Finiteness in Plateau's Problem

Michael Beeson

San José State University
beeson@cs.sjsu.edu
www.MichaelBeeson.com/Research

January, 2009

Plateau's Problem

Given a Jordan curve Γ in R^3,
find a minimal surface bounded by Γ.

- Today we are interested only in real-analytic Γ
- And only in surfaces of the topological type of the unit disk.

$A(u)$ is the area of surface u. A *relative minimum* of area is a surface u bounded by Γ such that $A(u) \leq A(v)$ for all "nearby" surfaces v bounded by Γ. An *absolute minimum* of area has $A(u) \leq A(v)$ for *all* surfaces bounded by Γ.

- A relative minimum of area is a minimal surface
- But not necessarily conversely.

Soap films that don't disappear immediately correspond to relative minima.

Finiteness

Theorem: a real-analytic Jordan curve Γ cannot bound infinitely many relative minima of area.

Today I will give background material and, time permitting, an overview of the proof. There is no time to discuss related results and conjectures– see the last section of my paper.

The full paper is available on the Math ArXiv, and also at www.michaelbeeson.com (click *Research* and then *Publications* and go to the end of the list). Click *Talks* to find these slides.

Compactness

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.

The limit of a sequence of absolute minima is another absolute minimum. But that is not necessarily true for relative minima.

It turns out that the difficult case is when the sequence u_n converges to a surface u with a branch point; and the really difficult case is when the branch point is on the boundary.

Complex notation

$$u_z = \frac{\partial u}{\partial z} = \frac{1}{2}(u_x - iu_y)$$

$$u_{\bar{z}} = \frac{\partial u}{\partial z} = \frac{1}{2}(u_x + iu_y)$$

Then the harmonicity of u is expressed by $u_{\bar{z}} = 0$ and the conformality by

$$u_z^2 = 0$$

A minimal surface is a harmonic conformal map.

Branch points are the zeroes of u_z.

Weierstrass representation

We make use of the Enneper-Weierstrass representation of u:

$$u(z) = Re \left[\begin{array}{l} \frac{1}{2} \int f - f g^2 \, dz \\ \frac{i}{2} \int f + f g^2 \, dz \\ \int f g \, dz \end{array} \right]$$

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

The Gauss map and Gaussian area

The *Gauss map* is the unit normal

$$N(z) = \frac{u_x \times u_y}{|u_x \times u_y|}$$

considered as a map from the parameter domain to the sphere S^2. The *Gaussian image* is the range of this map, and the *Gaussian area* is the area of the range. In the case of a minimal surface, the Gauss map is conformal.

Total curvature

The Gaussian curvature K is the product $\kappa_1\kappa_2$ of the principal curvatures. W is the area element $\det g_{ij}$. KW is the Jacobian of the Gauss map N (considered as defined in the parameter domain); K is the Jacobian of N (considered as defined on the surface.

The *total curvature* of a surface is $\int KW\,dx\,dy$, or sometimes (loosely) the magnitude of this quantity.

The *total curvature* of a Jordan curve Γ is $\int_\Gamma \kappa\,ds$, where κ is the magnitude of the curvature vector. The *geodesic curvature*, which only makes sense relative to a surface u bounded by Γ, is the component of the curvature that is normal to u.

Gauss-Bonnet

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

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

Note that for minimal surfaces, KW is negative.

Gauss-Bonnet with Branch Points

If there are branch points, there is another term in the Gauss-Bonnet formula:

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

where M is the sum of the orders of the interior branch points and the half-orders of the boundary branch points. In words, m

hemispheres of Gaussian area can be replaced by one boundary branch point of order $2m$, and m full spheres of Gaussian area can be replaced by an interior branch point of order $2m$.

An eigenvalue problem on the sphere

Associated with the map N is a natural eigenvalue problem:

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

We are only interested in $\lambda = \lambda_{\min}$, the least eigenvalue.

Gaussian images and convergence

Suppose the minimal surfaces u_n converge to u.
Suppose that the u_n have no branch points, but u does have a branch point; for simplicity suppose it just has one branch point. In the Gauss-Bonnet formula, outside a tiny disk around the branch point, convergence is uniform. But an extra $M\pi$ appears in the limit. Hence that tiny disk around the branch point must contribute almost exactly M hemispheres worth of Gaussian area to the Gaussian image of u_n.

Convergence to a surface with an interior branch point

... that tiny disk around the branch point must contribute almost exactly M hemispheres worth of Gaussian area to the Gaussian image of u_n.

If the branch point is an interior branch point, then the sphere is covered almost exactly $M/2$ times in each neighborhood U of the branch point, for large n.

Convergence to a surface with a boundary branch point

... that tiny disk around the branch point must contribute almost exactly M hemispheres worth of Gaussian area to the Gaussian image of u_n.

Now the order of the branch point is $2M$.

If the branch point is a boundary branch and all the surfaces are bounded by Γ then the normal N to u_n lies almost in a plane (perpendicular to Γ at the branch point of u), so the Gaussian area is almost equal to that of M hemispheres.

How is that Gaussian area taken on? Is one hemisphere covered M times? What does the image of the boundary look like near the branch point? It can probably be incredibly wild.

The second variation of area

The surface u is varied to nearby surfaces $u + \varphi N$, where φ is zero on the boundary. The second variation of area is a bilinear functional on an appropriate Sobolev space of such ϕ. We write it $D^2A(u)$, and so we can write $D^2A(u)[\varphi, \psi]$, or when $\varphi = \psi$, for short just $D^2A(u)[\varphi]$.

The formula for the first variation of area is

$$DA(u)[\phi] = -\int\int H\phi W \, dx \, dy$$

Differentiating this we find after some calculation:

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

The eigenvalue associated with $D^2A[u]$

It is well known that the kernel of the second variation consists exactly in solutions of the eigenvalue equation we mentioned before:

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

Monotonicity of the least eigenvalue

If the domain shrinks, the eigenvalue increases:

$$\Omega \subset \Delta \qquad \text{implies} \qquad \lambda_\Delta < \lambda_\Omega$$

Eigenvalue problems on the sphere

An eigenfunction, composed with N, is still an eigenfunction (but not conversely). Therefore:

Passing to the sphere does not increase the least eigenvalue.

Therefore:

If the Gaussian image is contained in a hemisphere, then $\lambda_{min} > 2$.

(The eigenvalue of a hemisphere is 2).

N can be very complicated

N can in general define a multiple covering; in our work, N might cover, or nearly cover, both the "east" and "west" hemispheres but with many umbilical points (zeroes of ∇N), and it is difficult to prove theorems about eigenvalues in the general setting of a conformal map to the sphere.

The least eigenfunction has one sign

Otherwise, its absolute value could be entered as a candidate in computing the Rayleigh coefficient, and would not be smooth, violating regularity results.
Also, at the boundary, the normal derivative of the least eigenfunction is not zero. (Hopf lemma)

Dirichlet's integral

$$\begin{aligned} E(u) &:= \frac{1}{2}\int u_x^2 + u_y^2 \, dx\, dy \\ &= \frac{1}{2}\int u \cdot u_r \, d\theta \end{aligned}$$

The appropriate space is a Sobolev space of functions defined on the boundary (or harmonic extensions thereof).

Tangent vectors

A "tangent vector" in this space is a function k from S^1 to R^3 such that $k(\theta)$ is tangent to Γ at $u(\theta)$. These are the "directions" in which you can take the first or second variation of $E(u)$.
The "weak inner product" is given by

$$(h, k) := \int h_r k \, d\theta$$

The second variation of Dirichlet's integral

The second variation is given by

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

where $k = \lambda u_\theta$ and $\tilde{k} = \lambda u_r$.

The equation for the kernel of $D^2E(u)$ is then

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

This looks much prettier in the following form:

$$k_z \cdot u_z = 0$$

These formulas are due to Tromba.

The forced Jacobi directions

If the minimal surface u has branch points, let M be the number in the Gauss-Bonnet-Sasaki-Nitsche formula: the sum of the orders of interior branch points and half-orders of boundary branch points. Then there is an $2M$-dimensional subspace of the kernel of $D^2(u)$, of the form

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

where $i\omega z$ is meromorphic and has a pole of order $\leq m$ at each interior branch point of order m.
You only get one instead of two at a boundary branch point because the tangent vectors have to be tangent to the boundary.

Tangential variations

Although usually only normal variations are considered with area, you can also consider more general variations. It turns out that the first and second variation only depend on the normal part.
But with E, the forced Jacobi directions are all *exactly* tangential–they have zero normal component. Conversely, Tromba showed that a tangent vector k with no normal component lies in the space spanned by the forced Jacobi and conformal directions.

The connection between $D^2A[u]$ and $D^2E[u]$

(i) If k is in the kernel of $D^2E(u)$ then $\phi = k \cdot N$ is in the kernel of $D^2A(u)$.
(ii) Conversely, if ϕ is in the kernel of $D^2A(u)$, then ϕ arises as $k \cdot N$ for some k in the kernel of $D^2E(u)$.
We don't need part (ii), but it clarifies the situation.

Böhme and Tromba's global analysis

Böhme and Tromba applied nonlinear global analysis to the theory of minimal surfaces. This is deep and beautiful work, and they drew deep and beautiful consequences from it, but we need only one thing from it.

Compactness

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.
But thanks to Böhme, we know more: there must be (at least) a one-parameter family $u(t)$ of minimal surfaces, all bounded by Γ, and real-analytic in t, x, and y jointly, for some interval of t values.

Tomi's theorem on absolute minima

Tomi's theorem: finiteness for absolute minima and analytic boundaries

Proof. If Γ bounds infinitely many absolute minima, then by compactness, there is one which is not isolated; hence there is a one-parameter family of minimal surfaces u^t bounded by Γ. In some neighborhood of $t = 0$, they all have the same E and the same area, and hence are all absolute minima. By compactness this family must loop, i.e. for some positive t we have $u^t = u^0$. The set of t for which u^t is an absolute minima is open and closed; hence all the u^t are absolute minima.

Proof of Tomi's theorem, continued

For each t, u_t is in the kernel of $D^2E[u^t]$, so $\phi = u_t \cdot N$ is in the kernel of $D^2A[u^t]$, i.e., it is an eigenfunction. Here we use that u^t has no branch points, so u_t is not a forced Jacobi direction.

Pick a point P some ways from the surfaces u_n and consider the signed volume

$$V(t) := \int (u - P) \cdot (u_x \times u_y)\, dx\, dy = \int (u - P) \cdot N\, dA$$

For some value of t this has a minimum. But

$$\frac{dV}{dt} = \int u_t \cdot N\, dA = \int \phi\, dA$$

cannot be zero, since ϕ, as the eigenfunction of the least eigenvalue, has one sign. Contradiction, QED.

The starting point of my work

If there are infinitely many relative minima bounded by Γ, then again there is a one-parameter family of them. But now, there is no immediate reason why this family cannot run into a minimal surface with a branch point. The situation would be,

- a one-parameter family of minimal surfaces u^t defined in the upper half plane for t in some interval about 0.
- each u^t is bounded by Γ and $u^t(0) = 0$.
- Γ is tangent to the X-axis at the origin.
- when $t = 0$, u^0 has a boundary branch point at the origin.
- when $t > 0$, u^t is a relative minimum (of area and Dirichlet's integral)

Why can't that happen?

Some observations

- for $t > 0$ we have $\lambda = 2$ (since u^t is a relative minimum and u_t is in the kernel)
- for $t = 0$ we also have $\lambda = 2$, as the eigenvalue is continuous. But maybe u^0 is not a relative minimum.
- while u_t is in the kernel of $D^2A(u)$ for $t > 0$, it might be a forced Jacobi vector when $t = 0$, and hence have no normal component.

Some more observations

Let the boundary branch point have order $2m$. (It has to be even since u takes the boundary monotonically.)

- For $t > 0$, the Gaussian area is about m hemispheres more than for $t = 0$.
- The image of N lies very close to the plane $X = 0$ when t is small, but it might wrap wildly around that great circle, reversing direction many times, etc.
- if the image of the parameter domain includes the entire upper hemisphere and a bit more, then $\lambda > 2$, contradiction.

Plan of the proof

- Analyze the Weierstrass representation as a function of t
- Calculate u_t and the eigenfunction $\phi = u_t \cdot N$
- use the fact that ϕ has one sign for $t > 0$ to get information about u

Only after we carry out a few steps of this analysis can we discuss the real difficulties of the proof.

The Weierstrass representation as a function of t

Let $f = {}^1u_z - i^2u_z$ and g, the stereographic projection of N, be the functions in the Weierstrass representation. The branch points are the common zeroes of f and fg^2. If the branch point at origin when $t = 0$ is an interior branch point, there are no common zeroes, and for $t > 0$ the zeroes of f are double. Hence the order of the branch point is $M = 2m$ and we define

- $a_i = a_i(t)$ are the zeroes of f for $i = 1, \ldots, m$
- $b_i = b_i(t)$ are the zeroes of fg^2 for $i = 1, \ldots, m + k$

Attack of the alien branch points

In the boundary branch point case, there might be branch points for $t > 0$, lying outside the parameter domain, but converging to the origin as $t > 0$. In that case there will be some common zeroes s_i of f and fg^2, say N of them, and we let a_i and b_i be the *other* zeroes of f and fg^2, so that

$$
\begin{aligned}
f(z) &= \Pi(z-a_i)^2\Pi(z-s_i) \\
fg^2(z) &= \Pi(z-b_i)^2\Pi(z-s_i)
\end{aligned}
$$

and the a_i are not equal to the b_j for $t > 0$. All the a_i and b_j converge to 0 as $t \to 0$.

Analytic dependence of the a_i on t

We want to have the a_i, b_i, and s_i depend analytically on t. This follows from the theorem that "analytic sets are analytically triangulable". (An analytic set is locally the zeroes of a real analytic function.) That is, the zero set of f, for example, is triangulable, and the maps from the simplices to the zero set are real-analytic. Several references to proofs of this fact are given in my 1980 paper.

Then, after a reparametrization from the original t to a new parameter, the a_i depend analytically on t.

The rate of convergence of the a_i to zero

Each $a_i(t)$ then goes to zero as some power of t:

$$\begin{aligned} a_i(t) &= \alpha_i t^{\gamma_i} \\ b_i(t) &= \beta_i t^{\text{some power}} \\ s_i(t) &= \sigma_i t^{\text{some power}} \end{aligned}$$

and similarly for the b_i and s_i. Let γ be the smallest of these exponents. The "principal roots" are those that go to zero as t^γ. The rest (the "fast roots") go to zero faster.

The w-plane

We introduce

$$w := \frac{z}{t^\gamma}$$

so that

$$z = t^\gamma w$$

Then the positions of the α_i, β_i, and σ_i in the w-plane are of interest, and by removing powers of t, we can derive formulas for various quantities of interest that are valid in the w-plane when $t = 0$.

The unit normal in the w-plane

As an example of the principle: on compact subsets of the w-plane away from the α_i, the unit normal N converges to $(0,0,-1)$ as $t \to 0$.

Of course, for small positive t, the unit normal behaves wildly near the α_i. But this wild behavior is confined to smaller and smaller neighborhoods of α_i as $t \to 0$.

Even if there is a boundary branch point, for small $t > 0$ the unit normal is confined to be very near the YZ plane, as that tiny part of Γ is almost on the X-axis.

The case of an interior branch point

In the case of an interior branch point, the boundary is not relevant in the w-plane–it recedes to infinity as $t \to 0$. Hence, for small positive t, some disk D about α_i does not meet the boundary. For small enough t, the pole $\mathbf{a}_i(t) = a_i(t)/t^\gamma$, which converges to α_i as $t \to 0$, is inside that disk. Hence N takes on the north pole near α_i, for small positive t.

But on the boundary of D, N converges to $(0, 0, -1)$. Hence on the interior of D, N covers at least the upper three-quarters of the sphere for small enough t. Then $\lambda < 2$, since more than a hemisphere is covered. Contradiction!

Some special cases of finiteness

Eliminating the interior branch point case was the main result of my 1980 paper. Some cases of finiteness follow. For example, if Γ lies on the boundary of a convex body, there can't be boundary branch points, so we have finiteness.

To answer a question raised by Tromba: Note that the question of whether the branch point might be a false branch point did not arise. Although we did appeal to the regularity theorems for $t > 0$, we did not need anything about the hypothetical branch point at $t = 0$.

The case of a boundary branch point

Then we take the parameter domain to be the upper half-plane, and the branch point is at the origin when $t = 0$. Then we cannot finish the argument immediately, because all we know is that the Gauss map takes on approximately m hemispheres near the origin for $t > 0$. This "extra Gaussian area" disappears when t becomes zero and the branch point term appears in the Gauss-Bonnet formula. These coverings of hemispheres "pinch off" from the Gaussian image as t goes to zero.

Calculating in the w-plane

Since we cannot complete the "eigenvalue argument", the next step is to compute the eigenfunction, and use the fact that it must have one sign.

The eigenfunction is $\phi = u_t \cdot N$, and both u_t and N can be computed in terms of the roots a_i, b_i, and s_i. To get started we note that

$$\begin{aligned} z - a_i &= t^\gamma w - \alpha_i t^\gamma \\ &= t^\gamma (w - \alpha_i) \end{aligned}$$

Thus the "fast roots", that go to zero faster than t^γ, will each contribute just a factor of w^2 to $f(z)$, while the principal roots will each contribute $(w - \alpha_i)^2$.

Definition of $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{S}$

We define

$$\begin{aligned} \mathbb{A}(w) &= \prod (w - \alpha_i) \\ \mathbb{B}(w) &= \prod (w - \beta_i) \\ \mathbb{S}(w) &= \prod (w - \sigma_i) \end{aligned}$$

Everything of interest will be calculated in terms of $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{S}$.

The formula for the eigenfunction

It is not possible in a one-hour talk to go through computations. For example, to start the computation, we observe that $g = t^{k\gamma}\mathbb{B}/\mathbb{A}$ is the stereographic projection of the unit normal N, so $\mathbb{B}/\mathbb{A}$ contains a lot of information about how the normal behaves as $t \to 0$.

Here is the result of the computation of the eigenfunction:

$$\phi \;=\; -\frac{1}{2}t^{(2m+k+1)\gamma-1}\operatorname{Im}\mathbb{H}$$

where

$$\mathbb{H}(w) \;=\; \frac{\mathbb{B}}{\mathbb{A}}(2m+1)\int_0^w \mathbb{A}^2\mathbb{S}\,dw - (2m+k+1)\int_0^w \mathbb{A}\mathbb{B}\mathbb{S}\,dw$$

The main obstacle

The formula for ϕ has the form $-\frac{1}{2}t^{2m+k+1}\mathrm{Im}\,\mathbb{H}(w)$, for a function $\mathbb{H}$ that we can calculate in terms of the principal roots. But $\mathbb{H}$ might turn out to be constant, so that $\mathrm{Im}\,\mathbb{H}$ is identically zero.

Why can't that happen?

A discouraging example

There is a family of minimal surfaces u^t, all bounded by a straight line, having a branch point when $t = 0$. This family exhibits some of the local behavior we are trying to rule out. In 1980 I found this family and thought it was the end of the line. I did not return to the problem until 2000.
I have made 3d computer-graphics pictures of this family in which you can fly around and observe the wild curvature near where the branch point eventually appears. But $\phi = u_t \cdot N$ is not of one sign in the example, something I did not notice in 1982.

Equations for the case $\mathbb{H}$ constant

It turns out that in case $\mathbb{H}$ is constant, a rather remarkable equation must hold:

$$\frac{\mathbb{B}}{\mathbb{A}} = \frac{B_0\beta}{\alpha}\left(\int \mathbb{A}^2\mathbb{S}\,dw\right)^{k/(2m+1)}$$

where β is the product of all the β_i and α is the product of the α_i (over the principal roots, so no of these are zero); and B_0 is coefficient of the leading term of fg^2. This equation is not hard to derive, by differentiating the equation for $\mathbb{H}$.

What is true if $\mathbb{H}$ is not constant

By analyzing the asymptotic behavior of $\mathbb{H}$ in the vicinity of the origin, and again in the vicinity of each α_i and β_i, we prove that

- There is at least one principal a_i or s_i (whether or not $\mathbb{H}$ is constant)
- if $\mathbb{H}$ is not constant, *all* the roots α_i, s_i, and β_i are principal.
- if $\mathbb{H}$ is not constant, then $\mathbb{B}/\mathbb{A}$ has either a simple zero or a simple pole at each α_i, or is analytic non-vanishing there.

Gaussian area and boundary branch points

- $\mathbb{A}^2\mathbb{S}$ is real, so the α_i and σ_i come in complex-conjugate pairs.
- All the α_i (limits in the w-plane of poles of g) are on the real axis. (Otherwise we could use the eigenvalue argument.)
- Near each α_i, for small t, the Gaussian area in a small disk is about m spheres.
- But maybe it's all contributed in the lower half plane.
- We hope each α_i contributes one hemisphere in the upper half plane.
- That would make $2\pi m$ in the Gauss-Bonnet formula.
- Some of the α_i could be equal.
- Some of the a_i may be non-principal, i.e. go to zero faster than t^γ.

The unit normal near the a_i

- The "north pole" is $(0, 0, 1)$
- The "south pole" is $(0, 0, -1)$
- The "west pole" is $(-1, 0, 0)$
- The "east pole" is $(1, 0, 0)$

A root a_i *contributes Gaussian area* if N covers either the east or west pole near a_i for arbitrarily small t. Since N restricted to the boundary lies near the plane $X = 0$, this means it covers at least (almost) a hemisphere near a_i.

The eigenfunction and Gaussian area

- The unit normal N does not cover both the east and the west pole near any α_i (in the upper half plane).
- This is proved by calculating a more exact formula for the eigenfunction, valid even near the α_i.

It turns out that at a point ξ where N takes on the east or west pole, we have

$$\frac{2}{t^{2m+1}}\phi = {}^1Nc(1 + O(\xi - \alpha_i)) + O(t)$$

where the constant c is nonzero. Here ${}^1N$ is the first component of N. So for small t, the sign of ϕ is the same as the sign of ${}^1N$. But ϕ has to be of one sign in the upper half plane.

If a principal root contributes Gaussian area, then $\mathbb{H}$ is not constant

Suppose N covers the west pole near α_i. Since N is a covering map, we can go around the equator in either direction to reach either of the two points $(0,1,0)$ and $(0,-1,0)$, and trace out the pre-image in the parameter domain. We must hit the boundary, or else we will get past the plane $X = 0$ and reach the east pole, contradicting the previous slide. Let the places where we hit the boundary be called $eq_1(t)$ and $eq_2(t)$. These points converge to α_i as $t \to 0$ and since $\alpha_i \neq 0$, they both have the same sign.

Making use of $eq_1(t)$ and $eq_2(t)$

Now ${}^2N$ can be calculated in terms of $\mathbb{B}/\mathbb{A}$, and if $\mathbb{H}$ is constant, recall

$$\frac{\mathbb{B}}{\mathbb{A}} = \text{constant}\left(\int_0^w \mathbb{A}^2 \mathbb{S}\, dw\right)^{k/(2m+1)}$$

In a few steps of calculation, we show that ${}^2N$ has the same sign at $eq_1(t)$ and $eq_2(t)$, contradiction. Hence $\mathbb{H}$ is not constant.

So what?

Now if we knew that some principal root had to contribute Gaussian area, then we would know that H is not constant. But all we can conclude is that if H is constant, then all the Gaussian area must be contributed by the "fast" roots, the ones going to zero faster than t^γ.

After many mistakes, I eventually faced the fact that it is necessary to analyze the fast roots, too–not just the principal roots.

The rings of roots

Each root a_i goes to zero according to *some* power of t. Arrange these powers t_i in descending order, so $t_1 = \gamma$, the power of the principal roots. Think of the roots as constituting "rings" (although at least the a_i lie close to the x-axis), with the larger γ_i corresponding to the "inner rings".

A w-plane for each ring

We can introduce a w-plane for each ring, with $z = t^{\gamma_i} w$. Fixing one γ_i, there are "slow" roots, going to zero as the power γ_i or slower, and "fast" roots, going to zero faster than t^{γ_i}

The roots that go to zero slower that t^{γ_i} go off to infinity in the w-plane as t goes to zero, but they may have a big influence anyway.

The N-condition

The N-condition says that on compact subsets away from the α_i, the unit normal tends uniformly to the south pole:

$$N = (0, 0, -1) + O(t)$$

This is obvious for the principal roots, but not for the inner rings.

Descending the rings of roots

The key observation (only made in 2008) is this: *If* $\mathbb{H}$ *is constant on one ring, then the* N*-condition holds on the next ring down.*

Of course this requires some calculation to establish.

All that we did for the principal roots now carries over to the inner rings as well, with more complicated expressions $\mathcal{K}$ and $\mathcal{M}$ replacing k and $2m+1$. In the paper it is only done once, instead of twice, which may make it seem more complicated. In particular the theorem that if $\mathbb{H}$ is not constant, all the roots are principal, becomes, if $\mathbb{H}$ is not constant on a certain ring of roots, then no root goes to zero faster than that ring.

$\mathbb{H}$ is not constant on some ring

Some root (in *some* ring) must contribute Gaussian area, by the Gauss-Bonnet formula. So if $\mathbb{H}$ is constant on all the slower rings than that one, the N-condition will hold on that ring, and hence $\mathbb{H}$ will *not* be constant on that ring. In other words, either $\mathbb{H}$ is already not constant on some slower ring, or it is not constant on the first ring that contributes Gaussian area.

All the roots are principal and the alien branch points don't occur

Since N cannot cover both hemispheres near a_i (in the upper half plane), each a_i contributes at most one hemisphere to the Gaussian area. When $\mathbb{H}$ is not constant (on a given ring) then there are no faster roots than that ring. True, there can be multiple a_i converging to the same α_i, but the number of sheets over the north pole is the number of those roots, so it works out to one hemisphere per a_i. Hence: to get the required $2m\pi$, there must be m roots in the first ring where Gaussian area is contributed. Hence that ring must be the first one, and what is more, the alien branch points do not occur, since if they did, there would be fewer than m of the a_i.

$k/(2m+1)$ is an integer

This computation takes ten pages in section 12 of my paper, and the following slides, which must be skipped in a one-hour talk, just give the first few steps.

The parametrization of Γ

$$\Gamma' = \begin{bmatrix} 1 \\ C_2\tau^q(1+O(t)) \\ C_3\tau^p(1+O(t)) \end{bmatrix}$$

The first component is exactly 1 (no higher-order terms).
Γ is tangent to the X-axis, and p and q control how much it deviates from a straight line.

The plan to prove $k/(2m+1)$ is an integer

Compute u_t on the boundary in terms of w in two different ways:

- From the parametrization of Γ, since

$$u_t = u_\tau \frac{d\tau}{dt}$$

- From the Weierstrass representation, using

$$u_t \quad \cong \quad \left(\frac{\partial}{\partial t} - \gamma t^{-1} w \frac{\partial}{\partial w} \right) u$$

Setting the two results equal

$$u_t \cong \begin{bmatrix} 1 \\ C_2\tau^q \\ C_3\tau^p \end{bmatrix} \gamma t^{(2m+1)\gamma-1}\left\{(2m+1)\int_0^w \mathbb{A}^2\,dw - \mathbb{A}^2 w\right\}$$

$$u_t = \gamma t^{(2m+1)\gamma-1}\begin{bmatrix} (2m+1)\int_0^w \mathbb{A}^2\,dw - \mathbb{A}^2\,w \\ \mathrm{Im}\left((2m+1)\int_0^w \tilde{\mathbb{A}}^2\,dw - \tilde{\mathbb{A}}^2\,w\right) \\ t^{k\gamma}(\mathrm{Im}\left((2m+k+1)\int_0^w \tilde{\mathbb{A}}\tilde{\mathbb{B}}\,dw - \tilde{\mathbb{A}}\tilde{\mathbb{B}}\,w\right) \end{bmatrix}$$

We get a handle on the leading powers of t in $\mathrm{Im}\,\tilde{\mathbb{A}}^2$ and $\mathrm{Im}\,\tilde{\mathbb{A}}\tilde{\mathbb{B}}$.

Sigma

$$\sigma := \int_0^w \tilde{\mathbb{A}}^2 \, dw$$

Note that σ is defined using $\tilde{\mathbb{A}}$, not $\mathbb{A}$, so it depends on t and it has an imaginary part on the boundary. Except for a power of t, σ is close to arc length on Γ, which is close to X.

Equating the two expressions

The result of equating the two expressions for the second component of u_t is

$$C_2 t^{(2m+1)q\gamma}\sigma^q\Big((2m+1)\sigma - \mathbb{A}^2\, w\Big) \;\cong\; \mathrm{Im}\,\Big((2m+1)\sigma - \tilde{\mathbb{A}}^2\, w\Big)$$

The left side isn't zero, since when H is not constant, there is at least one principal a_i. (In fact all the a_i are principal.)

This is just the start of the ten-page computation. For the rest of it, you must read the paper. Here I emphasize that the whole use of the boundary parametrization is in proving that $k/(2m+1)$ is an integer, and that computation is independent of the rest of the paper, except for the fact that if $\mathbb{H}$ is not constant, then all the a_i are principal.

The $\mathbb{H}$ equations as a differential equation for $\mathbb{B}/\mathbb{A}$

Let $\tilde{\mathbb{H}}$ be the eigenfunction of u^t. (The tilde indicates t-dependence.) Then, differentiating the equation for $\tilde{\mathbb{H}}$, we find that $f = \tilde{\mathbb{B}}/\tilde{\mathbb{A}}$ satisfies the differential equation

$$\tilde{\mathbb{H}}_w = (2m+1)f' \int_0^w \tilde{\mathbb{A}}^2 \, dw - k\tilde{\mathbb{A}}^2 f$$

Let

$$\sigma := \int_0^w \tilde{\mathbb{A}}^2 \, dw.$$

Then the equation above is

$$\tilde{\mathbb{H}}_w = (2m+1)f'\sigma - k\sigma_w f \qquad (1)$$

A more convenient differential equation

After some manipulation we find another form of the equation.

$$\mathbb{H}_w = (2m+1)\sigma^{k/(2m+1)+1}\left(\frac{f}{\sigma^{k/(2m+1)}}\right)'$$

The fraction $f/\sigma^{k/(2m+1)}$ is not identically 1, since then $\mathbb{H}_w$ would be zero, and $\mathbb{H}$ would be constant, but we proved that $\mathbb{H}$ is not constant.

Asymptotic behavior at infinity

Looking at the behavior for large w, we have

$$\begin{aligned}
\sigma &= \int_0^w \mathbb{A}^2\,dw = \frac{w^{2m+1}}{2m+1} + O(w^{2m}) \\
\sigma^{k/(2m+1)} &= c_3 w^k + O(w^{k-1}) \qquad \text{for some } c_3 \\
\sigma^{k/(2m+1)+1} &= c_3 w^{2m+k+1} + O(w^{2m+k}) \\
f &= \mathbb{B}/\mathbb{A} = w^k + O(w^{k-1})
\end{aligned}$$

Hence $f/\sigma^{k/(2m+1)}$ has both numerator and denominator asymptotic to w^k.

How we use that $k/(2m+1)$ is an integer

Whether or not $k/(2m+1$ is an integer, we have for large w

$$\frac{f}{\sigma^{k/(2m+1)}} = 1 + cw^{-J} + O(w^{-J-1}) \tag{2}$$

for some $J \geq 1$ and constant c. But since $k/(2m+1)$ is an integer, we are dealing with a rational function, whose numerator and denominator are polynomials of degree k. In that case we have $J \leq k$, since the worst case (largest J) is when the numerator and denominator have all their corresponding coefficients the same except the constant coefficient, and in that case $J = k$. For a more rigorous proof, see the paper.

For example

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

illustrating that $J = k$ in this case. There is nothing special about 2 and 5.

Integrate the equation to calculate $\mathbb{H}$

$$\begin{aligned}
\mathbb{H}_w &= (2m+1)\sigma^{k/(2m+1)+1}\left(\frac{f}{\sigma^{k/(2m+1)}}\right)' \\
&= (2m+1)w^{2m+k+1}(1+cw^{-J}+O(w^{-J-1}))' \\
&= (2m+1)w^{2m+k+1}(-Jcw^{-J-1}+O(w^{-J-2})) \\
&= -J(2m+1)cw^{2m+k-J}+O(w^{2m+k+J-1})
\end{aligned}$$

Integrating, we have

$$\mathbb{H} = \frac{-J(2m+1)c}{2m+k+J+1}w^{2m+k-J+1}+O(w^{2m+k-J})$$

The final contradiction

Since $\mathrm{Im}\,\mathbb{H}$ must have one sign in some neighborhood of infinity, we must have $2m + k - J + 1 = \pm 1$. That is, $J = 2m + k + 1 \pm -1$. Since $k/(2m+1)$ is an integer, and hence $J \leq k$ as shown above, subtracting J from each side we have

$$\begin{aligned} 0 &= 2m + 1 + (k - J) \pm 1 \\ &\geq 2m + (k - J) \\ &\geq 2m \qquad \text{since } k - J \geq 0 \\ &\geq 1 \end{aligned}$$

contradiction, since $m \geq 1$. This contradiction completes the proof.