Finding Hausdorff Dimension of Limit Sets Using Spectral Theory

Alex Karlovitz

1 A First Example

We are interested in the Hecke group Γ_R for $R > 0$ given by

$$\Gamma_R = \left\langle \begin{pmatrix} 1 & R \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\rangle$$

This of course corresponds to the maps $z \mapsto z + R$ and $z \mapsto -1/z$. In particular, we will look at the case $R = 20/7$.

It is straightforward to show that $\Gamma_{20/7}$ is conjugate to the group

$$\Gamma = \left\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -\frac{7}{20} \\ \frac{20}{7} & 0 \end{pmatrix} \right\rangle$$

For our purposes, it is sufficient to consider Γ, since conjugation preserves the Hausdorff dimension of the limit set.

In Figure 1, we see a fundamental domain for $\Gamma \backslash \mathbb{H}$. This is clear because $z \mapsto z + 1$ gives the horizontal shift by 1, and $z \mapsto -R^2/z$ results in inversion through the circle of radius R composed with reflection across the imaginary axis. Notice that this is an infinite volume fundamental domain, which is one way in which this case differs from the original $SL(2, \mathbb{Z})$ case.

We are interested in finding the Hausdorff dimension δ of the limit set of $\Gamma \backslash \mathbb{H}$. Patterson-Sullivan theory tells us that δ is related to the smallest eigenvalue λ_0 achieved by Maass cusp forms on Γ. The relation is $\lambda_0 = \delta(1 - \delta)$. So, we will compute δ by applying Hejhal's algorithm to Γ. (One can even show that in the case of Γ_R for $R \geq 2$, there is exactly one eigenvalue, and it falls between 0 and $1/4$).

Now let's discuss applying Hejhal's algorithm to Γ. Suppose f is a Maass cusp form for Γ. (See notes from my learning seminar talk "Maass Cusp Forms and Hejhal's Algorithm" for more details). Since $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix} \right) \in \Gamma$, we have the same Fourier expansion as in the $SL(2, \mathbb{Z})$ case. Thus, the same argument where we applied the Laplacian to the Fourier expansion applies here. So we have that

$$f(z) = a_0(y) + \sum_{n \neq 0} a_n \sqrt{y} K_\nu(2\pi|n|y) e(nx)$$

where $\nu \in \mathbb{C}$ is chosen so that f is an eigenfunction of the Laplacian with eigenvalue $1/4 - \nu^2$. The only difference in this case is that we are no longer working with a cusp form. This is actually simpler than the non-constant terms case. Since there is no x in the constant term, the differential equation is

$$\frac{1}{y^2} a_0''(y) - \lambda a_0(y) = 0$$

where $\lambda = \delta(1 - \delta)$ and δ is chosen to be greater than $1/2$. It is straightforward to see that the only two solutions are y^δ and $y^{1-\delta}$. By the assumption that our Maass form is square-integrable around the cusp, we are able to throw out the y^δ term. Hence,
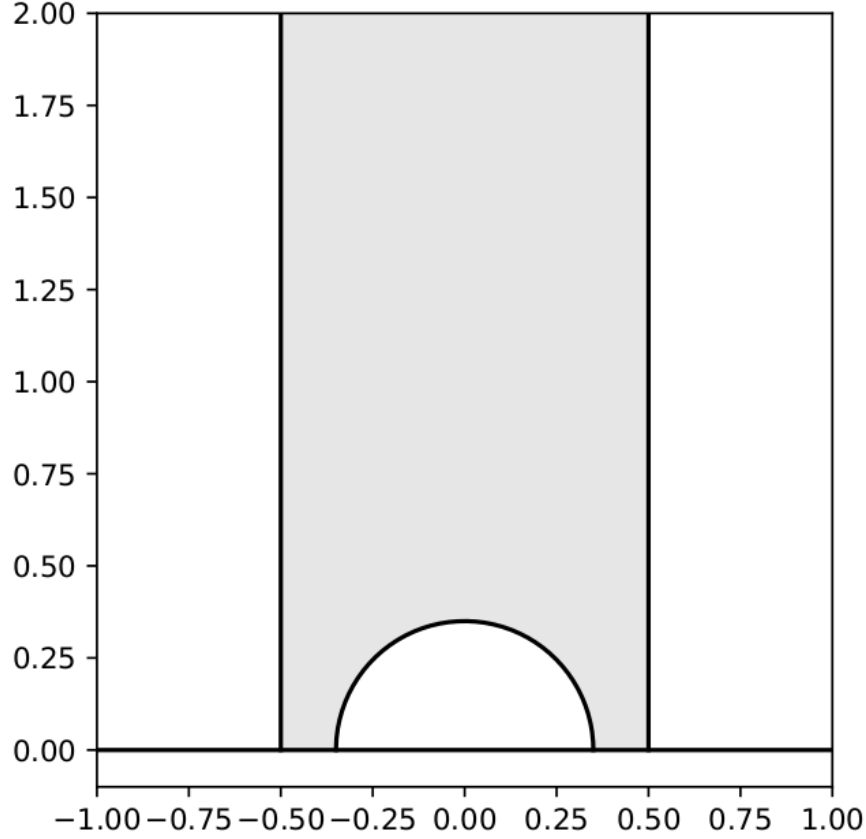
$$a_0(y) = a_0 y^{1-\delta} = a_0 y^{1/2 - \nu}$$

Figure 1: Fundamental Domain for $\Gamma\backslash\mathbb{H}$

It is straightforward to show (again, see the learning seminar talk) that $\nu = ir$ for $r > 0$ or $\nu \in \mathbb{R}$ with $|\nu| < 1/2$. It turns out that in this case, there is indeed a real choice for ν in the spectrum. **Why is this true?** This is another way in which our current case will differ from the original $SL(2, \mathbb{Z})$ case.

The only other way this case will differ from the $SL(2, \mathbb{Z})$ case is in the different fundamental domain. To modify the existing code to fit this case, I will need to

- Modify the $\mathcal{F}$-pullback function to work with $\Gamma\backslash\mathbb{H}$ instead of $SL(2, \mathbb{Z})\backslash\mathbb{H}$

- Make sure K-Bessel computations use subscript ν, not just ir

- Add in the a_0 term

1.1 Testing

Let's discuss how we can test the code. Kontorovich and Strömbergsson's result is that the Hausdorff dimension is

$$\delta = 0.7670524170091020567715020886425950627668(7)$$

The last digit is in parentheses to indicate that it may be off by ± 1. This suggests that the smallest eigenvalue $\lambda_0 = \delta(1 - \delta)$ has the following first 25 digits:

$$\lambda_0 \approx 0.1786830065695966584789154$$

Finally, recall that Hejhal's algorithm steps towards a value for ν where $\lambda_0 = 1/4 - \nu^2$. Hence, we are actually looking for the value

$$\nu \approx 0.2670524170091020567715020886425950627668(7)$$

This is just $\nu = \delta - 1/2$.

1.2 Results

June 27 2019 I used the following parameters:

- $D = 50$

- $M_0 = 30$

- $Q = 100$

- $Y_1 = 1/10$ and $Y_2 = 1/300$

I started with $\nu = 0.26$ and let the algorithm run. Immediately, the program chose to step towards 0.19. Since the true value starts out $\nu = 0.26...$, the code is not working as it should. This is perhaps due to this method requiring a much higher degree of precision to observe the correct eigenvalue. Another possibility is that $1/10$ is too high (meaning too many sample points are in the standard fundamental domain). Of course, there is also a chance that there is an error in the code.

I also checked what would happen if I changed Y_1 to $1/20$ and M_0 to 80. The same results were observed.

July 9 2019 I tried using a higher precision with a larger number of Fourier coefficients to see if this would help. I used the following parameters:

- $D = 80$

- $M_0 = 80$

- $Q = 100$

- $Y_1 = 1/10$ and $Y_2 = 1/100$

This also stepped towards 0.19, interestingly. However, this is still incorrect. Perhaps an even higher precision...?

2 A Disk Model Approach to the Same Problem

Recall that we can map our work in the upper half plane to the unit disk via

$$z \mapsto \frac{z - i}{z + i}$$

(The inverse map is $\tau \mapsto i(1+\tau)/(1-\tau)$, where τ is taken in the unit disk). Under this map, we can deduce the image of the fundamental domain from the previous section as follows. First, notice that

$$\pm \frac{1}{2} \mapsto -\frac{3}{5} \mp \frac{4}{5}i \qquad \pm \frac{7}{20} \mapsto -\frac{351}{449} \mp \frac{280}{449}i \qquad \infty \mapsto 1$$

Next, since the geodesics in the disk model are circles normal to the unit circle, we can find the image of the geodesics from the upper half plane using these endpoints (see Figure 2).

The new idea here is to apply Hejhal's algorithm to the fundamental domain in $\mathbb{D}$. To do this, we begin by writing a Maass form f in polar form $f(\rho, \theta)$. Since f is 2π-periodic in θ, it has a Fourier expansion

$$f(\rho, \theta) = \sum_n a_n(\rho) \exp(in\theta)$$

Next, since f is a Maass form, it is an eigenfunction of the hyperbolic Laplace operator, say $\Delta f = s(1-s)f$. After mapping to $\mathbb{D}$ and converting to polar coordinates, this operator is

$$\Delta = \frac{(1-\rho^2)^2}{4} \frac{\partial^2}{\partial \rho^2} + \frac{(1-\rho^2)^2}{4\rho} \frac{\partial}{\partial \rho} + \frac{(1-\rho^2)^2}{16\rho^2} \frac{\partial^2}{\partial \theta^2}$$
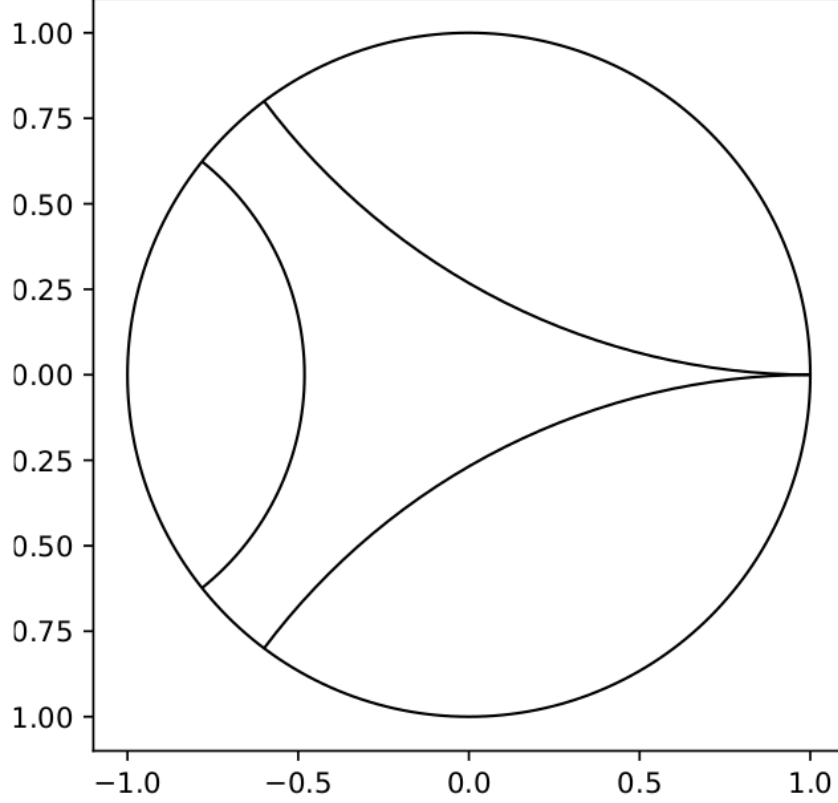
Figure 2: Fundamental Domain in Disk Model

Applying this map to the Fourier expansion, then using uniqueness of Fourier coefficients, we have that the $a_n(\rho)$'s satisfy the differential equation

$$\frac{(1-\rho^2)^2}{4}a_n''(\rho) + \frac{(1-\rho^2)^2}{4\rho}a_n'(\rho) - \frac{n^2(1-\rho^2)^2}{16\rho^2}a_n(\rho) = s(1-s)a_n(\rho)$$

As with the original case (where we worked in $\mathbb{H}$ and used the Fourier expansion at the cusp ∞), one finds that this differential equation has two solutions. Only one of the solutions exhibits the appropriate decay, and hence we are able to deduce the form of the $a_n(\rho)$'s. The result is that

$$a_n(\rho) = a_n(1-\rho^2)^2\rho^{|n|}\,_2F_1(s, s+|n|, 1+|n|; \rho^2)$$

where $_2F_1$ is a hypergeometric function.

To get a linear system, we first approximate f by a finite Fourier expansion with $|n| \leq M$ (we are no longer assuming f is a cusp form). Then, we fix a distance $\rho = P$ from the origin and take $N \geq 2M + 1$ equally spaced θ's to produce sample points. For example, we could take the points (P, θ_j) where

$$\theta_j = \frac{2\pi j}{N} \qquad \text{for} \qquad 0 \leq j < N$$

We can then extract an expression for the Fourier coefficients by taking an appropriate linear combination of the function evaluated at the sample points. This expression is

$$a_n(P) = \frac{1}{N}\sum_{j=0}^{N-1} f(P, \theta_j)e^{-in\theta_j}$$

4

Finally, we can set up a $(2M+1) \times (2M+1)$ linear system by using the Γ-automorphicity of f. Let (P_j^*, θ_j^*) be the point in the chosen fundamental domain which is equivalent to (P, θ_j) mod Γ. Some straightforward computation gives the following system

$$a_n c_n(P) = \sum_{|\ell| \leq M} a_\ell V_{n\ell} \tag{1}$$

where

$$c_n(\rho) = (1 - \rho^2)^2 \rho^{|n|} \, {}_2F_1(s, s + |n|, 1 + |n|; \rho^2) \quad \text{and} \quad V_{n\ell} = \frac{1}{N} \sum_{j=0}^{N-1} c_\ell(P_j^*) e^{i(\ell \theta_j^* - n\theta_j)}$$

Note: recall that we set $s = 1/2 + \nu$ so that $\lambda = s(1-s) = 1/4 - \nu^2$.

2.1 Results

I plan to first test the disk method on $SL(2, \mathbb{Z})$ to make sure it is working correctly. This, of course, is the Hecke group with parameter $R = 1$.

August 24 2019

I tested the disk method on $SL(2, \mathbb{Z})$ with the following inputs:

- $D = 50$

- $M_0 = 50$

- $N = 100$

- $\rho_1 = 9/10, \rho_2 = 49/50$

The final output was 9.619465, which is incorrect. The correct r-value near this is $9.53\ldots$. Perhaps we require higher precision...

August 27 2019

I redid the test from August 24 with the following changes:

- $D = 150$

- $M_0 = 100$

- $N = 250$

This also failed, as it found the value $8.9\ldots$. We will have to investigate what is causing this to fail. One immediate thought is to print the values which go into the linear system and see if any of these are "too large" or "too small." Perhaps this causes the numerics to be unstable.

I ran another test today with the same ρ_j's, but this time with

- $D = 50$

- $M_0 = 50$

- $N = 101$

I printed out the values which are the pieces to build the linear system, and it appears the problem comes from the asymptotics of the hypergeometric function (See Section 2.2 below). I observed some values from the hypergeometric function as small as 10^{-60}, while others were on the order of 10^{-3}.

August 29 2019

Today, I attempted to use the fix detailed in Section 3 below. The inputs were

- $D = 50$

- $M_0 = 50$

- $N = 101$

- $\rho_1 = 1/2, \rho_2 = 1/3$

Surprisingly, this resulted in a "matrix is numerically singular" error, even though this did not occur in the tests (described above) with larger ρ values. Inspecting the values of the hypergeometric function, I noticed some as small as 10^{-40}. These tiny values trickled into the linear system, where I saw entries as small as 10^{-20}. This, of course, is probably what caused the numerics to fail. The question is: where are these tiny values coming from?

Another question is this: if I observed even tinier values when ρ was closer to 1, why didn't I get a "matrix is numerically singular" error then? This could simply be randomness, since the algorithm fails in both cases anyway.

September 3 2019

Today, I attempted to use the fix detailed in Section 4 below. The inputs were

- $D = 50$

- $M_0 = 50$

- $N = 101$

- $\rho_1 = 4999/1000, \rho_2 = 49/50$

And the matrix I used to map out of the standard fundamental domain was (the corresponding disk version of)

$$A = T^{20}$$

where T is the shift map $z \mapsto z + 1$.

The scaling worked beautifully here! All the hypergeometric function values had norm close to 1 (by "close," I mean the norm was between 10^{-1} and 10). We still have the issue that the polynomial in ρ that appears in the expression for c_n can be very small. I did notice values as small as 10^{-6} in the final linear system.

Did it work?

2.2 Scaling the Hypergeometric Function

The linear system defined by Equation 1 above could be given some fine-tuning before it can be run in the algorithm. This is due to the hypergeometric function, which will be huge for some values of s and $|n|$, and very tiny for others. This leads to the matrix being "numerically singular" (that is, close enough to singular that Python gets confused).

One possible method of dealing with this is to scale the hypergeometric function by premultiplying by a function of s and $|n|$. The asymptotics of the function are derived here.

Recall that the hypergeometric function $_2F_1$ is defined

$$_2F_1(a, b, c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$$

where $(q)_n$ is the Pochhammer symbol

$$(q)_n = \begin{cases} 1 & n = 0 \\ q(q+1)\cdots(q+n-1) & n \neq 0 \end{cases}$$

From this definition, we see that if $z \sim 0$, then $_2F_1(a, b, c; z) \sim 1$.

For our application, we will be using $z = \rho^2 \sim 1$. To get an approximate size for the hypergeometric function near $z = 1$, we use the transformation formula

$$_2F_1(a, b, c; z) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \, _2F_1(a, b, a+b+1-c; 1-z)$$

$$+ \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b} \, _2F_1(c-a, c-b, 1+c-a; 1-z)$$

Since $_2F_1(a, b, c; z) \sim 1$ when $z \sim 0$, this formula implies that

$$_2F_1(a, b, c; z) \sim \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b} \qquad \text{when } z \sim 1 \qquad (2)$$

Now let's move to our application. Recall that we are interested in the function

$$_2F_1(s, s+|n|, 1+|n|; \rho^2)$$

where $s = 1/2 + \nu$ (given that $\lambda = 1/4 - \nu^2$ is the eigenvalue of a Maass form), $n \in \mathbb{Z}$ is bounded by $|n| \leq M$ (M is how far we go out in the finite Fourier series approximation), and ρ is the magnitude of z (which we take to be near 1). If we plug these values into Equation 2, we get

$$_2F_1(s, s+|n|, 1+|n|; \rho^2) \sim \frac{\Gamma(1+|n|)\Gamma(1-2s)}{\Gamma(1+|n|-s)\Gamma(1-s)} + \frac{\Gamma(1+|n|)\Gamma(2s-1)}{\Gamma(s)\Gamma(s+|n|)}(1-\rho^2)^{1-2s} \qquad (3)$$

for $\rho \sim 1$. Next, we use Stirling's formula. This is stated in Iwaniec-Kowalski as follows:

$$\Gamma(s) = \left(\frac{2\pi}{s}\right)^{1/2}\left(\frac{s}{e}\right)^s\left(1 + O\left(\frac{1}{|s|}\right)\right)$$

This is valid in the sector $|\arg s| \leq \pi - \epsilon$ for any $\epsilon > 0$ (the implied constant depends on ϵ). Applying Stirling's formula to Equation 3 gives

$$_2F_1(s, s+|n|, 1+|n|; \rho^2) \sim \frac{(1+|n|)^{\frac{1}{2}+|n|}(1-2s)^{\frac{1}{2}-2s}}{(1+|n|-s)^{\frac{1}{2}+|n|-s}(1-s)^{\frac{1}{2}-s}} + \frac{(1+|n|)^{\frac{1}{2}+|n|}(2s-1)^{2s-\frac{3}{2}}}{s^{s-\frac{1}{2}}(s+|n|)^{s+|n|-\frac{1}{2}}}(1-\rho^2)^{1-2s} \quad (4)$$

I wrote code to test the accuracy of this statement, and it verified my claims.

To finish, we wish to apply Equation 4 to our disk model approach. Recall that a key step in deriving the linear system (1) was removing the dependence of a_n on ρ. Hence, our scaling is not allowed to depend on ρ. On the other hand, if we scale the $c_n(\rho)$'s with respect to s and n, the a_n's will be scaled accordingly. This will *not* break the derivation of Equation 1, and hence is a permissible scaling.

2.3 Why Scaling May be in Vain

If we take sample points at a fixed ρ and with the θ_j's defined above, the $\mathcal{F}$-pullbacks could conceivably be anywhere in $\mathcal{F}$. Since the Hecke groups always have a cusp, this means we can expect points with ρ anywhere from 0 to 1.

For points with $\rho \sim 1$, we would like to use Equation 4 to scale. For points with ρ closer to 0, we would like to use no scaling (since the Taylor expansion of the hypergeometric function suggests it will be close to 1). However, any choice of scaling must be used for *all* points.

3 First Idea to Fix Problems With Disk Method

The first idea we had to fix the problems described above is this: don't use ρ's close to 1. In other words, choose the radii of the two circles on which we choose our sample points to be relatively small, say $\rho_1, \rho_2 \leq 1/2$.

Here are the pro's:

- Since the Taylor series defining the hypergeometric function suggests that it is about 1 when ρ is close to 0, we require no scaling.

- Not having to worry about scaling should fix the "matrix is numerically singular" error.

- Given a point z on the circle of radius ρ, its pullback z^* to the standard fundamental domain for any Hecke group has $|z^*| \leq \rho$. See proof below.

And the con's:

- Choosing ρ values close to 1 was done for a reason; specifically, this will cause the sample points to touch a large variety of fundamental domains.

- If too many sample points have neighboring sample points in the same fundamental domain, their pullbacks will too close together.

- These considerations may still result in a "matrix is numerically singular" error (just for a different reason than before).

3.1 Proof of Claim About Pullbacks from Circle

First, recall that the Möbius transformation which sends the disk model to the upper half plane model is

$$z \mapsto i\frac{1+z}{1-z}$$

One possible representative for this map in $SL(2, \mathbb{R})$ is

$$A = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$$

Writing S_ρ for the circle of radius $0 < \rho < 1$ centered at the origin, we are interested in the image $A(S_\rho)$ in the upper half plane.

We claim that $A(S_\rho)$ is the circle with center $i(1+\rho^2)/(1-\rho^2)$ and radius $2\rho/(1-\rho^2)$. To see this, one may check that

- $A(\rho) = i(1+\rho)/(1-\rho)$

- $A(-\rho) = i(1-\rho)/(1+\rho)$

- $i(1+\rho^2)/(1-\rho^2)$ is the midpoint of the two image points just described

- $A^{-1}(2\rho/(1-\rho^2) + i(1+\rho^2)/(1-\rho^2))$ has modulus equal to ρ

These facts, together with the fact that Möbius transformations take circles to circles, imply that $A(S_\rho)$ is the circle we claimed it to be.

Next, for any $0 < \rho < \rho' < 1$, we have that the circle $A(S_\rho)$ is contained entirely within the interior of the disk whose boundary is $A(S_{\rho'})$. Proving this is a straightforward computation. Moreover, it is clear that the circles may described by a continuous function of ρ. So the circles grow continuously as ρ ranges from 0 to 1.

Finally, we conclude the proof by showing that the pullback of every point on $A(S_\rho)$ to the standard fundamental domain is in the interior of the disk whose boundary is $A(S_\rho)$. This will finish the proof by the previous paragraph.

SHOW THIS!!!

3.2 Results

9/21/2019

Running this method with both $\rho = 1/2$ and $\rho = 1/3$ still resulted in a "matrix is numerically singular" error. However, after inspecting the matrices involved, it appears that the issue is still a scaling issue. Specifically, some of the matrix entries are still very tiny while others are still very huge.

I will need to look into this.

4   Second Idea to Fix Problems With Disk Method

Another idea to fix the issues described above is this: replace the standard fundamental domain with another one. In particular, perhaps we could take a fundamental domain which is contained entirely outside of the closed disk of radius ρ for some ρ "close" to 1.

Here are the pros of this approach:

- We can return to using ρ values close to 1 for the sample points.

- Since all ρ values involved will be close to 1, we can use the scaling of the hypergeometric function introduced in Equation 4.

- Since all points will be using an appropriate scaling, perhaps this will fix the "matrix is numerically singular" error.

And the cons:

- Although all fundamental domains have the same hyperbolic area, this choice of fundamental domain will have a small Euclidean area.

- The hypergeometric function is continuous in each of its entries, so the lack of variety in ρ values could result in a lack of information, again causing a "matrix is numerically singular" error.