}} \tilde{\mathbb{B}} \tilde{\mathbb{S}} dw - 2it^{\mathcal{M}+\mathcal{K}-1} \gamma \tilde{\mathbb{A}} \tilde{\mathbb{B}} \tilde{\mathbb{S}} w + O(t^{\mathcal{M}+\mathcal{K}}) \end{aligned} \right]
\end{aligned}$$

We may also drop the tildes over \mathbb{A} , \mathbb{B} , and \mathbb{S} , since the terms with nonzero powers of t can be absorbed in the error terms. Dropping the tildes and dividing by the leading power of t , we have

$$\frac{2u_t}{t^{\mathcal{M}-1}} = \operatorname{Re} \left[\begin{aligned} & \mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w + O(t) \\ & \mathcal{M} i \int_0^w \mathbb{A}^2 \mathbb{S} dw - i \gamma \mathbb{A}^2 \mathbb{S} w + O(t) \\ & 2it^{\mathcal{K}} (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w + O(t^{\mathcal{K}+1}) \end{aligned} \right]$$

Taking the dot product with N , we have the following formula for the eigenfunction, valid in the whole w -plane, including subsets that contain the zeroes of \mathbb{A} :

$$\begin{aligned} \frac{2}{t^{\mathcal{M}-1}}\phi &= \tag{21} \\ & [{}^1N] \operatorname{Re} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w + O(t) \right) \\ & - [{}^2N] \operatorname{Im} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w + O(t) \right) \\ & - t^{\mathcal{K}} (\mathcal{M} + \mathcal{K}) [{}^3N] 2 \operatorname{Im} \left(\int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w + O(t) \right) \end{aligned}$$

Lemma 10 (The limit of the eigenfunction) *The following two formulas for the least eigenfunction of u^t in terms of N are valid on each compact subset of the w plane away from the zeroes of \mathbb{A} , provided $\mathcal{K} > 0$.*

$$\begin{aligned} \frac{-1}{t^{(\mathcal{M}+\mathcal{K})\gamma-1}}\phi &= \operatorname{Im} \left\{ \frac{\mathbb{B}}{\mathbb{A}} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w \right) \right. \\ & \quad \left. - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w \right\} + O(t) \\ &= \operatorname{Im} \left\{ \frac{\mathbb{B}}{\mathbb{A}} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw \right) - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw \right\} + O(t) \end{aligned}$$

Remark. The provision that $\mathcal{K} > 0$ is the same as providing that N approach $(0, 0, -1)$ on compact subsets away from the zeros of \mathbb{A} . It is certainly true when $\gamma = \gamma_1$, and later we will inductively show it is true on the inner rings of roots as well.

Proof. Starting with formula (21), we substitute for 1N , 2N , and 3N the values obtained in (17), to express the formula entirely in terms of $\tilde{\mathbb{B}}$ and $\tilde{\mathbb{A}}$. We get (after multiplying by the denominator and changing the signs of both sides)

$$\begin{aligned} \frac{-(1 + t^{2\mathcal{K}} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2)}{t^{\mathcal{M}-1}}\phi &= \\ & \left(t^{\mathcal{K}} \operatorname{Im} (\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) \right) \operatorname{Re} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w + O(t) \right) \\ & + \left(t^{\mathcal{K}} \operatorname{Re} (\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) \right) \operatorname{Im} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w + O(t) \right) \\ & + t^{\mathcal{K}\gamma} (\mathcal{M} + 2\mathcal{K}) \left(-1 + t^{2\mathcal{K}} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2 \right) 2 \operatorname{Im} \left(\int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w + O(t) \right) \end{aligned}$$

Now we separate out the error terms:

$$\frac{-(1 + t^{2\mathcal{K}} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2)}{t^{\mathcal{M}-1}}\phi =$$

$$\begin{aligned}
& \left(t^\mathcal{K} \operatorname{Im}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) \right) \operatorname{Re} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w \right) \\
& + \left(t^\mathcal{K} \operatorname{Re}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) \right) \operatorname{Im} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w \right) \\
& + t^\mathcal{K} (\mathcal{M} + 2\mathcal{K}) \left(-1 + t^{2\mathcal{K}\gamma} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2 \right) 2 \operatorname{Im} \left(\int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w \right) \\
& + t^\mathcal{K} \operatorname{Im}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) O(t) + t^\mathcal{K} (\mathbb{B}/\mathbb{A}) O(t) + t^\mathcal{K} O(t) + |\mathbb{B}/\mathbb{A}|^2 O(t^{3\mathcal{K}})
\end{aligned}$$

Now we combine real and imaginary parts according to the rule

$$\operatorname{Re}(u)\operatorname{Im}(v) + \operatorname{Re}(v)\operatorname{Im}(u) = \operatorname{Im}(uv),$$

applied to the first two terms on the right, and divide both sides by $t^\mathcal{K}$. That yields

$$\begin{aligned}
\frac{-(1 + t^{2\mathcal{K}} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2)}{t^{\mathcal{M}+\mathcal{K}-1}} \phi &= \operatorname{Im} \left\{ (\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w \right) \right\} \\
&- \left(1 - t^{2\mathcal{K}} |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2 \right) 2 \operatorname{Im} \left\{ (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w \right\} \\
&+ \operatorname{Re}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) O(t) + \operatorname{Im}(\tilde{\mathbb{B}}/\tilde{\mathbb{A}}) O(t) + |\tilde{\mathbb{B}}/\tilde{\mathbb{A}}|^2 O(t^{2\mathcal{K}})
\end{aligned}$$

That formula for the least eigenfunction of u^t in terms of N is valid everywhere, even near the zeroes of \mathbb{A} , in the sense that the big- O terms are analytic in t and w .

Now replace $t^\mathcal{K} \tilde{\mathbb{B}}/\tilde{\mathbb{A}}$ with $O(t)$ on the left, and noting that $\tilde{\mathbb{B}} = \mathbb{B} + O(t)$, and $\tilde{\mathbb{A}} = \mathbb{A} + O(t)$, and on a given compact subset away from zeros of \mathbb{A} , we have $\tilde{\mathbb{B}}/\tilde{\mathbb{A}} = \mathbb{B}/\mathbb{A} + O(t)$. That yields the first formula of the lemma. The second formula follows by canceling two terms in the first formula. That completes the proof of the lemma.

We now define the function \mathbb{H} by

$$\mathbb{H}(w) = \frac{\mathbb{B}}{\mathbb{A}} \mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw \quad (22)$$

Then by the previous lemma we have

$$\operatorname{Im} \mathbb{H} = \lim_{t \rightarrow 0} \frac{-1}{t^{\mathcal{M}+\mathcal{K}-1}} \phi$$

on compact subsets of the w -plane away from the zeroes of \mathbb{A} . A crucial question is whether \mathbb{H} is constant or not. If it is constant, then all the visible terms cancel out, and we can thus get no information from them. If it is not constant, we can get some information by the asymptotic behavior of \mathbb{H} near the origin or near α_i .

7 Gaussian area and the eigenfunction

We next draw some consequences from the calculations of the eigenfunction and the properties of \mathbb{H} . Recall that by Lemma 9, all the α_i are real. Hence we may guess that the X -component of u_t does not change sign near α_i . Now consider the “east pole” and the “west pole” on the Riemann sphere, namely $(1, 0, 0)$ and $(-1, 0, 0)$. Can the unit normal take on both values near α_i ? If it did, then since the eigenfunction $\phi = u_t \cdot N$, that would imply that the X -component of u_t would have to change sign near α_i , which seems unlikely. Thus, we hope to show that N cannot take on both the east pole and the west pole near α_i . But since N is confined to be perpendicular to Γ on the boundary, and α_i is real, it intuitively seems that N must then be confined to one hemisphere. These ideas are the basis of the following rigorous developments.

Lemma 11 *Suppose that on compact subsets of the upper half of the w -plane away from the zeroes of \mathbb{A} , N converges to $(0, 0, -1,)$ as t goes to zero. Then exactly one of the points $(1, 0, 0)$ and $(-1, 0, 0)$ (the “east pole” and “west pole”) can be taken by N at points in the upper half plane in the vicinity of one α_i , and that point is taken on at most J times, where J is the number of \mathbf{a}_j that converge to α_i .*

Proof. Fix i , and fix a disk U centered at α_i and excluding all other α_i and zero. Consider a point ξ in U at which the normal takes on $(1, 0, 0)$ or $(-1, 0, 0)$. Let U^+ be the intersection of U with the upper half plane and let U^- be the intersection of U with the lower half plane. Since $N = t^{\mathcal{K}} \tilde{\mathbb{B}} / \tilde{\mathbb{A}}$, the hypothesis about the convergence of N is equivalent to $\mathcal{K} > 0$. Hence N is $O(t)$ also on the boundary of U . There are J values of j for which \mathbf{a}_j lies in U (for sufficiently small t), and at these points N takes on the north pole. Hence in the whole neighborhood U , the map N is a J -sheeted covering of the portion of the Riemann sphere lying north of a small circle of latitude within $O(t)$ of the south pole. Hence each of the east pole and the west pole can be taken on at most J times over the whole neighborhood U .

We will show that if one of the east or west poles is taken on in U^+ , the other cannot be. At ξ we have ${}^1N = \pm 1$, and ${}^2N = {}^3N = 0$. The formula (21) calculated above for the eigenfunction ϕ in terms of the components of N simplifies greatly when we substitute ${}^2N = {}^3N = 0$. According to that formula, at ξ we have

$$\frac{2}{t^{\mathcal{M}-1}} \phi = {}^1N \operatorname{Re} \left(\mathcal{M} \int_0^\xi \mathbb{A}^2 \mathbb{S} dw - \mathbb{A}^2(\xi) \mathbb{S}(\xi) \xi + O(t) \right)$$

Note that

$$\int_0^\xi \mathbb{A}^2 \mathbb{S} dw = \int_0^{\alpha_i} \mathbb{A}^2 \mathbb{S} dw + \int_{\alpha_i}^\xi \mathbb{A}^2 \mathbb{S} dw$$

The first term is a constant and the second is $O(\xi - \alpha_i)$. Hence the previous equation can be written as

$$\frac{2}{t^{\mathcal{M}-1}} \phi = {}^1N c(1 + O(\xi - \alpha_i)) - \tilde{\mathbb{A}}^2(\xi) \tilde{\mathbb{S}}(\xi) \xi + O(t)$$

where c is the constant $\mathcal{M} \int_0^{\alpha_i} \mathbb{A}^2 \mathbb{S} d\xi$. The constant c is nonzero since (i) α_i is real, by Lemma 9, and (ii) the integrand $\mathbb{A}^2 \mathbb{S}$ is nonnegative on the real axis, because u takes the boundary monotonically and $u_x = t^{\mathcal{M}} \mathbb{A}^2 \mathbb{S} (1 + O(t))$, and (iii) $\alpha_i \neq 0$, since a_i goes to zero as t^γ .

As t goes to zero, ξ will converge to α_i , since by assumption outside every disk centered at α_i , N converges to $(0, 0, -1)$. We do not require that ξ depend smoothly on t ; a sequence of values of t going to zero will suffice, or even a single choice of t depending appropriately on the constants hidden in the $O(t)$ and $O(w)$ terms. Therefore the $O(\xi - \alpha_i)$ term goes to zero too; that is, $\xi - \alpha_i(t) = O(t)$. We have

$$\frac{2}{t^{\mathcal{M}-1}} \phi(\xi) = {}^1N(\xi) c + O(t) \quad \text{where } c = \mathcal{M} \int_0^{\alpha_i} \mathbb{A}^2 \mathbb{S} \quad (23)$$

For small enough t , we see that the sign of $\phi(w)$ is equal to the sign of 1N if $\alpha > 0$, or is opposite to the sign of 1N if $\alpha < 0$. In particular, for all such points in the vicinity of a fixed α_i , the sign of 1N is the same, since the sign of ϕ is the same in the entire upper half plane, because ϕ is the eigenfunction of the least eigenvalue. That proves that not both the east and the west pole can be taken on over U^+ near α_i . But the west (or east) pole can be covered at most J times, since N is a J -sheeted covering over U . That completes the proof.

Discussion. The significance of this lemma is that the Gauss map, considered as a covering map, covers at most one hemisphere (plus $O(t)$ in a narrow band near the great circle dividing the two hemispheres, whose plane is perpendicular to the X -axis), since if it could be extended across the narrow strip near the YZ plane where the image of the boundary must lie, it would reach the other of the two points in question. Of course that hemisphere could be covered multiple times; or possibly not at all, as it is conceivable that \tilde{a}_i has a slightly negative imaginary part for $t > 0$ and that N then covers the entire sphere in the lower half plane near a_i , even though \tilde{a}_i tends to the real number α_i . At this point we have not ruled out the possibility that one or more sheets of the covering might lie entirely over the lower half U_- of U . The argument does not work for U_- , because although ϕ continues analytically into U_- , we do not know that it has the same sign throughout U_- .

However, the proof does permit us to extract a bit more information than is stated in the above lemma. For each α_i , we have shown that there is one ‘‘permitted hemisphere’’, either the west hemisphere or the east hemisphere, according as the east pole or the west pole may possibly be a value of N near that α_i . We can say more about which hemisphere is permitted:

Lemma 12 *All positive α_i have the same permitted hemisphere, and all negative α_i have the opposite permitted hemisphere.*

Proof. Let ξ be a point near α_i where the east or west pole is taken on. The formula for $\phi(\xi)$ given in (23) shows that $\phi(\xi)$ has the same sign as ${}^1N(\xi)$ when $\alpha_i > 0$, since then $\int_0^{\alpha_i} \mathbb{A}^2 \mathbb{S} dw > 0$, and the opposite sign with $\alpha_i < 0$,

since then the integral is negative. But ϕ , since it is the eigenfunction for the least eigenvalue, has the same sign throughout the upper half plane. Hence the permitted sign of 1N is the same for all positive α_i , and the opposite sign for all negative α_i . That completes the proof.

Definition 2 *We say that α_i contributes Gaussian area if, in the vicinity of α_i , the normal restricted to a small upper half-disk covers an area at least $2\pi - O(t)$ on the sphere. To make this definition precise, let D_ρ be a disk of radius ρ about α_i in the w -plane, excluding all other α_j and zero, and let D_ρ^+ be its intersection with the upper half plane. Then α_i contributes Gaussian area if, for every $\rho > 0$, $N(D_\rho^+)$ has area at least $2\pi - O(t)$ for sufficiently small t . In case $N(D_\rho^+)$ has area $J\pi - O(t)$ then we say α_i contributes Gaussian area $J\pi$.*

Since the normal N is a projection mapping (away from its ramification points), and since the image of the boundary is confined to within $O(t)$ of the YZ -plane, as soon as N takes on a value more than $O(t)$ from the YZ -plane (in D_t^+) then α_i contributes Gaussian area, and in that case it must contribute Gaussian area $2J\pi$ for some J , i.e. area equal to an integer number of hemispheres. In particular, if N takes on the ‘‘east pole’’ $(1, 0, 0)$ or the ‘‘west pole’’ $(-1, 0, 0)$ in D_t^+ for every $t > 0$, then α_i contributes Gaussian area.

Discussion. How much Gaussian area is contributed by each a_i ? We hope to show just one hemisphere, rather than one sphere. We know that no a_i can contribute a whole sphere in the most obvious way, as then the Gaussian image would contain more than a hemisphere and the eigenvalue would be less than 2. Moreover in the previous lemma we showed that the image of the boundary sticks up at least to the equator, hinting that it divides the sphere into hemispheres. But this still does not rule out the possibility that somehow a whole sphere of Gaussian area might be contributed. In the end, we are going to count the Gaussian area, and it will be vital that each a_i contributes only a hemisphere. The next lemma addresses this crucial issue. The key to the proof is the fact that near a_i , and in the upper half-plane, N can take on the east or the west pole, but not both. That means it is confined to a hemisphere. Here are the details:

Lemma 13 *Assume that N goes to $(0, 0, -1)$ on compact subsets of the w -plane away from the zeroes of \mathbb{A} . Then every α_i contributes Gaussian area at most equal to one hemisphere for each \mathbf{a}_j that converges to α_i .*

Proof. According to the Gauss-Bonnet theorem (discussed in Section 3.4), we have when $t = 0$

$$\int KW \, dx \, dy + \int_{\Gamma} \kappa_g = 2\pi + 2M\pi$$

where M is the sum of the orders of the boundary branch points and twice the sum of the orders of the interior branch points. When $t > 0$, there are no branch points in the closed upper half plane, so we have

$$\int KW \, dx \, dy + \int_{\Gamma} \kappa_g = 2\pi$$

We let t go to zero in this formula. Since κ_g is bounded by the curvature of Γ , the integral containing κ_g converges as t goes to zero to the corresponding term in the first formula. Hence the first term, which represents the Gaussian area, must jump upwards by $2M\pi$. That is, each interior branch point contributes one sphere (times the order) and each boundary branch point contributes a hemisphere (times the order).

Near each a_i , disregarding the boundary, one whole sphere of Gaussian area (times the order) will exist in a small neighborhood of a_i for small positive t , within $O(t)$. Since the Gaussian image of the boundary is confined to lie near the YZ plane, and since (in the case of several nearby a_i 's) the Gauss map may be a multi-sheeted covering of the sphere, there are (*a priori*) several possibilities. For example, there may be no Gaussian area contributed (in the upper half plane), if the whole "bubble" is taken on in the lower half plane. Or, the "bubble" may be divided by the image of the boundary, so that one hemisphere is taken on in the upper half plane, and one in the lower half plane. Or, if there is more than one sheet of the covering in a small neighborhood of α_i , which is possible if several \mathbf{a}_j converge to the same α_i , then some sheets may be divided by the boundary and others not. *A priori* it is also possible that the image of the boundary may rise part way up the sphere and then reverse direction, so as to only partly divide the sphere.

However, we can now rule out some of these possibilities. Let p be one of the α_i or zero. Fix a neighborhood U of p in the w -plane small enough to exclude all other α_i , or all α_i if p is zero, and let U^+ be the intersection of U with the upper half-plane. We know that each α_i is real, so U^+ is half of U . Let J be the number of \mathbf{a}_j that converge to p . Then the total extra Gaussian area over U is $4\pi J$, and over U^+ , the Gauss map is a J -sheeted branched covering of the part of the Riemann sphere north of any fixed circle of latitude (say in the southern hemisphere). That is, fixing such a circle of latitude, for t sufficiently small N will be a branched covering of the upper part of the Riemann sphere, north of the fixed circle of latitude.

If this α_i contributes any Gaussian area at all (in the upper half plane), then N takes on either the east or west pole somewhere in U^+ . Let us suppose that N takes on the west pole W (the other case is symmetric). It may do so on more than one sheet. On each sheet (that is, starting from each different pre-image of W , we can continue a path π such that $N(\pi(\tau))$ lies on the equator (even if the path must pass through ramification points of N). But the image $N(\pi(\tau))$ cannot extend into the eastern hemisphere by more than $O(t)$, or else it could be continued all the way to the east pole, contradicting Lemma 12. Over the whole neighborhood U there are J sheets covering the sphere J times, for a total area of $4\pi J$. But now we have proved that each of these J sheets has at most the west hemisphere covered by $N(U^+)$.

Hence the total Gaussian area contributed by α_i is at most $2\pi J$; that is, one hemisphere per a_i at the most. That completes the proof of the lemma.

8 What is true if H is not constant

In this section, we suppose that γ is one of the γ_n , that $z = t^\gamma w$, and that N goes to $(0, 0, -1)$ on compact subsets away from the zeroes of \mathbb{A} . This condition we call the “ N -condition”. Since

$$\frac{B}{A} = t^\mathcal{K} \frac{\mathbb{B}}{\mathbb{A}}$$

the N -condition can be succinctly expressed as $\mathcal{K} > 0$. The α_i are the coefficients of the a_i in this ring of roots, i.e., $a_i = \alpha_i t^\gamma (1 + O(t))$, and we define $\alpha_i = 0$ for roots that go to zero faster than t^γ . Similarly for the β_i and ζ_i . We refer to the roots that go to zero slower than t^γ as “slow roots”, and the roots that go to zero faster as “fast roots”.

Lemma 14 *If \mathbb{H} is not constant, then there are no fast roots; that is, no a_i or b_i or s_i goes to zero faster than t^γ .*

Proof. Recall that Q , S , and R , are the numbers of a_i , s_i , and b_i going to zero as fast or faster than t^γ . But now we need notation for the numbers of fast roots. Let Q_f , S_f , and R_f be respectively the number of fast a_i , fast s_i , and fast b_i . The total number of roots going to zero faster than t^γ , is

$$P := Q_f + R_f + S_f.$$

Let $\hat{\alpha}$ be the product of the nonzero α_i (all of which are real) over the a_i that go to zero as t^γ . Let $\hat{\beta}$ be the product of the nonzero β_i . If the β_i are not real, they come in complex conjugate pairs, by Lemma 9, so the product $\hat{\beta}$ is real. Let $\hat{\zeta}$ be the product of the nonzero ζ_i , which is real since all the ζ_i are real. (Do not confuse $\hat{\alpha}$ with α , which is the product of the α_i over the slow roots.) We will analyze the asymptotic form of \mathbb{H} near the origin (without yet assuming \mathbb{H} is not constant.) We will show that, if \mathbb{H} is not constant, its leading term is a nonzero real constant times βw^{P+1} .

We analyze the first formula of Lemma 10, rather than the second one. Letting t go to zero in that formula and using the definition of \mathbb{H} we have

$$\begin{aligned} \mathbb{H} = \frac{\mathbb{B}}{\mathbb{A}} & \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w \right) \\ & - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w \end{aligned} \quad (24)$$

Consider the second term in that equation, namely

$$(\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w. \quad (25)$$

Asymptotically for small w , $\mathbb{A} \mathbb{B} \mathbb{S}$ is a constant times w^P , since each fast root contributes a factor of w and there are P fast roots in all. Hence, asymptotically for small w , the second term (25) has the form

$$(\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw - \gamma \mathbb{A} \mathbb{B} \mathbb{S} w = c w^{P+1} + O(w^{P+2}) \quad (26)$$

It does not matter whether the coefficient c is zero or not (although a closer analysis shows that it is not).

Now consider the behavior of the first term in (24) near the origin, namely

$$\frac{\mathbb{B}}{\mathbb{A}} \left(\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w \right). \quad (27)$$

We have

$$\begin{aligned} \mathbb{A} &= \alpha \hat{\alpha} w^{Q_f} + O(w^{Q_f+1}) \\ \mathbb{S} &= \zeta \hat{\zeta} w^{S_f} + O(w^{S_f+1}) \\ \mathbb{A}^2 \mathbb{S} &= \alpha^2 \hat{\alpha}^2 \zeta \hat{\zeta} w^{2Q_f+S_f} + O(w^{2Q_f+S_f+1}) \end{aligned}$$

To simplify this equation we introduce a new constant:

$$c_1 = \alpha^2 \hat{\alpha}^2 \zeta \hat{\zeta}.$$

Then we have

$$\begin{aligned} \mathbb{A}^2 \mathbb{S} &= c_1 w^{2Q_f+S_f} + O(w^{2Q_f+S_f+1}) \\ \mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw &= \frac{\mathcal{M} c_1}{2Q_f + S_f + 1} w^{2Q_f+S_f+1} + O(w^{2Q_f+S_f+2}) \\ \gamma \mathbb{A}^2 \mathbb{S} w &= \gamma c_1 w^{2Q_f+S_f+1} + O(w^{2Q_f+S_f+2}) \end{aligned}$$

Subtracting the last two equations,

$$\mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - \gamma \mathbb{A}^2 \mathbb{S} w = c_1 \left(\frac{\mathcal{M}}{2Q_f + S_f + 1} - \gamma \right) w^{2Q_f+S_f+1} + O(w^{2Q_f+S_f+2})$$

We have

$$\begin{aligned} \frac{\mathbb{B}}{\mathbb{A}} &= \frac{\mathbb{B}}{\alpha \hat{\alpha} w^{Q_f} + O(w^{Q_f+1})} \\ &= \frac{\beta \hat{\beta}}{\alpha \hat{\alpha}} w^{R_f-Q_f} (1 + O(w)) \\ &= \frac{\beta \hat{\beta}}{\alpha \hat{\alpha}} w^{R_f-Q_f} + O(w^{R_f-Q_f+1}) \end{aligned}$$

so the first term of (24), that is (27), comes to

$$\begin{aligned} &c_0 \left(\frac{\mathcal{M}}{2Q_f + S_f + 1} - \gamma \right) w^{R_f+Q_f+S_f+1} + O(w^{R_f+Q_f+S_f+2}) \\ &= c_0 \left(\frac{\mathcal{M}}{2Q_f + S_f + 1} - \gamma \right) w^{P+1} + O(w^{P+2}) \end{aligned} \quad (28)$$

Subtracting the contributions of the two terms as given in (28) and (26), we have

$$\begin{aligned}\mathbb{H} &= w^{P+1} \left(c_0 \left(\frac{\mathcal{M}}{2Q_f + S_f + 1} - \gamma \right) - c \right) + O(w^{P+2}) \\ &= O(w^{P+1})\end{aligned}$$

It does not matter whether the leading term is zero or not. \mathbb{H} is a rational function of w , whose poles are at the α_i , by definition all different from zero. If \mathbb{H} is not identically zero, then it has a leading term with power at least $P + 1$; but since $\text{Im } \mathbb{H}$ takes only one sign in the upper half plane, we must then have $P = 0$. But $P = 0$ means there are no fast roots. That completes the proof of the lemma.

Remark. A similar analysis near each α_i yields the following lemma, which we do not prove here because it turns out not to be needed. Nevertheless, stating it may assist the reader to understand the situation.

Lemma 15 *If H is not constant, and the N -condition holds, then at each α_i , either \mathbb{B}/\mathbb{A} has a simple zero or a simple pole, or is analytic non-vanishing at α_i .*

Remark. This means that the number of \mathbf{b}_j converging to α_i and the number of \mathbf{a}_i converging to α_j differ by at most 1.

Proof. Omitted, because the lemma is not needed.

9 What is true if \mathbb{H} is constant

In this section we draw out some consequences of the assumption that \mathbb{H} is constant. We start with a formula for the derivative of \mathbb{H} :

Lemma 16

$$\frac{d}{dw} \mathbb{H} = \frac{d}{dw} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) \int_0^w \mathbb{A}^2 \mathbb{S} dw - \mathcal{K} \mathbb{A} \mathbb{B} \mathbb{S}$$

Proof. By the definition in (22) we have

$$\mathbb{H}(w) = \frac{\mathbb{B}}{\mathbb{A}} \mathcal{M} \int_0^w \mathbb{A}^2 \mathbb{S} dw - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} \mathbb{S} dw$$

Differentiating we have

$$\frac{d}{dw} \mathbb{H} = \frac{d}{dw} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) \int_0^w \mathbb{A}^2 \mathbb{S} dw + \left(\frac{\mathbb{B}}{\mathbb{A}} \right) \mathbb{A}^2 \mathbb{S} - (\mathcal{M} + \mathcal{K}) \mathbb{A} \mathbb{B} \mathbb{S}$$

Simplifying, we have the formula of the lemma. That completes the proof.

Lemma 17 *If H is constant, then*

$$\frac{\mathbb{B}}{\mathbb{A}} = C \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\kappa/\mathcal{M}}$$

for some constant C , and \mathbb{B}/\mathbb{A} is a polynomial.

Proof. Let \mathbf{Q} be the rational function \mathbb{B}/\mathbb{A} . By Lemma 16, we have

$$\mathbb{H}' = 0 = \mathcal{M}\mathbf{Q}' \int \mathbb{A}^2 \mathbb{S} dw - \mathcal{K}\mathbb{A}^2 \mathbb{S} \mathbf{Q}$$

The equation has one solution

$$\mathbf{Q} = \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\kappa/\mathcal{M}}$$

Since it is a linear equation, every solution is a multiple of that one, so the formula in the lemma is proved. Whenever C occurs in the context of this integral, it will mean this constant.

Multiplying by \mathbb{A} we have

$$\mathbb{B} = C\mathbb{A} \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\kappa/\mathcal{M}}.$$

This equation shows that \mathbb{B} vanishes wherever \mathbb{A} does, and to at least as great an order. Hence the rational function \mathbb{B}/\mathbb{A} is actually a polynomial. That completes the proof of the lemma.

The converse is also true:

Lemma 18 *If for some constant C , we have*

$$\frac{\mathbb{B}}{\mathbb{A}} = C \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\kappa/\mathcal{M}}$$

then \mathbb{H} is constant.

Proof. By Lemma 16,

$$\frac{d}{dw} \mathbb{H} = \mathcal{M} \frac{d}{dw} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) \int \mathbb{A}^2 \mathbb{S} dw - \mathcal{K} \mathbb{A} \mathbb{B} \mathbb{S}$$

Substituting the expression assumed in the lemma for \mathbb{B}/\mathbb{A} we have

$$\begin{aligned} \frac{d}{dw} \mathbb{H} &= \mathcal{M} C \frac{d}{dw} \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\kappa/\mathcal{M}} \int \mathbb{A}^2 \mathbb{S} dw - \mathcal{K} \mathbb{A} \mathbb{B} \mathbb{S} \\ &= \mathcal{K} C \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\kappa/\mathcal{M}-1} \mathbb{A}^2 \mathbb{S} \int \mathbb{A}^2 \mathbb{S} dw - \mathcal{K} \mathbb{A} \mathbb{B} \mathbb{S} \end{aligned}$$

$$\begin{aligned}
&= \mathcal{K}C \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\mathcal{K}/\mathcal{M}} \mathbb{A}^2 \mathbb{S} - \mathcal{K} \mathbb{A} \mathbb{B} \mathbb{S} \\
&= \mathcal{K} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) \mathbb{A}^2 \mathbb{S} - \mathcal{K} \mathbb{A} \mathbb{B} \mathbb{S} \\
&= 0
\end{aligned}$$

Since \mathbb{H} has derivative 0, \mathbb{H} is constant. That completes the proof of the lemma.

Lemma 19 *Let R , Q , and S again be the numbers of b_i , a_i , and s_i going to zero as t^γ or faster. (In other words, the degrees of \mathbb{B} , \mathbb{A} , and \mathbb{S} respectively.) Let R_f , Q_f , and S_f be the numbers of b_i , a_i , and s_i going to zero strictly faster than t^γ (the “fast roots”). If \mathbb{H} is constant then*

- (i) $R - Q = (2Q + S + 1)\mathcal{K}/\mathcal{M}$.
- (ii) $R_f - Q_f = (2Q_f + S_f + 1)\mathcal{K}/\mathcal{M}$.
- (iii) if $\mathcal{K} > 0$ then $R_f > Q_f$.

Proof. By the previous lemma we have

$$\frac{\mathbb{B}}{\mathbb{A}} = C \left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\mathcal{K}/\mathcal{M}}$$

We will derive (i) and (ii) by looking at this formula for large w and small w , respectively. Assertion (iii) follows immediately from (ii), so it suffices to prove (i) and (ii).

Ad (i). We look at the asymptotic expansion for large w . We note that as w goes to infinity, $\mathbf{Q} = \mathbb{B}/\mathbb{A}$ must be asymptotic to $(\beta/\alpha)w^{R-Q}$, where R and Q are degrees of \mathbb{B} and \mathbb{A} respectively. The right hand side gives us

$$\begin{aligned}
\left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\mathcal{K}/\mathcal{M}} &= \left(\frac{\alpha^2 \zeta}{2Q + S + 1} (w^{2Q+S+1} + O(w^{2Q+S})) \right)^{\mathcal{K}/\mathcal{M}} \\
&= \left(\frac{\alpha^2 \zeta}{2Q + S + 1} \right)^{\mathcal{K}/\mathcal{M}} w^{(2Q+S+1)\mathcal{K}/\mathcal{M}} + O(w^{(2Q+S+1)\mathcal{K}/\mathcal{M}-1})
\end{aligned}$$

The exponents of the leading power of w must be equal; that yields the equation in (i). Comparing the coefficients also enables us to determine the constant C ; it turns out to be

$$C = \frac{\beta}{\alpha} \left(\frac{2Q + S + 1}{\alpha^2 \zeta} \right)^{\mathcal{K}/\mathcal{M}}$$

Ad (ii). Now we look at the behavior for small w . Asymptotically near the origin we have

$$\int \mathbb{A}^2 \mathbb{S} dw = \frac{\alpha^2 \zeta}{2Q_f + S_f + 1} w^{2Q_f + S_f + 1} (1 + O(w))$$

and hence

$$\left(\int \mathbb{A}^2 \mathbb{S} dw \right)^{\mathcal{K}/\mathcal{M}} = \left(\frac{\alpha^2 \zeta}{2Q_f + S_f + 1} \right)^{\mathcal{K}/\mathcal{M}} w^{(2Q_f + S_f + 1)\mathcal{K}/\mathcal{M}} (1 + O(w))$$

and hence

$$\begin{aligned} \mathbb{B} &= C \left(\frac{\alpha^2 \zeta}{2Q_f + S_f + 1} \right)^{\mathcal{K}/\mathcal{M}} \mathbb{A} w^{(2Q_f + S_f + 1)\mathcal{K}/\mathcal{M}} (1 + O(w)) \\ &= c w^{(2Q_f + S_f + 1)\mathcal{K}/\mathcal{M} + Q_f} (1 + O(w)) \quad \text{for some constant } c \neq 0 \\ \beta w^{R_f} &= c w^{(2Q_f + S_f + 1)\mathcal{K}/\mathcal{M} + Q_f} (1 + O(w)) \end{aligned}$$

Equating the exponents of the leading terms on left and right we have

$$R_f = (2Q_f + S_f + 1)\mathcal{K}/\mathcal{M} + Q_f$$

as required in (ii). That completes the proof of the lemma.

10 Descending through the rings of roots

Recall that the roots a_i , b_i , and s_i fall into “rings”, where the n -th ring of roots goes to zero as t^{γ_n} , and $\gamma_1 > \gamma_2 \dots$. For each ring we can consider the corresponding w -plane, where $z = t^{\gamma_i} w$. In each ring there is a different \mathbb{A} , \mathbb{B} , \mathbb{S} , and \mathbb{H} . Up until now, we have considered only one ring of roots at a time. Now, we must consider more than one ring at a time.

The results of this section are (i) the branch points s_i do not occur, and (ii) \mathbb{H} is constant on all the rings but the innermost, and (iii) \mathbb{H} is not constant on the innermost ring, and (iv) there is at least one α_i on that innermost ring.

As above, we abbreviate by the phrase “ N -condition” the proposition that N goes to $(0, 0, -1)$ on compact subsets of the w -plane apart from the zeros of \mathbb{A} . The N -condition depends on n , since the power of t used to define the w -plane for the n -th ring of roots is γ_n . If the n -th ring satisfies the N -condition, and \mathbb{H} for ring n is constant, then by Lemma 19, there are more fast b_i ’s than there are fast a_i ’s, so the $n + 1$ -st ring also will satisfy the N -condition. This is the fundamental observation behind the following lemma:

Lemma 20 *Suppose that the N -condition holds on all rings from the first up to and including the n -th ring, and that \mathbb{H} is constant on all those rings. Then the N -condition holds on ring $n + 1$.*

Proof. The essential point is that on each of the first n rings, there are more b_i roots than a_i roots, as we shall see.

The N -condition is equivalent to $\tilde{\mathbb{B}}/\tilde{\mathbb{A}} = O(t)$ on compact subsets of the w -plane away from the α_i , and also (by Lemma 8) to $\mathcal{K} > 0$. Fix one of the rings, say the one given by γ_j , and let $\gamma = \gamma_j$. Recall that when we set $z = t^\gamma w$,

and convert $z - b_i$ to the w -plane, we get $t^\gamma(w - \beta_i)$ if b_i is on the j -th ring, or if b_i goes to zero as t^{γ_i} with $\gamma_i < \gamma$, we get t_i^γ times a constant, or if b_i goes to zero faster, we just get $t^\gamma w$. The power of contributed by the slow roots is t^{γ_i} (whatever γ is), so if there are more b_i than a_i on every ring slower than γ , then when B/A is expressed in the w -plane, we get a positive power of t from the slow roots. Each root going to zero as t^γ or faster contributes a power of t^γ , so if the degree of \mathbb{B} is more than the degree of \mathbb{A} , we will get a positive power of t from those roots too.

Now suppose that for all rings up to and including the j -th one, there are more b_i roots than a_i roots on that ring. To make the argument precise requires numerous subscripts, but it seems unavoidable:

Let R_n , Q_n , and S_n be the number of roots b_i , a_i and s_i that go to zero as or faster than γ_n ; in other words the degrees of \mathbb{B} , \mathbb{A} , and \mathbb{S} respectively. Let e_n be the sum of $n_i \gamma_i$ for $i < n$, where n_i is the number of b_i that go to zero as t_i^γ ; and let d_n be the sum of $n_i \gamma_i$ for $i < n$, where n_i is the number of a_i that go to zero as t_i^γ . Then in the n -th ring we have

$$\begin{aligned} \frac{B}{A} &= t^{\mathcal{K}} \frac{\mathbb{B}}{\mathbb{A}} + O(t^{\mathcal{K}+1}) \\ &= t^{e_n - d_n + \gamma(R_n - Q_n)} \frac{\mathbb{B}}{\mathbb{A}} + O(t^{\mathcal{K}+1}) \end{aligned}$$

Since the N -condition holds for ring n , we have

$$e_n - d_n + \gamma(R_n - Q_n) > 0.$$

When we pass to ring $n + 1$, the roots that go to zero slower than t_{n+1}^γ include those that contributed $t^{e_n - d_n}$ at ring n , but also those that go to zero as t^{γ_n} . Those now contribute to $t^{e_{n+1} - d_{n+1}}$ the powers of t that they were contributing to B/A , making no net change in the power of t in B/A . The remaining roots, the roots that go to zero faster than t^{γ_n} , now contribute a factor $t^{\gamma_{n+1}}$ instead of the former t^{γ_n} .

Therefore the N -condition will hold at ring $n + 1$ if there are more of the b_i going to zero faster than t^{γ_n} than there are a_i going to zero faster than t^{γ_n} . But that has been proved in Lemma 19, part (ii), which is applicable since \mathbb{H} is constant on ring n , and the N -condition is equivalent to $\mathcal{K} > 0$. That completes the proof of the lemma. (Actually $\mathcal{K} \geq 0$ in Lemma 19, part (ii), would have been enough.)

Lemma 21 *On the “innermost” ring of roots, i.e. the roots that go to zero as the highest power of t , \mathbb{H} is not constant, and the N -condition holds (i.e. $\mathcal{K} > 0$) on that ring, and indeed on all the rings. On all but the innermost ring \mathbb{H} is constant. On all the rings we have*

$$\frac{\mathcal{K}}{\mathcal{M}} = \frac{k}{2m + 1}$$

Proof. There are only finitely many rings, because there are only finitely many of the a_i , b_i , and s_i . Consider the first one on which \mathbb{H} is not constant, if any. By Lemma 20, each ring of roots down to and including that ring satisfies the N -condition, i.e. $\mathcal{K} > 0$. On a ring where \mathbb{H} is constant and the N -condition is satisfied, there are some fast roots, by Lemma 19, part (iii). Assume, for proof by contradiction, that \mathbb{H} is constant on all the rings. Then the N -condition holds on the innermost ring. But by Lemma 19, part (iii) there are some fast roots on that ring, which is impossible for the innermost ring, as the fast roots would lie on another ring of roots. That contradiction shows that \mathbb{H} cannot be constant on all the rings, so there is a ring on which \mathbb{H} is not constant. By Lemma 20, the N -condition holds on the slowest (outermost) ring for which \mathbb{H} is not constant. Then by Lemma 14, there are no fast roots on that ring. Hence, it is the innermost ring. It now remains only to prove the last equation of the lemma.

On the outermost ring we have $\mathcal{K} = k$ and $\mathcal{M} = 2m + 1$, so $\mathcal{K}/\mathcal{M} = k/(2m + 1)$ on the outermost ring. Now look at the formulas in Lemma 19. These involve the numbers Q , R , and S (numbers of roots that go to zero as t^γ or faster) and Q_f , R_f , and S_f (numbers of roots that go to zero faster than t^γ). When we pass from a ring on which \mathbb{H} is constant to the next ring, the Q_f , R_f , and S_f of the first ring become the Q , R , and S on the next ring in. According to Lemma 19, we have \mathcal{K}/\mathcal{M} on the first ring equal to $(R_f - Q_f)/(2Q_f + S_f + 1)$. Now passing to the second ring, the values on the right become $(R - Q)/(2Q + S + 1)$, which, according to the other formula in Lemma 19, is equal to \mathcal{K}/\mathcal{M} on the next ring. Hence, the value of \mathcal{K}/\mathcal{M} never changes as we pass from one ring to the next, all the way down to the innermost ring. Hence, it retains the value $k/(2m + 1)$ that it has on the outermost ring. That completes the proof of the lemma.

Lemma 22 *The s_i do not occur.*

Proof. According to the Gauss-Bonnet-Sasaki-Nitsche formula, a total of $2\pi m$ in extra Gaussian area must be contributed by the a_i , since the branch point has order $2m$, and each boundary branch point contributes an amount equal to a hemisphere of Gaussian area, times half the order. If any s_i exist, there are fewer than $2m$ of the a_i . We shall show that each a_i can contribute at most a hemisphere, so the total will be insufficient unless there are actually m of the a_i , in which case, no s_i exist. By Lemma 20, on the innermost ring \mathbb{H} is not constant, and the N -condition holds on each ring. By Lemma 11, all the a_i that belong to that ring or to slower rings contribute only one hemisphere of Gaussian area per a_i . Thus every a_i contributes only one hemisphere. That completes the proof of the lemma.

Remark. Tony Tromba asked whether the proof (of our main theorem about finiteness) requires an appeal to external references to know that the branch point is not a false branch point. It does not, and the key reason why not is the application of the Gauss-Bonnet formula. The surfaces u^t for $t > 0$ must contribute Gaussian area. Therefore \mathbb{H} is not constant, as we have shown. That

means that the surfaces u^t are not simply “sliding” over the same geometrical surface as t becomes positive; there is a normal component, with *some* power of t , to the motion.

Corollary 6 *We have*

$$\mathbb{H} = \frac{\mathbb{B}}{\mathbb{A}} \left(\mathcal{M} \int_0^w \mathbb{A}^2 dw \right) - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} dw$$

and $\text{Im } \mathbb{H}$ cannot take two different signs in the upper half plane.

Proof. By the previous lemma we can now put $\mathbb{S} = 1$ in the formula 22. Since $\text{Im } \mathbb{H}$ is (according to Lemma 10) the limit of a power of t times the first eigenfunction, and the first eigenfunction has just one sign in the upper half plane, either $\text{Im } \mathbb{H}$ takes just one sign or possibly is identically zero. That completes the proof.

Lemma 23 *Suppose that on the outermost ring of roots, \mathbb{H} is constant. Then on the innermost ring, we have $R - Q = (2Q + 1)\mathcal{K}/\mathcal{M}$, where Q and R are respectively the degrees of \mathbb{A} and \mathbb{B} ; that is, the number of a_i and b_i belonging to the innermost ring.*

Proof. If on the outermost ring \mathbb{H} is constant, then the innermost ring (where \mathbb{H} is not constant) is not the outermost ring. Consider the next-to-innermost ring. On that ring, the “fast roots” are the roots that belong to the innermost ring. We apply Lemma 19 on the next-to-innermost ring, where \mathbb{H} is constant. Formula (ii) of that lemma tells us that

$$R_f - Q_f = (2Q_f + S_f + 1) \frac{\mathcal{K}}{\mathcal{M}}$$

where the numbers Q_f , R_f , and S_f are the numbers of a_i , b_i , and s_i that go to zero faster than the roots of the next-to-innermost ring. But those are exactly the roots of the innermost ring. By Lemma 22, we have $S_f = 0$. Hence

$$R - Q = (2Q + 1) \frac{\mathcal{K}}{\mathcal{M}}$$

That completes the proof of the lemma.

Lemma 24 *On the innermost ring of roots, there exists some a_i ; that is, some a_i goes to zero as t^γ , where γ is the largest number such that some a_i , b_i , or s_i goes to zero as t^γ .*

Proof. On the innermost ring, \mathbb{H} is not constant, and $K > 0$, as shown in Lemma 21. Suppose, for proof by contradiction, that there is no a_i on this ring. Then \mathbb{A} is constant, $\mathbb{A} = \alpha$. By Corollary 6, we have

$$\mathbb{H}(w) = \frac{\mathbb{B}}{\mathbb{A}} \mathcal{M} \int_0^w \mathbb{A}^2 dw - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} dw$$

Substituting $\mathbb{A} = \alpha$ we have

$$\mathbb{H}(w) = \mathbb{B}\alpha\mathcal{M}w - \alpha(\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{B} dw$$

Looking at the asymptotic behavior for large w , we have $\mathbb{B} = \beta w^R + O(w^{R-1})$, where R is the degree of \mathbb{B} , and β is a nonzero constant. Then

$$\begin{aligned} \mathbb{H}(w) &= w^R \alpha \beta \mathcal{M} w - \alpha \beta (\mathcal{M} + \mathcal{K}) \frac{w^{R+1}}{R+1} + O(w^R) \\ &= \alpha \beta w^{R+1} \left(\mathcal{M} - \frac{\mathcal{M} + \mathcal{K}}{R+1} \right) + O(w^R) \end{aligned}$$

We have $R > 0$ since there must be *some* root on the innermost ring, and we have supposed there are no a_i and proved in Lemma 22 that there are no s_i . Hence, unless the coefficient is zero, \mathbb{H} will have a leading term in a power at least 2, and hence the eigenfunction, which is a multiple of $\text{Im } \mathbb{H}$, will not have one sign in the upper half plane. But it must have one sign. Hence the coefficient is zero:

$$\mathcal{M} = \frac{\mathcal{M} + \mathcal{K}}{R+1}$$

Clearing the denominator we have

$$\mathcal{M}(R+1) = \mathcal{M} + \mathcal{K}$$

Subtracting \mathcal{M} from both sides we have

$$\mathcal{M}R = \mathcal{K} \tag{29}$$

Now we recall the definitions of \mathcal{M} and \mathcal{K} from (12) and 10. Since there is no a_i on this ring, we have $Q = 0$ in those definitions; also since there no s_i at all, we have $S = 0$ and $f = 0$. The equations for \mathcal{K} and \mathcal{M} become

$$\begin{aligned} \mathcal{K} &= R\gamma + (e - d) \\ \mathcal{M} &= 2d + \gamma \end{aligned}$$

Putting those expressions into 29 we get

$$(2d + \gamma)R = R\gamma + e - d$$

Subtracting $R\gamma$ from both sides we have

$$\begin{aligned} 2dR &= e - d \\ R &= \frac{e - d}{2d} \end{aligned} \tag{30}$$

Now recall what e and d are. Let γ be the exponent associated with the innermost ring, i.e., the largest number such that some root goes to zero as t^γ . The number d is, according to its definition in (7), the product of one factor $t^{\gamma i}$

for each a_i that goes to zero on a slower ring, i.e., as t^{γ_j} for $\gamma_j < \gamma$. Since we are working on the innermost ring and assuming that there are no a_i on that ring, that is one factor of t^{γ_j} for *every* a_i . Similarly, according to (8), e is a product of one factor t^{γ_j} for each root b_i on any but the innermost ring. Now, on all the rings but the innermost, \mathbb{H} is constant, so on those rings, each a_i is matched by a b_i converging to the same α_i . That means that for each factor t^{γ_j} in the product defining d , the same factor occurs in the product defining e , and with at least as great a multiplicity, since when \mathbb{H} is constant, \mathbb{B} is a polynomial multiple of \mathbb{A} . Since all factors are powers of t , the more factors, the smaller the number. Hence $e \leq d$. Therefore $e - d \leq 0$ and since $d > 0$, (30) gives us $R \leq 0$. But as remarked above, $R > 0$, so this is a contradiction. That completes the proof of the lemma.

11 Finiteness

Now we know that on the innermost ring of roots, \mathbb{H} is not constant. The rest of the proof will focus on that ring of roots only. Henceforth, γ will be the number such that the roots on that ring go to zero as t^γ . We have already established the following facts:

- The possible branch points s_i in the lower half plane do not occur, so $\mathbb{S} = 1$.
- \mathbb{A} and \mathbb{B} do not vanish at the origin (i.e., there are no fast roots).
- The degree R of \mathbb{B} and the degree Q of \mathbb{A} are related by

$$R - Q = \frac{\mathcal{K}}{\mathcal{M}}(2Q + 1).$$

- $Q > 0$, i.e. there is at least one α_i .

We define

$$\sigma := \int_0^w \mathbb{A}^2 dw \tag{31}$$

Then σ is a polynomial of degree $2Q + 1$. As an example of the use of σ , the formula for \mathbb{H} in Corollary 6 becomes

$$\mathbb{H}(w) = \frac{\mathbb{B}}{\mathbb{A}} \mathcal{M} \sigma - (\mathcal{M} + \mathcal{K}) \int_0^w \mathbb{A} \mathbb{B} dw \tag{32}$$

To assist the intuition: σ is, except for a factor of a power of t , the first term in the arc length along the boundary Γ from the origin.

A second example of the use of σ : we can express the condition shown in in Lemma 17 and Lemma 18 to be equivalent to “ \mathbb{H} is constant” as

$$\frac{\mathbb{B}}{\mathbb{A}} = c \sigma^{\mathcal{K}/\mathcal{M}} \tag{33}$$

We have dropped the factor \mathbb{S} occurring in the cited lemmas, because it is now known to be 1.

11.1 A differential equation for \mathbb{B}/\mathbb{A} and its solution

Starting with the formula for \mathbb{H} , we will differentiate, change variables, and integrate. The result is a formula that connects the (stereographic projection of the) normal \mathbb{B}/\mathbb{A} with the derivative of the eigenfunction. The formula involves a complex path integral. The basic idea of this final section of the paper is to analyze the behavior of the terms in this formula for \mathbb{B}/\mathbb{A} , especially the complex integral, at zero, at the complex zeroes of σ , and finally at infinity. Using the facts that \mathbb{H} must have just one sign in the upper half plane, and is not constant, we are able to reach a contradiction. Here is the key formula on which these results are based:

Lemma 25 *Where both sides are defined we have for some constant c*

$$\frac{\mathbb{B}}{\mathbb{A}} = c\sigma^{\mathcal{K}/\mathcal{M}} \exp\left(\int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} dw\right)$$

The path of integration begins somewhere on the positive real axis, at a point greater than all the α_i and β_i . For real w we assume the path of integration is on the real axis except for small semicircles in the upper half plane around the poles of \mathbb{B}/\mathbb{A} .

Remarks. If \mathbb{H} is constant, then $\mathbb{H}_w = 0$, so the exponential factor is 1, and the formula simplifies to the formula given in (33) for \mathbb{B}/\mathbb{A} when \mathbb{H} is constant. In some sense, this new formula measures the deviation from what is true when \mathbb{H} is constant. The lemma mentions the zeroes of σ ; of course 0 is the only real zero of σ , but it may have complex zeroes. Even for real w , we must use a complex path of integration, that deviates from the real axis to make small detours in the upper half-plane around the poles of \mathbb{B}/\mathbb{A} . Then the constant c depends only on the starting point of the path of integration.

Proof. By Lemma 16 (with $\mathbb{S} = 1$) we have

$$\mathbb{H}_w = \mathcal{M} \int_0^w \mathbb{A}^2 dw \frac{d}{dw} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) - \mathcal{K} \mathbb{A} \mathbb{B} \quad (34)$$

Recall that $\sigma = \int_0^w \mathbb{A}^2 dw$, so $\sigma_w = \mathbb{A}^2$. Introducing σ , the previous equation becomes

$$\mathbb{H}_w = \mathcal{M} \sigma \frac{d}{dw} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) - \mathcal{K} \sigma_w \frac{\mathbb{B}}{\mathbb{A}}$$

This can be considered as a differential equation satisfied by $f = (\mathbb{B}/\mathbb{A})$:

$$\mathbb{H}_w = \mathcal{M} \sigma f_w - \mathcal{K} \sigma_w f$$

with w as the independent variable. We now want to regard σ as the independent variable. By definition $\sigma = \int_0^w \mathbb{A}^2 dw$. Hence $\sigma_w = \mathbb{A}^2$. Then σ is a one-one function of w at least locally, away from the α_i (which are zeroes of \mathbb{A}), so

$f(w)$ can be regarded (at least locally) as a function of σ . We write f_σ for the derivative of this function. By the chain rule $f_w = f_\sigma \sigma_w = \mathbb{A}^2 f_\sigma$. Then the previous equation becomes

$$\begin{aligned} \mathcal{M}\sigma f_\sigma \sigma_w &= \mathcal{K}\sigma_w f + \mathbb{H}_w \\ \mathcal{M}\mathbb{A}^2 \sigma f_\sigma &= \mathcal{K}\mathbb{A}^2 f + \mathbb{H}_w \quad \text{since } \sigma_w = \mathbb{A}^2 \end{aligned}$$

Dividing both sides by $\mathcal{M}\mathbb{A}^2 \sigma f$ we have

$$\frac{f_\sigma}{f} = \frac{\mathcal{K}}{\mathcal{M}\sigma} + \frac{\mathbb{H}_w}{\mathcal{M}\mathbb{A}^2 \sigma f}$$

Replacing f by \mathbb{B}/\mathbb{A} in the last term we have

$$\frac{f_\sigma}{f} = \frac{\mathcal{K}}{\mathcal{M}\sigma} + \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{A}\mathbb{B}}$$

We are going to integrate; this will be a complex path integral and we want to specify the path. Fix a starting point q on the positive real w -axis away from 0 and the α_i . (We take $q > 0$ so that $\sigma(q)$ will be positive.) Let P be any path from q to w , avoiding 0, any complex zeroes of σ , and the zeroes and poles of \mathbb{B}/\mathbb{A} ; to avoid those zeroes and poles it suffices to avoid the α_i and β_i . By $P\sigma$ we mean the path in the σ -plane defined by composing P with σ . We integrate with respect to σ along the path $P\sigma$ (introducing $\log c$ as a constant of integration):

$$\log f = \frac{\mathcal{K}}{\mathcal{M}} \log \sigma + \int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{A}\mathbb{B}} d\sigma + \log c$$

where the constant c is explicitly given by

$$\log c = \log f(q) - \log \sigma(q) = \log \frac{\mathbb{B}}{\mathbb{A}\sigma} \quad \text{evaluated at } w = q$$

Since \mathbb{B}/\mathbb{A} is finite and non-zero at q , and σ is non-zero at q , the constant $\log c$ is well-defined; in particular $c \neq 0$. (If $f(q) < 0$ then both $\log c$ and $\log f$ have an imaginary part $i\pi$.) We have

$$\begin{aligned} \log f &= \log \left(c\sigma^{\mathcal{K}/\mathcal{M}} \right) + \int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{A}\mathbb{B}} d\sigma \\ &= \log \left(c\sigma^{\mathcal{K}/\mathcal{M}} \right) + \int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{A}\mathbb{B}} \mathbb{A}^2 dw \quad \text{since } d\sigma = \mathbb{A}^2 dw \\ &= \log \left(c\sigma^{\mathcal{K}/\mathcal{M}} \right) + \int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} dw \end{aligned}$$

Here the path of integration must avoid the singularities of the integrand, which occur at $\sigma = 0$ (the only real zero of σ is $w = 0$, but there may be complex zeroes), at zeroes of \mathbb{B}/\mathbb{A} , and possibly at singularities of \mathbb{H}_w . The branch of the logarithm is chosen to be continuous along the path of integration.

Applying the function $\exp(x) = e^x$ to both sides of the equation, we have

$$f = c\sigma^{\mathcal{K}/\mathcal{M}} \exp\left(\int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} dw\right)$$

Recalling that $f = \mathbb{B}/\mathbb{A}$ we have proved the lemma.

11.2 \mathcal{K}/\mathcal{M} is an integer

This is an important step in the proof. Among other things, it implies that $\sigma^{\mathcal{K}/\mathcal{M}}$ is a polynomial. We prove it by analyzing the behavior of the path integral in Lemma 25 at the origin.

Lemma 26 \mathcal{K} is a multiple of $\mathcal{M} = 2m + 1$; in other words \mathcal{K}/\mathcal{M} is an integer.

Proof. First we calculate \mathbb{H}_w at the origin. We have by (34)

$$\mathbb{H}_w = \mathcal{M}\sigma \frac{d}{dw} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) - \mathcal{K}\mathbb{A}\mathbb{B}$$

At the origin, σ is zero. Let α and β be the constants such that $\mathbb{A} = \alpha \prod (w - \alpha_i)$ and $\mathbb{B} = \beta \prod (w - \beta_i)$. Let η be the product of all the $-\alpha_i$, and δ the product of all the $-\beta_i$. Then the value of $\mathbb{A}\mathbb{B}$ at the origin is $\alpha\beta\eta\delta$. Hence the value of \mathbb{H}_w at the origin is $-\mathcal{K}\alpha\beta\eta\delta \neq 0$.

By Lemma 25,

$$\frac{\mathbb{B}}{\mathbb{A}} = c\sigma^{\mathcal{K}/\mathcal{M}} \exp\left(\int \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} dw\right)$$

As w moves from small positive values to small negative ones, the left side remains real, and since neither \mathbb{B} nor \mathbb{A} is zero at the origin, the left side does not change sign. Since $\mathbb{H}_w \neq 0$ at the origin, the integrand has a simple pole, due to the simple zero of $\sigma = \int_0^w \mathbb{A}^2 dw$, as we now show in detail. Near the origin, $\sigma = \int_0^w \mathbb{A}^2 dw = \alpha^2\eta^2w + O(w^2)$. Hence the denominator is

$$\begin{aligned} \mathcal{M}\sigma\mathbb{B}/\mathbb{A} &= \mathcal{M}(\alpha^2\eta^2w) \frac{\beta\delta}{\alpha\eta} + O(w^2) \\ &= \mathcal{M}\alpha\eta\beta\delta w + O(w^2) \end{aligned}$$

Since the value of the numerator at the origin is $-\mathcal{K}\alpha\beta\eta\delta$, the residue is $-\mathcal{K}/\mathcal{M}$. In order that the right side remain real when w passes from small negative to small positive values, this residue must be an integer. That completes the proof.

Remark. The factor $\sigma^{\mathcal{K}/\mathcal{M}}$ changes sign if and only if k is odd, because $\mathcal{K}/\mathcal{M} = k/(2m + 1)$. Hence the exponential factor changes sign if and only if k is odd. For the exponential factor to change sign, the complex path integral in the exponent must pick up an odd multiple of $i\pi$ on a small semicircle in the upper half plane around the origin. That also happens if and only if \mathcal{K}/\mathcal{M} is odd, so in either case, the right side does not change signs.

11.3 The non-real zeroes of σ are also zeroes of \mathbb{B}

In this section, we analyze the behavior of the formula in Lemma 25 at the non-real zeroes of σ . The conclusion is that these zeroes are also zeroes of \mathbb{B} , and moreover, they have the same multiplicity as zeroes of \mathbb{B} as they do as zeroes of $\sigma^{\mathcal{K}/\mathcal{M}}$ (which is a polynomial because \mathcal{K}/\mathcal{M} is an integer).

We start with the following fact from complex analysis:

Lemma 27 *Let $f(w)$ be a rational function with a pole at q . Let $\gamma : [0, 1] \rightarrow \mathbf{C}$ be a path in the complex plane avoiding the poles of w , except that at the last point of γ , γ reaches q , i.e. $\gamma(1) = q$. Let $p = \gamma(0)$. Then*

$$\lim_{\tau \rightarrow 1} \int_p^{\gamma(\tau)} f(w) dw$$

does not have a value independent of γ ; indeed, for appropriate choices of γ it can be made $-\infty$, ∞ , or 0 .

Proof. By considering the new rational function $f(w - q)$, we can assume without loss of generality that the pole q is at $w = 0$ and the function f has the asymptotic form

$$f(w) = Cw^{-j} + O(w^{-j+1}) = Cw^{-j}(1 + O(w))$$

for some constant $C \neq 0$. Also without loss of generality, we can assume that C is real. We seek a change of variable $w \mapsto \zeta$ such that as a function of ζ we have

$$f(w) = C\zeta^{-j} \quad \text{exactly}$$

Once we have such a variable ζ , the desired paths of integration in the ζ plane are simply lines given by $g(\tau) = (1 - \tau)e^{i\theta}$ for various θ . Along such a path we have

$$\begin{aligned} \int f(w) dw &= \int_0^{g(1-\tau)} \zeta^{-j} d\zeta \\ &= \int_0^{1-\tau} g(t)^{-j} g'(t) dt \\ &= \int_0^{1-\tau} (1-t)^{-j} e^{-ij\theta} g'(t) dt \\ &= e^{-ij\theta} \int_0^{1-\tau} (1-t)^{-j} g'(t) dt \\ &= e^{-ij\theta} \int_0^{1-\tau} (1-t)^{-j} (-e^{i\theta}) dt \\ &= -e^{-(j+1)i\theta} \int_0^{1-\tau} (1-t)^{-j} dt \\ &= e^{-(j+1)i\theta} \int_1^\tau x^{-j} dx \end{aligned}$$

Since the pole at q is at least a double pole, we have $j \neq 1$, and

$$\int f(w) dw = e^{(j-1)i\theta} \left(1 - \frac{\tau^{-j+1}}{-j+1} \right)$$

and choosing θ appropriately we can make the right side approach $\pm\infty$ or 0 as desired. If $j = 1$, we have

$$\int f(w) dw = e^{-2i\theta} \log \tau$$

which also can be made to approach $\pm\infty$ or 0 by an appropriate choice of θ .

It remains to show how to find ζ as a function of w . We have

$$f(w) = Cw^j(1 + g(w))$$

for some analytic function $g(w)$ with $g(w) = 0$. Then by the binomial series, $(1 + g(w))^{1/j}$ is analytic in some neighborhood of the origin. That is, there is a function h analytic in some neighborhood of the origin such that

$$(h(w))^j = 1 + g(w)$$

Now define $\zeta = wh(w)$. Then $\zeta^j = w^j(h(w))^j = w^j(1 + g(w)) = f(w)$, so we have found the required change of variables. That completes the proof of the lemma.

Next we extract information by looking at the formula of Lemma 25 at the non-real zeroes of σ . Those zeroes occur in complex-conjugate pairs, but our argument does not depend on which half-plane contains the zero in question.

Lemma 28 *Every non-real zero of σ is also a zero of \mathbb{B} , and as a zero of \mathbb{B} it has multiplicity \mathcal{K}/\mathcal{M} times greater than its multiplicity as a zero of σ .*

Proof. By (34) we have

$$\mathbb{H}_w = \mathcal{M}\sigma \frac{d}{dw} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) - \mathcal{K}\mathbb{A}\mathbb{B}.$$

Suppose that q is a non-real zero of σ . Since the zeroes of \mathbb{A} are all real, \mathbb{B}/\mathbb{A} is analytic at q . Then at q we have

$$\mathbb{H}_w = -\mathcal{K}\mathbb{A}\mathbb{B}$$

Therefore

$$\frac{\mathbb{H}_w}{\mathbb{B}/\mathbb{A}} = -\mathcal{K}\mathbb{A}^2 \neq 0.$$

Therefore

$$\frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} \quad \text{has a pole at } q$$

By Lemma 25, we have

$$\frac{\mathbb{B}}{\mathbb{A}} = c\sigma^{\mathcal{K}/\mathcal{M}} \exp\left(\int \frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}}\right).$$

Now assume, for proof by contradiction, that \mathbb{B} is not zero at q . Then the left side, \mathbb{B}/\mathbb{A} , is not zero at q and is continuous there, since q is not real but the zeroes of \mathbb{A} are real. On the right side, $\sigma^{k/\mathcal{M}}$ is zero at q . Therefore the exponential factor has to tend to infinity as w approaches q . That means that the integral also tends to infinity as w approaches q . But by Lemma 27, the integral does not have a limit as w approaches q ; choosing an appropriate path approaching q the real part of the (absolute value of the) integrand will remain bounded. Hence the exponential factor remains bounded away from both zero and infinity. Hence the right side of the equation goes to zero. That is a contradiction, since the left side approaches a nonzero value. Hence in fact \mathbb{B} is zero at q .

It remains to prove that q has the same multiplicity as a zero of \mathbf{B} as it does as a zero of $\sigma^{\mathcal{K}/\mathcal{M}}$. Suppose, for proof by contradiction, that it has a different multiplicity. Because the exponential factor has a non-zero finite limit (for a suitable approach to q), along that approach \mathbb{B}/\mathbb{A} and $\sigma^{\mathcal{K}/\mathcal{M}}$ must approach zero at the same rate. Since \mathbb{A} is not zero at q , it follows that q has the same multiplicity as a zero of \mathbb{B} and as a zero of $\sigma^{\mathcal{K}/\mathcal{M}}$. That completes the proof of the lemma.

11.4 Laurent expansions at infinity

Recall that R is the degree of \mathbb{B} , Q is the degree of \mathbb{A} , and

$$R - Q = \frac{\mathcal{K}}{\mathcal{M}}(2Q + 1).$$

Looking at the formula of Lemma 25, we see that the left side \mathbb{B}/\mathbb{A} and the factor $\sigma^{\mathcal{M}}$ on the right both are asymptotic to w^{R-Q} for large k . The ratio of the two therefore goes to 1 as w goes to infinity. But what is the next power in its Laurent expansion? We need to show that it can't be too very far out. Our answer is that it is no farther out than the term in $w^{-(R-Q)}$. It is important that $\sigma^{\mathcal{K}/\mathcal{M}}$ is a polynomial—otherwise we would have no hope of answering this question! Therefore it is natural to begin with a lemma about the Laurent expansions of rational functions.

Lemma 29 *Let f and g be two unequal rational functions of the same degree n with the same leading coefficient. Let $n - J$ be the highest power of w such that the coefficients of w^J in f and g are different. Then the Laurent expansion of f/g at infinity is given by*

$$\frac{f(w)}{g(w)} = 1 + \mu w^{-J} + O(w^{-J-1})$$

for some constant $\mu \neq 0$. In particular $J \leq n$.

Proof. The worst case (largest J) occurs when the numerator and denominator have their corresponding coefficients equal. Here is an example:

$$\begin{aligned}\frac{w^n + 3}{w^n + 5} &= \frac{1 + 3w^{-k}}{1 + 5w^{-k}} \\ &= (1 + 3w^{-k})(1 - 5w^{-k})(1 + O(w^{-1})) \\ &= (1 - 2w^{-k})(1 + O(w^{-1}))\end{aligned}$$

The general proof is only notationally more complicated: Let $f(w) = a_0w^n + \dots + a_n$ and $g(w) = b_0w^n + \dots + b_n$, with $a_0 = b_0$. Let J be the smallest integer such that $a_J \neq b_J$. Then with $h(w)$ defined as the sum of the terms with lower-indexed coefficients we have

$$\frac{\sum a_i w^{n-i}}{\sum b_i w^{n-i}} = \frac{h(w) + \sum_{i=J}^n a_i w^{n-i}}{h(w) + \sum_{i=J}^n b_i w^{n-i}}$$

Dividing numerator and denominator by $h(w)$ we have

$$\begin{aligned}\frac{\sum a_i w^{n-i}}{\sum b_i w^{n-i}} &= \frac{1 + (a_J/a_0)w^{-J} + O(w^{-J-1})}{1 + (b_J/b_0)w^{-J} + O(w^{-J-1})} \\ &= 1 + \frac{a_J}{a_0} - \frac{b_J}{b_0}w^J + O(w^{-J-1}) \\ &= 1 + \frac{a_J - b_J}{a_0}w^{-J} + O(w^{-J-1}) \quad \text{since } a_0 = b_0\end{aligned}$$

Since by hypothesis $a_J \neq b_J$, we have found the first nonzero term in the Laurent expansion. That completes the proof of the lemma.

Lemma 30 *For some constant $\mu \neq 0$ and some integer J with $1 \leq J \leq R - 2Q$ we have the Laurent expansion at infinity:*

$$\sigma^{-k/\mathcal{M}} \left(\frac{\mathbb{B}}{\mathbb{A}} \right) = 1 + \mu w^{-J} + O(w^{-J-1}).$$

Remark. Whether or not \mathcal{K}/\mathcal{M} is an integer, we always have a Laurent expansion. The point is that since \mathcal{K}/\mathcal{M} is an integer, we have a rational function on the left and can therefore say where the first nonzero term in the Laurent expansion must occur, while without knowing that \mathcal{K}/\mathcal{M} is an integer, the first nonzero term could occur very far out in the Laurent series.

Proof. Since \mathcal{K}/\mathcal{M} is an integer, the function $\sigma^{\mathcal{K}/\mathcal{M}}$ is a polynomial. Since the degree of σ is $2Q + 1$, the degree of $\sigma^{\mathcal{K}/\mathcal{M}}$ is $(\mathcal{K}/\mathcal{M})(2Q + 1) = R - Q$. The expression in the lemma can be written as

$$\frac{\mathbb{B}}{\mathbb{A}\sigma^{\mathcal{K}/\mathcal{M}}}.$$

This is a rational function. It is not constant, since if it were constant, then by Lemma 18, \mathbb{H} would be constant, but \mathbb{H} is not constant. On the face of it, the

numerator has degree R , and the denominator has degree $Q + (\mathcal{K}/\mathcal{M})(2Q+1) = Q + (R-Q) = R$, the same as the numerator. But appearances are deceptive: by Lemma 28, the non-real zeroes of the denominator are all zeroes of the numerator with the same multiplicity. Now σ is of degree $2Q+1$ and has only one real zero (namely zero), so there are $2Q(\mathcal{K}/\mathcal{M})$ of these zeroes, counting multiplicities. After these common zeroes are cancelled out, the degrees of numerator and denominator will be

$$n := R - \frac{2Q\mathcal{K}}{\mathcal{M}}$$

Since \mathcal{K}/\mathcal{M} is an integer, the last term being subtracted off is at least $2Q$. Hence

$$n \leq R - 2Q.$$

Applying Lemma 29 to this rational function, there exists a term w^{-J} in the Laurent expansion with $J \leq R - 2Q$. That completes the proof of the lemma.

11.5 The final contradiction

Now we are ready to analyze the behavior of the formula of Lemma 25 at infinity. The numerator of the integrand is \mathbb{H}_w . On the face of it, by (34), \mathbb{H}_w is a rational function asymptotic to w^{R-Q} . One can easily calculate that the leading coefficient vanishes, but since \mathbb{H} has only one sign in the upper half plane, we must have \mathbb{H} asymptotic to w or w^{-1} . That means that the integrand goes to zero as a large negative power of w , as those higher powers of \mathbb{H} are not there to cancel the denominator. That makes the integral (to infinity) finite; that is the first key point of the proof. Then we divide the formula of Lemma 25 by $\sigma^{k/\mathcal{M}}$, isolating the exponential factor on the right:

$$\sigma^{-\mathcal{K}/\mathcal{M}} \frac{\mathbb{B}}{\mathbb{A}} = c \exp \left(\int \frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} \right).$$

Now we have worked out the Laurent expansion of the left side, and the order of the pole in the integrand. Differentiating this equation, we arrive at a comparison between the first non-zero term of the Laurent expansion, and the order of the pole. The first term on the left is w^J (or w^{J-1} after differentiating), where J is at most so much, and the order of the pole on the right is at least so much, and it turns out that the two estimates are incompatible. The details are below.

Theorem 8 (Finiteness) *Let Γ be a real-analytic Jordan curve in \mathbb{R}^3 . Then Γ cannot bound infinitely many immersed disk-type minimal surfaces whose least eigenvalue satisfies $\lambda_{\min} \geq 2$.*

Proof. By Theorem 7, Γ bounds a one-parameter family of minimal surfaces u^t with u^t immersed for $t > 0$, and each u^t having least eigenvalue 2, so the

assumptions that have been in force for all our calculations are satisfied. By Lemma 25, we have

$$\frac{\mathbb{B}}{\mathbb{A}} = c\sigma^{\mathcal{K}/\mathcal{M}} \exp\left(\int \frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}}\right).$$

Dividing both sides by $\sigma^{\mathcal{K}/\mathcal{M}}$ we have

$$\sigma^{-\mathcal{K}/\mathcal{M}} \frac{\mathbb{B}}{\mathbb{A}} = c \exp\left(\int \frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}}\right).$$

We consider the behavior for large w . By Lemma 30, there is an integer J with $1 \leq J \leq R - 2Q$ such that the Laurent expansion of the left side has a term in w^{-J} . The leading term is β , the leading coefficient of \mathbb{B} . Then we have for some $\mu \neq 0$

$$\beta + \mu w^{-J} + O(w^{-J-1}) = c \exp\left(\int \frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}}\right). \quad (35)$$

Since $\text{Im } \mathbb{H}$ has one sign in the upper half plane, the Laurent expansion of \mathbb{H} has to begin with w or w^{-1} . Hence the Laurent expansion of H_w begins with a constant term or with w^{-2} . Therefore the first term of H_w can be written $\nu w^{-1 \pm 1}$ for some constant $\nu \neq 0$. The integrand can then be written like this:

$$\begin{aligned} \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} &= \frac{\mathbb{A}\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}} \\ &= \frac{\alpha\nu w^{Q-1 \pm 1}}{\beta w^{2Q+1} w^R} (1 + O(w^{-1})) \\ &= \frac{\alpha\nu}{\beta} w^{-1 \pm 1 - Q - R - 1} (1 + O(w^{-1})) \\ &= \frac{\alpha\nu}{\beta} w^{-(2+Q+R \pm 1)} (1 + O(w^{-1})) \end{aligned}$$

Summarizing that calculation,

$$\frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} = \frac{\alpha\nu}{\beta} w^{-(2+Q+R \pm 1)} (1 + O(w^{-1})) \quad (36)$$

Since Q and R are the degrees of \mathbb{A} and \mathbb{B} , respectively, we have $Q + R \geq 1$. Therefore the exponent is negative and has magnitude at least $Q + R + 1$, which is at least 2. It follows that the integral has a finite limit as w goes to infinity. Let that finite limit be L . Then the integral is $L + O(w^{-1})$.

Differentiating (35), which is legal since the $O(w^{-J-1})$ term stands for a Laurent series, we have

$$\begin{aligned} -J\mu w^{-J-1} + O(w^{-J-2}) &= c \exp\left(\int \frac{\mathbb{H}_w dw}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}}\right) \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} \\ &= c \exp\left(L + O(w^{-1})\right) \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} \\ &= ce^L (1 + O(w^{-1})) \frac{\mathbb{H}_w}{\mathcal{M}\sigma\mathbb{B}/\mathbb{A}} \end{aligned}$$

Putting in the result from (36) we have

$$-J\mu w^{-J-1} + O(w^{-J-2}) = ce^L(1 + O(w^{-1}))\frac{\alpha\nu}{\beta}w^{-(2+Q+R\pm 1)}(1 + O(w^{-1}))$$

On the left, we have a negative exponent $J + 1 \leq R - 2Q + 1$ (by Lemma 30). On the right, we have a negative exponent of magnitude at least $R + Q + 1$. If these two exponents were equal, we would have $R - 2Q + 1 \geq R + Q + 1$. Subtracting the right side from both sides we would have $-Q \geq 0$. Since Q is the degree of \mathbb{A} , which is non-negative, we would have $Q = 0$. But that contradicts Lemma 24, so we have reached a contradiction.

That contradiction depends on no other assumptions than the existence of a one-parameter family of minimal surfaces u^t bounded by Γ , such that $\lambda_{\min} = 2$ for $t > 0$. That completes the proof by contradiction that no such family of minimal surfaces exists.

12 Relative minima in which topology?

We are going to prove a theorem limiting the number of relative minima of area with a given boundary. To state such a theorem precisely, we must fix a topology to use in defining “relative minimum.” Once a topology is fixed, a relative minimum is a surface u such that there exists a neighborhood of u (in the chosen topology) such that no surface in that neighborhood has smaller area than u . As usual in Plateau’s problem, we work with harmonic surfaces.

We wish to use the C^n topology. First consider whether it is a stronger theorem for smaller n or larger n . *A priori* it might be possible to decrease area by a C^0 variation but not by a C^n variation, but if area can be decreased by a C^n variation, that same variation counts as a C^0 variation. Therefore a C^0 relative minimum is automatically a C^n relative minimum, but not vice-versa. Thus if there were infinitely many C^0 relative minima, there would automatically be infinitely many C^n relative minima. Taking the contrapositive, if there cannot be infinitely many C^n relative minima, there cannot be infinitely many C^0 relative minima. Thus the theorem will be stronger, when we use the C^n topology for as large an n as possible.

Unfortunately, our result depends on the regularity result that a relative minimum cannot have an interior or boundary branch point, which at present is known only for the C^0 topology.

When and if this regularity result is proved for larger n , our proof will immediately apply to the C^n topology; but for now we can prove it for the C^0 case.

Theorem 9 (Main theorem) *Let Γ be a real-analytic Jordan curve in R^3 . Suppose that n is a positive integer and every relative minimum of area in the C^n topology bounded by Γ has no (interior or boundary) branch points. Then Γ cannot bound infinitely many disk-type minimal surfaces that are relative minima of area in the C^n topology.*

Remark: As usual in Plateau’s problem, we are counting two surfaces that differ only by a conformal reparametrization as the same surface.

Proof: Suppose, for proof by contradiction, that Γ bounds infinitely many disk-type minimal surfaces furnishing C^n relative minima of area. Then those surfaces are immersed, by hypothesis, and have $\lambda_{\min} \geq 2$. Then by Theorem 7, Γ bounds a one-parameter family of minimal surfaces u^t with u^t immersed for $t > 0$, and each u^t having least eigenvalue 2. But that contradicts Theorem 8. That completes the proof.

Corollary 7 *Let Γ be a real-analytic Jordan curve in R^3 . Then Γ cannot bound infinitely many disk-type minimal surfaces that are relative minima of area in the C^0 topology.*

Proof. By Theorem 1 and the known regularity results, which are fully discussed in section 3.6 above.

13 Open Problems and Conjectures

The theorem proved here could possibly be generalized by weakening any of these hypotheses: (i) that the boundary is real-analytic; (ii) that the surfaces considered are of the topological type of the disk; or (iii) that the surfaces considered are relative minima of area. All questions formed by weakening one or more of these hypotheses are open. These questions are all “equal conclusion” strengthenings of the theorem, formed by weakening the hypotheses. We will comment on these questions one by one below.

There is also an interesting “stronger conclusion” version of the theorem, in which we ask for an explicit bound on the number of relative minima as a function of the geometry of the boundary curve Γ . Our ignorance is scandalous: Not counting curves for which uniqueness is known, there is only one Jordan curve Γ known for which anyone can give an explicit bound on the number of solutions of Plateau’s problem (of disk type) bounded by Γ . That curve is “Enneper’s wire” Γ_r (the image of the circle of radius r in the usual parametrization of Enneper’s surface), for values of r slightly larger than 1, and the bound is three. This bound follows from the theorem of Ruchert [32] that uniqueness holds up to $r = 1$ and the theorem of [8] characterizing the space of minimal surfaces bounded by Γ_r in the vicinity of Enneper’s surface for $r = 1$. The following conjecture, for example, seems presently out of reach: if the real-analytic Jordan curve Γ has total curvature less than or equal to 6π , then it bounds at most two relative minima of area (of disk type). We do know that there can be no such estimate for the number of minimal surfaces (minimizing or not), since Böhme proved in [11] that for any $\epsilon > 0$ and any n , there is a real-analytic Jordan curve with total curvature less than $4\pi + \epsilon$ bounding more than n minimal surfaces. But these are branched surfaces, not relative minima, so they do not show the impossibility of a bound on the number of immersed minimal surfaces bounded by Γ .

Since Enneper’s wire bounds two relative minima when the total curvature exceeds 4π , we could expect to string together sequences of Enneper’s wires with small bridges connecting them, obtaining for each n a Jordan curve with total curvature $4n\pi + \epsilon$ bounding (at least) 2^n relative minima (as well as some unstable minimal surfaces). Since we seem to have to increase the total boundary curvature to get many solutions of Plateau’s problem, we might hope that for some explicitly given function F , we might have, for all real-analytic Jordan curves Γ of total curvature less than $2\pi n$, there are at most $F(n)$ relative minima of area. As far as we know, this might be true with $F(n) = 2^{n-1}$. Thus: (when $n = 2$) curvature less than 4π implies uniqueness (which is known); when $n = 3$, curvature less than 6π implies at most two relative minima (the conjecture mentioned above); when $n = 4$, curvature less than 8π implies at most 4 relative minima (also, of course, not known).

Next we take up the possible “equal conclusion” versions. First, what about weakening the smoothness of the boundary to C^n ? Then we have some troubles about the boundary branch point representation, but perhaps that can be solved adequately. A more pressing difficulty is that we then cannot control the dependence of the zeroes, poles, and ramification points of the Gauss map N on the parameter t . There is also a third difficulty: even the proof of Tomi’s theorem that there can’t be infinitely many absolute minima does not go through for C^∞ boundaries, because the regularity result that a minimum of area cannot have a boundary branch point is still an open problem for C^∞ boundaries.

Second, what about generalizing the theorem to other topological types of surfaces or boundaries? Frank Morgan has given an example [27] of a boundary consisting of four circles, such that it bounds a continuum of unstable minimal surfaces of arbitrarily high genus. That still leaves open the finiteness problem for relative minima. I conjecture that a system of real analytic Jordan curves cannot bound infinitely many relative minima of any fixed orientable topological type (that is, any fixed number of boundary components and Euler characteristic). For non-orientable surfaces, there is a problem in that the least eigenfunction no longer has one sign. Instead, it has just one nodal line. Thus even Tomi’s argument that there cannot be a loop of minimal surfaces all of which are absolute minima of area fails to go through for surfaces of the type of the Möbius strip. That is the simplest open question for non-orientable surfaces: can there be a one-parameter family of absolute minima of surfaces of the type of the Möbius strip?

Incidentally, another paper of Morgan [26] gives an example of a curve in R^4 that bounds a continuum of absolute minima of disk type. Therefore we cannot generalize the theorem to higher dimensional spaces.

Third, what about dropping the requirement that the surfaces be relative minima of area? Then our methods say nothing at all, since they all depend on analysis of the least eigenvalue and its eigenfunction. The only known results (other than uniqueness theorems) along these lines are two “ 6π theorems”, both of which assume that Γ is a real-analytic Jordan curve whose total curvature is less than or equal to 6π . In [6], it is proved that such a curve cannot bound infinitely many minimal surfaces (stable or unstable). Earlier, Nitsche proved

that same conclusion (see §A29 of [29], p. 447) using the additional hypothesis that Γ bounds no minimal surface with branch points. At the time of Nitsche's proof, the forced Jacobi fields had not yet been discovered. In [6], the theory of the forced Jacobi fields is used.

Appendix: Dictionary of Notation

To assist the reader we provide in one place a summary of the notations used in more than one section of this paper. This summary is provided as a memory aid, not as a list of complete definitions. The symbols are more or less in order of introduction; there is no natural "lexicographical" order for such a variety of symbols.

X, Y, Z	coordinates in R^3
Γ	a real-analytic Jordan curve in R^3 , tangent to the X -axis at the origin
$u = u^t$	a one-parameter family of minimal surfaces, analytic in t in some interval $[0, t_0]$ u^t is analytic as a function of z in a closed upper half-disk
t	the parameter defining that one-parameter family
2u	the second component of the vector $u = ({}^1u, {}^2u, {}^3u)$.
$z = x + iy$	the parameter used to define u^t
N	the unit normal to u^t
g	the stereographic projection of the Gauss map N
u_t	subscript denotes differentiation: $\frac{\partial u}{\partial t}$
λ_{\min}	least eigenvalue
ϕ	$u_t \cdot N$. If nonzero, this is an eigenfunction for eigenvalue 2.
n and χ	$\phi = t^n \chi$.
f	${}^1u_z - {}^2u_z$, occurs in the Weierstrass representation of u^t
$2m$	order of the boundary branch point of u (when $t = 0$).
s_i	branch points in the lower half plane for $t > 0$ that converge to 0 as $t \rightarrow 0$. There are S of these.
N	There are $2N$ of the s_i . Hopefully no confusion will arise with N the unit normal.
a_i	zeroes of f . The a_i depend analytically on t . There are Q of these.
b_i	zeroes of fg^2 . The b_i depend analytically on t . There are R of these.
γ_n	The n -th ring of roots consists of a_i, b_i , and s_i that go to zero as t^{γ_n} .
γ	One of the γ_n .
α_i	coefficient such that $a_i = \alpha_i t^\gamma + O(t^{\gamma+1})$
β_i	coefficient such that $b_i = \beta_i t^\gamma + O(t^{\gamma+1})$
ζ_i	coefficient such that $s_i = \zeta_i t^\gamma + O(t^{\gamma+1})$
	α_i, β_i , and ζ_i are used only when non-zero, i.e.

	for the roots that go to zero as t^γ .
α	product of all the $-\alpha_i$ over roots going to zero slower than t^γ
β	product of all the $-\beta_i$ over roots going to zero slower than t^γ times $B_0(0,0)$
ζ	product of all the $-\zeta_i$ over roots going to zero slower than t^γ
A	$A = A_0 \prod_{i=1}^{m-N} (z - a_i)$ so $f = A^2$, and $A_0 = 1$ when $t = 0$
B	$B = \beta \prod_{i=1}^{m+k-N} (z - b_i)$ so $fg^2 = B^2$.
C	complex constant, value of B_0 when $t = 0$
S	$S = \prod_{i=1}^{2N} (z - s_i(t))$
$w = z/t^\gamma$	a “blow-up” of the parameter domain
$O(t)$	refers to uniform convergence on compact subsets of the w -plane away from the α_i
\mathbb{A}	$\mathbb{A} = \alpha \prod_{i=1}^Q (w - \alpha_i)$ so $A = t^{(m-N)\gamma} \mathbb{A}(1 + \mathbf{O}(t))$ \mathbb{A} does not depend on t
\mathbb{B}	$\mathbb{B} = \beta \prod_{i=1}^R (w - \beta_i)$ so $B = t^{(m+k-N)\gamma} \mathbb{B}(1 + \mathbf{O}(t))$ \mathbb{B} does not depend on t , and B_0 turns out to be real.
$\tilde{\mathbb{A}}$	$\alpha \prod (w - \mathbf{a}_i)$. The product is over i such that $a_i = O(t^\gamma)$
$\tilde{\mathbb{B}}$	$i\lambda B_0 \beta \prod (w - \mathbf{b}_i)$
$\tilde{\mathbb{S}}$	$\zeta \prod (w - \mathbf{s}_i)$
\mathbf{a}_i	$\mathbf{a}_i = a_i/t^\gamma$
\mathbf{b}_i	$\mathbf{b}_i = b_i/t^\gamma$
\mathbf{s}_i	$\mathbf{s}_i = s_i/t^\gamma$
\mathcal{K}	number such that $t^{\mathcal{K}} \mathbb{B}/\mathbb{A} = B/A + O(t^{\mathcal{K}+1})$
\mathcal{M}	number such that $t^{\mathcal{M}} \int \mathbb{A}^2 \mathbb{S} dw = \int A^2 S dz + O(t^{\mathcal{M}+1})$
$\tilde{\mathbb{A}}_0(t, w)$	$\tilde{\mathbb{A}}_0(t, w) = A_0(t, t^\gamma w)$
$\tilde{\mathbb{B}}_0(t, w)$	$\tilde{\mathbb{B}}_0(t, w) = B_0(t, t^\gamma w)$
χ	$\chi(w) = \phi(t^\gamma w) / \gamma t^{(2m+k+1)\gamma-1}$ χ shows what happens to the eigenfunction in the w -plane.
χ^0	limit of χ when t goes to 0
\mathbb{H}	complex analytic function such that $\chi^0 = -\text{Im } \mathbb{H}$
σ	abbreviation for $\int_0^w \tilde{\mathbb{A}}^2 dw$ in the last two sections
τ	the parameter used to parametrize the boundary curve, equal to the X -coordinate

References

- [1] Beeson, M., On interior branch points of minimal surfaces, *Math Zeitschrift* **171** (1980) 133-154.
- [2] Beeson, M., Some results on finiteness in Plateau’s problem, Part I, *Math. Zeitschrift* **175** (1980) 103-123.
- [3] Preprint version of [2], preprint no. 286, Sonderforschungsbereich 72, Universität Bonn (1977).

- [4] Beeson, M. Some results on finiteness in Plateau's problem, Part II, *Math. Zeitschrift* **181** (1982) 1-30.
- [5] Beeson, M., Notes on Minimal Surfaces, unpublished manuscript available on the author's website at
www.michaelbeeson.com/research/papers/introminimal.pdf
- [6] Beeson, M., The 6π theorem about minimal surfaces, *Pacific Journal of Mathematics* **117** No. 1, 1985.⁷
- [7] Beeson, M.,
www.MichaelBeeson.com/research/minimalsurfaces/BranchPointAnimations
- [8] Beeson, M., and Tromba, A. J., The cusp catastrophe of Thom in the bifurcation of minimal surfaces, *Manuscripta Mathematica* **46** (1984) 273-308.
- [9] Böhme, R., Die Zusammenhangskomponenten der Lösungen analytischer Plateauprobleme. *Math. Zeitschrift* **133**, 31-40 (1973).
- [10] Böhme, R., New results on the classical problem of Plateau on the existence of many solutions. Séminaire Bourbaki 24 année **579** (1981–82) 1–20.
- [11] Böhme, R., A Plateau problem with many solutions for boundary curves in a given knot class, in Feris, Künel, Simon, and Wegner (eds.), *Global Differential Geometry and Global Analysis*, Lecture Notes in Mathematics **838**, pp. 36–41. Springer, Berlin Heidelberg (1981).
- [12] Böhme, R., and Tomi, F., Zur structure der Lösungsmenge des Plateauprobblems, *Math. Zeitschrift* **133** (1973) 1–29.
- [13] Böhme, R., and Tromba, A. J., The index theorem for classical minimal surfaces, *Annals of Mathematics* **113** 2, 447-499 (1981).
- [14] Courant, R., *Dirichlet's Principle, Conformal Mapping, and Minimal Surfaces*, Interscience, New York (1950), reprinted by Springer-Verlag (1977).
- [15] Courant, R., and Hilbert, D., *Methods of Mathematical Physics, vol. I*, First English Edition, Interscience, New York (1953).
- [16] Dierkes, U., Hildebrandt, S., Küster, A., Wohlrab, O., *Minimal Surfaces*, 2nd edition, Grundlehren der mathematischen Wissenschaften **339**, Springer-Verlag, Berlin Heidelberg New York (1992, 2010).
- [17] Dierkes, U., Hildebrandt, S., Tromba, A., *Regularity of Minimal Surfaces*, 2nd edition, Grundlehren der mathematischen Wissenschaften **340**, Springer-Verlag, Berlin Heidelberg New York (1992, 2010).

⁷See also a revised version available on the author's web site, in which a direct argument is given that removes any dependence on [4].

- [18] Evans, L., *Partial Differential Equations*, Graduate Studies in Mathematics, vol. 19, American Mathematical Society, Providence, R. I. (1998).
- [19] Gilbarg, D., and Trudinger, N. S., *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag, Berlin/Heidelberg/New York (1977).
- [20] Gulliver, R. A minimal surface with an atypical boundary branch point, in: B. Lawson and K. Tenenblat (eds.), *Differential Geometry: a Symposium in honor of Manfredo P. do Carmo*, pp. 211-228, Longman, Harlow (1991).
- [21] Gulliver, R., and Lesley, F., On boundary branch points of minimizing surfaces, *Arch. Rational Mech. Anal.* **52**, 20–25 (1973).
- [22] Gulliver, R., Osserman, R., and Royden, H., A theory of branched immersions of surfaces, *Amer. J. Math.* **95**, 750–812 (1973).
- [23] Kato, T., *Perturbation Theory for Linear Operators*, second edition. Springer-Verlag, Berlin/Heidelberg/New York (1980).
- [24] Lewy, H. On the boundary behaviour of minimal surfaces, *Proc. Natl. Acad. Sci. USA* **37**, 102-110 (1951).
- [25] Micaleff, M. J., and White, B., The structure of branch points in minimal surfaces and in pseudoholomorphic curves. *Ann. Math.* **139**, 35–85 (1994).
- [26] Morgan, F., A smooth curve in R^4 bounding a continuum of area minimizing surfaces, *Duke Math. J.* **43** (1976), 867–870.
- [27] Morgan, F., A smooth curve in R^3 bounding a continuum of minimal manifolds, *Arch. Rat. Mech. Anal.* **75** (1981), 193–197.
- [28] Morrey, C., Jr., *Multiple Integrals in the Calculus of Variations*, Springer-Verlag, Berlin/Heidelberg/New York (1966).
- [29] Nitsche, J. C. C. *Lectures on Minimal Surfaces, Volume 1*, Cambridge University Press, Cambridge (1988).
- [30] Osserman, R., *A Survey of Minimal Surfaces*, Dover Publications, New York (1986).
- [31] Osserman, R., A proof of the regularity everywhere of the classical solution of Plateau’s problem, *Annals of Mathematics* **91**, 550-569 (1970).
- [32] Ruchert, H., A uniqueness result for Enneper’s minimal surface, *Indiana Univ. Math. J.*, **30(3)**, 427–431 (1981).
- [33] Tomi, F., On the local uniqueness of the problem of least area, *Arch. Ration. Mech. Anal.* **52**, 312-381 (1978).
- [34] Tomi, F., On the finite solvability of Plateau’s problem, *Lecture Notes in Mathematics* **597**, 679–695, Springer, Berlin Heidelberg New York (1977).

- [35] Tomi, F., and Tromba, A. J., On the structure of the set of curves bounding minimal surfaces of prescribed degeneracy, *J. für die reine und angewandte Mathematik* **316** (1980) 31–43.
- [36] Tromba, A. J., On the number of simply connected minimal surfaces spanning a curve, *Mem. Am. Math. Soc. No. 194*, **12** (1977).
- [37] Tromba, A. J., *A Theory of Branched Minimal Surfaces*, Springer-Verlag, Berlin Heidelberg 2012
- [38] Classical area minimizing surfaces with real analytic boundaries, *Acta Math.* **179**, no. 2, 295–305 (1997).