Complex Geometries and Aperiodicity

Ben Sinclair
Supervisor: John Hunton

July - September 2019

1 Introduction

This project brings together two distinct mathematical topics of recent interest to the mathematical community, *Aperiodic Tiling* and the mathematics underlying in the artwork of M.C. Escher, with particular focus on the lithograph *Print Gallery* (Figure 1) and Escher's Grid (Figure 3). This project will demonstrate that there is a way of representing uniquely hierarchical, aperiodic substitution tilings on variations of Escher's Grid, (explained further in Section 1.1).

Figure 1: *Print Gallery* by M.C. Escher

1.1 Summary of Research

The aim of this section is to provide a non-mathematical summary of this report.

A significant amount of time during this research project was devoted to a literature review of the article *The Mathematical Structure of Escher's Print Gallery* by B.de Smit and H.W. Lenstra Jr. [1]. Escher's Grid (Figure 3) is the grid that the lithograph *Print Gallery* was imposed on and this grid was carefully constructed (via trial and error) by Escher so that a square grid invariant under magnification by 256 (remains the same grid under magnification by 256) can be distorted in a way that matches Escher's Grid. In particular Escher wanted an anticlockwise loop where the start and end points differ by a magnification of 256 in the square grid to represent a single anticlockwise loop (i.e. the start and end points of this loop are the same) in Escher's Grid. Consequently, Escher's Grid gives a way of representing images that remain the same under a magnification of 256. The article [1] gives the mathematical formalisation of the construction of Escher's Grid.

The analysis given in [1] specifies how to mathematically map square grids that are invariant under magnification by 256 to Escher's Grid. However, this value 256 was chosen arbitrarily by Escher, and can be changed to some other value which we will call x. By following the mathematical analysis in [1], we can take square grids that are invariant under magnification by x and distort them to some variation of Escher's Grid. Code provided by Christopher Prior (Durham University), enabled me to map rectangular grids to distorted grids, allowing creation of some new distorted grids which can be found in Section 5 for various values of x. This exercise was a successful attempt to produce some distorted grids and has been a useful learning experience which will act as a platform to build on and develop, and support my next research period.

Now we will move on to the concept of tilings. A tiling can be described as a 2D plane, occupied by tiles (which can be any geometric shape), such that there are no gaps on the plane and there are no overlaps between the individual tiles (like a jigsaw).

Given a collection of tiles that are unique, we can create a tiling by the *inflate and subdivide rule.* This rule involves taking a tile from the collection of unique tiles and inflating it by a certain scale factor then trying to fill this inflated tile with the tiles in the unique collection of tiles such that are no gaps in the inflated tile and no tiles overlap with each other. The inflate and subdivide rule can be repeated, and the significance of this is described later.

The concept of strange loops utilised by Escher, can be shown in Figure 2. If we pick any point on the top layer of stairs in Figure 2, then travel up the stairs (i.e. moving up the hierarchy), one finds themselves returning to the point where they started. This demonstrates an example of a strange loop.

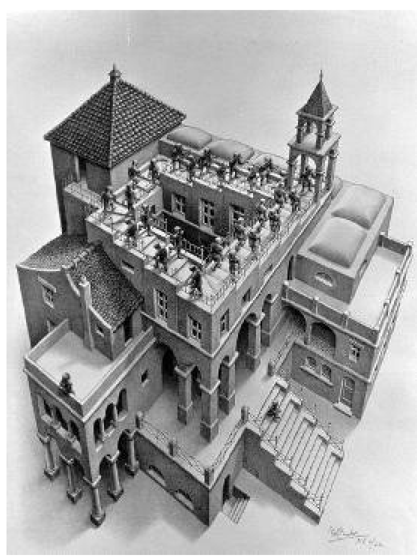

Figure 2: Ascending and Descending by M.C. Escher

A hierarchy level of a tiling can be described as the number of times the inflate and subdivide rule is applied to a tile.

In conclusion, this research includes taking tilings produced by the inflate and subdivide rule that are invariant under magnification by some value x and dividing these tilings into different hierarchy levels (a possible example of a divided tiling template is given in Figure 9) so that a strange loop between the tiling hierarchies is produced in Escher's Grid.

The tiling template produced in Figure 9 is constructed such that the anticlockwise loop (where the start and end points of this loop differ by a magnification of x) represents an increase in hierarchical levels from level 0 to level 3 and then returns to level 0. When mapping this anticlockwise loop (as above) to a variation of Escher's Grid, this loop now represents a single anticlockwise loop where the start and end points of the loop are the same. Since the start and end points are on the same level in the square grid (invariant under magnification by x) and since these points are the same point in the variation of Escher's Grid, this now produces a strange loop in the variation of Escher's Grid.

In summary, we have found a way to represent tilings with different hierarchy levels on variations of Escher's Grid. In the next research period, I will attempt to produce visual representations of this concept, based on my analysis.

2 The Complex Geometry of Escher's Grid

A large proportion of time during this research project was devoted to a literature review of the article *The Mathematical Structure of Escher's Print Gallery* by B.de Smit and H.W. Lenstra Jr. [1]. An additional step was to rewrite this article, as some areas required additional explanation in order to enhance the researcher's understanding of this topic. In particular, further understanding was needed with regards to the mathematical field of *Topology* (i.e. the fundamental group of a topological space), which required a brief reading of *Basic Topology* by M.A. Armstrong [2]. The rewritten article can be found in Appendix A.

To summarise the content of [1], a conformal bijection h (conformal as in each cell of the newly distorted grid maintains its square appearance) was constructed which mapped a distorted grid (namely Escher's Grid see Figure 3) that is invariant under multiplication by $\gamma \in \mathbb{C}^*$, to an undistorted grid (a standard square grid) that is invariant under scalar multiplication by 256. The bijection h represents how the print *Print Gallery* (Figure 1) is mapped to the undistorted grid and mathematically speaking, h is a map from $\mathbb{C}^*/<\gamma>$ to $\mathbb{C}^*/<256>$. The fundamental groups of $\mathbb{C}^*/<\gamma>$ and $\mathbb{C}^*/<256>$ are L_γ and L_{256} respectively where L_δ is defined as $L_\delta := \mathbb{Z}2\pi i + \mathbb{Z}log\delta$. The article then concludes:

- $h : \mathbb{C}^*/<\gamma> \to \mathbb{C}^*/<256>$ where $h(w) = w^\alpha$
- $\gamma = e^{\alpha^{-1}log256}$
- $\alpha = (2\pi i + log256)/2\pi i$ where α satisfies $\alpha L_\gamma = L_{256}$

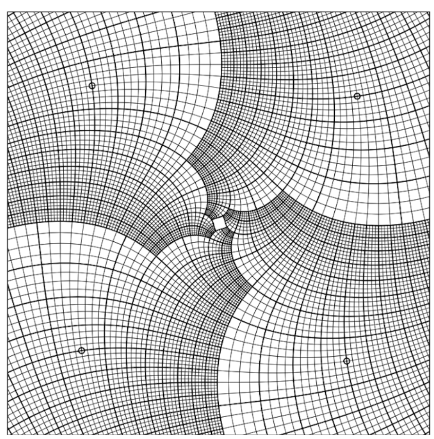

Figure 3: Escher's Grid

Remark 1. *The value $\alpha = (2\pi i + log256)/2\pi i$ is derived from the fact that Escher's procedure in constructing his distorted grid was to take an anticlockwise loop around the origin in the distorted grid and map this to an anticlockwise* *loop around the origin that is multiplied by 256. This represents the element $2\pi i \in L_\gamma$ being mapped to the element $2\pi i + log256 \in L_{256}$ and using the fact that $\alpha L_\gamma = L_{256}$ implies that $\alpha = (2\pi i + log256)/2\pi i$.*

2.1 Producing New Grids

In the article [1], the value 256 was chosen to represent a decision made by Escher in constructing his distorted grid. Namely, "as one goes clockwise around the center, the grid folds onto itself, but expanded by a factor of $4^4 = 256$" ([1] pg 447). However from a mathematical perspective, 256 was chosen arbitrarily. So one can ask if we replace 256 by some $x \in \mathbb{R}_{>0} \setminus \{1\}$, what happens to the bijection between the distorted grid (a variation of Escher's Grid) and the undistorted grid invariant under multiplication by x?

Remark 2. *$x = 1$ is not considered since the analysis given in* [1] *is not applicable for $x = 1$.*

So now instead of considering the conformal bijection h as given above, one should consider a conformal bijection:

$$h_x : \mathbb{C}^*/<\gamma_x> \to \mathbb{C}^*/<x> \tag{1}$$

Applying the same methodology as given in [1] pg 449-450 concludes that:

- $h_x(w) = w^{\alpha_x}$
- $\gamma_x = e^{\alpha_x^{-1} logx}$
- $\alpha_x = (2\pi i + logx)/2\pi i$ where α_x satisfies: $\alpha_x L_{\gamma_x} = L_x$

Remark 3. *The methodology in calculating h_x replicated a part of Escher's procedure as given in Remark 1, namely that the value α_x reflects the idea Escher wanted to map an anticlockwise loop in the distorted grid to an anticlockwise loop that is multiplied by x in the undistorted grid. So mathematically speaking, one can choose where the anticlockwise loop of L_{γ_x} in the distorted grid is mapped to in the undistorted grid's fundamental group L_x; however this has already been examined for $x = 256$, by the authors of* [1] *and the corresponding grids can be found on the website* [3].

To summarise, the bijection h_x is a function that maps a distorted grid that is invariant under multiplication by γ_x to an undistorted grid that is invaraint under multiplication by x. However, the researcher is interested in mapping undistorted images (in particular undistorted tilings) to distorted images. This implies that the map h_x^{-1} (h_x^{-1} exists since h_x is a bijection) must be used to map these undistorted images to distorted ones. One can show that:

$$h_x^{-1} : \mathbb{C}^*/<x> \to \mathbb{C}^*/<\gamma_x> \text{ where } h_x^{-1}(w) = w^{\alpha_x^{-1}} := e^{\alpha_x^{-1} logw}$$

A proof of this is given in Appendix B.

2.2 Mathematica Work

In order to produce any visual outputs for this project, the use of some mathematical software or programming language was required. One of the steps was to map undistorted square grids to distorted grids under conformal mappings. The researcher was provided code for the software *Mathematica* by Christopher Prior, Durham University which allowed this step to be achieved by taking the undistorted grid's nodes and mapping each individual node under a conformal map, then joining up the distorted nodes by using a straight line.

Initially the idea was to take a square, undistorted grid and map it to a distorted grid under the map h_x^{-1} for various x. However, the function h_x^{-1} includes a complex logarithm and for a complex logarithm to be holomorphic, it must be only defined away from some branch cut of the complex plane (the function must be holomorphic with a non-zero derivative for the function to also be conformal). In Mathematica, the complex logarithm *Log* is defined on the complex plane away from the negative real axis, so any undistorted grids must be defined away from the negative real axis. So consider the set:

$$A = \{z \in \mathbb{C} | (Re(z) \in [-1, 1/20] \text{ and } Im(z) \in [-1, -1/20] \cup [1/20, 1]) \text{ or } \\ (Re(z) \in [1/20, 1] \text{ and } Im(z) \in [-1, 1])\}$$

Then divide this set A into grid squares of length 1/20. This gives an undistorted grid (will refer to this undistorted grid as 'Grid A' from now on) and a rough sketch of this undistorted grid can be seen in Figure 4.

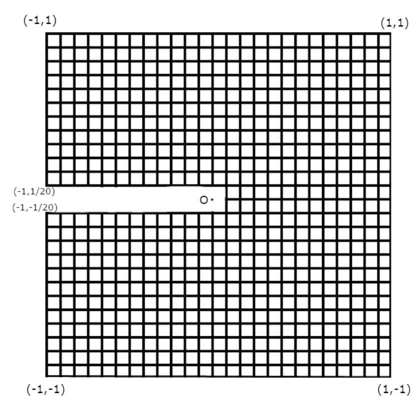

Figure 4: Sketch of Grid A

The researcher then observed how Grid A is mapped under the function h_x^{-1} for various x. For example, when $x = 256$ the distorted grid produced is given in Figure 5.

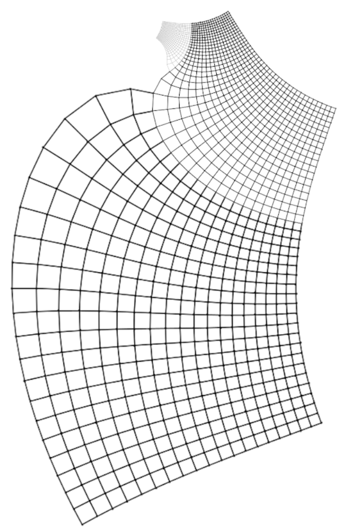

Figure 5: Grid A mapped under h_{256}^{-1}

Other values of x were considered when mapping grid A under h_x^{-1} and a collection of these distorted grids can be found in Section 5 of this report.

3 Aperiodic Substitution Tilings

This section is dedicated to providing definitions related to aperiodic, substitution tilings, so that the reader is familiar with the terminology used in the subsequent section of this report (Section 4). *Note: only tilings that cover* $\mathbb{R}^2$ *(i.e. the 2-D Euclidean plane), will be considered in this report.*

Definition 1. *A tiling T is a subdivision of the plane $\mathbb{R}^2$ into pieces called tiles. Specifically, we have a set of tiles t_i that intersect only on their boundaries, and whose union is the entire plane* ([4] pg 1).

Definition 2. *A prototile set P is a finite set of inequivalent tiles* ([5] pg 299).

For example, one can tile the plane $\mathbb{R}^2 \cong \mathbb{C}$ by using black and white unit squares as prototiles to give a checkerboard tiling (Figure 6). Figure 6 is a tiling since the black and white tiles only intersect on their boundaries.

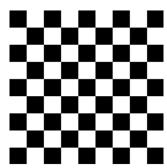

Figure 6: Checkerboard Tiling

Definition 3. *Let $t \in P$ where P is a set of prototiles. A linear transformation σ (that is not the identity map) is a substitution map if $\sigma(t)$ is a union of tiles belonging to the prototile set P. If $\sigma(t)$ is similar to t (i.e. same geometric shape but a different size), then σ can be represented as $\sigma(z) = \lambda z$ where λ is defined as the expansion factor of the prototiles and any tiling T generated by this substitution map σ is called a self-similar tiling.*

One can conclude that $|\lambda| > 1$ since if $|\lambda| < 1$, this would show that $\sigma(t)$ has a smaller area than t. This implies that $\sigma(t)$ cannot be the union of prototiles and hence σ is not a substitution map. ($|\lambda| \neq 1$ since σ is not the identity map).

Substitution maps are commonly refered to as the 'inflate and subdivide rule' since the process involves taking a tile, inflating this tile by the expansion factor λ, and then subdividing this inflated tile into its corresponding prototiles. For instance, Figure 7 gives the substitution map σ for the chair tiling (each prototile has unit length). This substitution rule has an expansion factor of $\lambda = 2$ since for all $t \in P$, $\sigma(t)$ is inflated by a scale factor of 2.

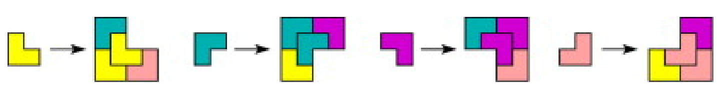

Figure 7: Substitution Map for Chair Tiling ([5] pg 301)

Definition 4. *A patch of a tiling is a finite subset of tiles in a tiling* ([4] pg 1)

Definition 5. *A patch of a tiling corresponding to $\sigma^n(t_i)$ (for some $t_i \in P$ and for some substitution map σ) is called a supertile of level n and type i* ([4] pg 11)*.*

For example, Figure 8 shows the supertiles level 2,3,4 of the yellow prototile with respect to the substitution map given in Figure 7.

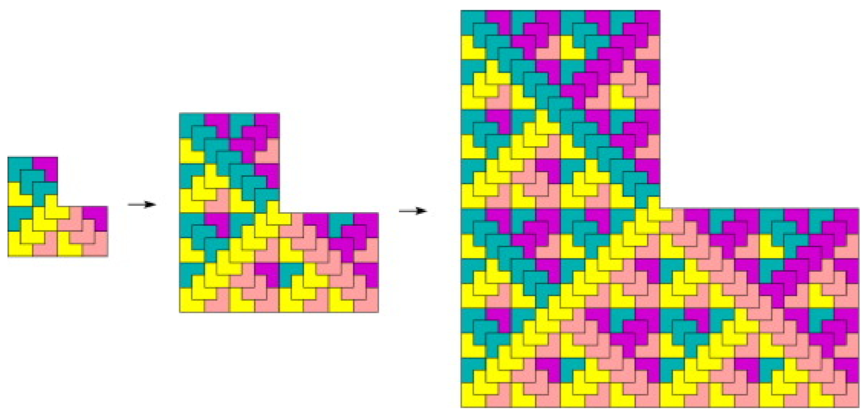

Figure 8: Chair Tiling Substitution: Supertile of the yellow prototile (of Figure 7) level 2 (Left), level 3 (Middle) and level 4 (Right) ([5] pg 302)

In Section 4, the researcher considers self-similar tilings T that are fixed points of some similar substitution map σ, that is $\sigma(T) = T$. To construct these tilings using σ such that the above condition is satisfied, an origin must be defined on the plane $\mathbb{R}^2$ such that a set of prototiles P joined together and $\sigma(P)$ agree on the interior and boundary of P and the boundary of $\sigma(P)$ is further away from the origin than the boundary of P in all directions. By repeating this, we have that $\sigma^n(P)$ and $\sigma(\sigma^n(P))$ agree on the interior and boundary of $\sigma^n(P)$ and the boundary of $\sigma(\sigma^n(P))$ is further away from the origin than the boundary of $\sigma^n(P)$ in all directions. Now the tiling T can be defined as:

$$T = \lim_{n \to \infty} \sigma^n(P)$$

and it immediately follows that $\sigma(T) = T$.

4 Strange Loops and Tiling Hierarchies

To discuss how certain aperiodic substitution tilings can be represented on variations of Escher's Grid, one can discuss the phenomenon of a *Strange Loop*. A Strange Loop can be described as when "moving upwards (or downwards) through the levels of some hierarchical system, we unexpectedly find ourselves right back where we started" ([6] pg 10). A good example of this is shown in Print Gallery (Figure 1). "What we see is a picture gallery where a young man is standing, looking at a picture of a ship in the harbor of a small town, ... upon one of [the buildings] ... sits a boy, ... while two floors below him is a woman ... [who] sits directly above a picture gallery where a young man is standing, looking at a picture of a ship in the harbor of a small town" ([6] pg 715). The start and end points are on the same level, despite travelling through increasing levels of the hierarchy.

Remark 4. *A hierarchy of a substitution map can be defined based on the level of a supertile. For example, a supertile of level $n+1$ is one level higher on the hierarchy than a supertile of level n.*

Since a Strange Loop is produced in Figure 1, the underlying complex geometry of Escher's Grid (Figure 3) has the potential to allow an undistorted image which contains multiple hierarchy levels to be mapped to and produce a Strange Loop on Escher's Grid. So one can ask, is there a way to divide a substitution tiling into different hierarchy levels such that a Strange Loop between these hieraracby levels is produced in Escher's Grid (or some variation of Escher's Grid)?

This question was first answered empirically by McCoyd for Escher's Grid [7], however based on the analysis of Escher's Grid (and variations of Escher's Grid) in Section 2, there is a more formal approach to answer this question.

4.1 Constructing a Strange Loop of Tiling Hierarchies

Recall from Section 2.1 that the bijection h_x^{-1} maps an undistorted grid invariant under scalar multiplication by x to a distorted grid (a variation of Escher's Grid) invariant under complex multiplication by γ_x. So for an undistorted self-similar substitution tiling T produced by a self-similar substitution map σ to be mapped to a variation of Escher's Grid, it is required that T is a fixed point of σ ($\sigma(T) = T$) and that the expansion factor of σ is $\lambda = x$.

Next, one must figure out how to divide T so that different hierarchy levels are displayed on this tiling. By Remark 4, this implies that some tiles in T are replaced by their corresponding supertiles level n scaled down by factor x^n for some $n \in \mathbb{N}$ (this fully replaces the original tile since σ is a self-similar map). A *supertile line* can be defined as a line on a tiling that follows the boundaries of the tiles such that when crossing this line, tiles are replaced with their corresponding supertiles level n for some $n \in \mathbb{N}$. The supertile line is defined to follow the boundaries of the tiles to preserve the overall tiling structure. For example, given a self similar substitution tiling generated by square prototiles of equal length, Figure 9 shows how supertile lines may be constructed on the tiling (note that the supertile lines will follow the boundaries of the square tiles).

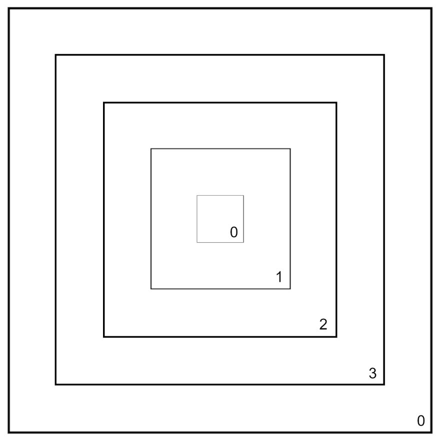

Figure 9: Supertile lines constructed on a tiling genereated by square prototiles (numbers represent the level of supertile)

The remainder of this report will consider a self-similar square tiling T generated by a substitution map σ with square prototiles of equal length and invariant under mulitiplication by x. Figure 9 gives a possibility of the levels of the supertiles in each region. So for example, each tile in the region labelled 2 is replaced with its corresponding supertile level 2 scaled down by a factor of x^2. In this case, when the supertile line labelled 3 is crossed moving away from the origin, the levels of the hierarchy are reset and the next region contains the original tiles of the tiling (supertiles level 0) and the process is repeated (the next supertile line will be labelled 1, then 2, then 3, then 0). It is a requirement that both supertile lines labelled 0 in Figure 9 are similar to each other by a scale factor of x.

The method above has constructed an undistorted tiling with multiple hierarchy that is invariant under scalar multiplication by x. So how is this tiling mapped under the function h_x^{-1} to a variation of Escher's Grid? Recall from Section 2.1 that the map h_x^{-1} takes a single anticlockwise loop around the origin that is multiplied by x in the undistorted grid and maps it to a single anticlockwise loop around the origin in the variation of Escher's Grid. This means the end point of the loop in the undistorted grid is a multiplication of the start point of the loop by x.

Therefore if this loop in the undistorted grid starts at a point given by a complex number on the inner supertile line labelled 0 of Figure 9, then as this loop travels anticlockwise around the origin, the distance of the points on the loop from the origin increases. This implies that this loop crosses all of the supertile lines 1 to 3 and since the outer supertile line 0 is similar to the inner supertile line 0 by a scale factor of x, this means that this loop will end at a point on the outer supertile line 0 of Figure 9.

Under the map h_x^{-1}, this loop is mapped to a single anticlockwise loop around the origin on a variation of Escher's Grid. Following this loop on the variation of Escher's Grid as described above, the loop starts at a point on the supertile line 0 and crosses the supertile lines 1 to 3 as one travels anticlockwise around this loop. However, since this loop is now a single anticlockwise loop on the variation of Escher's Grid, the loop returns to the intial starting point of this loop on the supertile line 0. When crossing the supertile lines on the variation of Escher's Grid, this represents an increase in the hierarchy of the tiles on the tiling. Consequently, if one keeps travelling anticlockwise around this loop multiple times, the hierarchy levels of the tiles change in the following way:

$$0 \to 1 \to 2 \to 3 \to 0 \to 1 \to 2 \to 3 \to 0$$

and so on. This demonstrates that a *Strange Loop* has been produced since as one travels up this hierarchy from level 0, we eventually return to level 0.

Therefore Figure 9 gives a possibility of how to divide an undistorted square tiling into different hierarchy levels so that a Strange Loop is produced in a variation of Escher's Grid. *Note: Other geometric shape prototiles that produce self-similar tilings will work if the supertile lines of Figure 9 are adjusted accordingly to match the boundaries of the tiles in the tiling.*

5 Gallery

This section provides a range of distorted grids produced by mapping Grid A (see Section 2.2) under the map h_x^{-1} for various x. The methodology for producing these grids is given in Section 2.2.

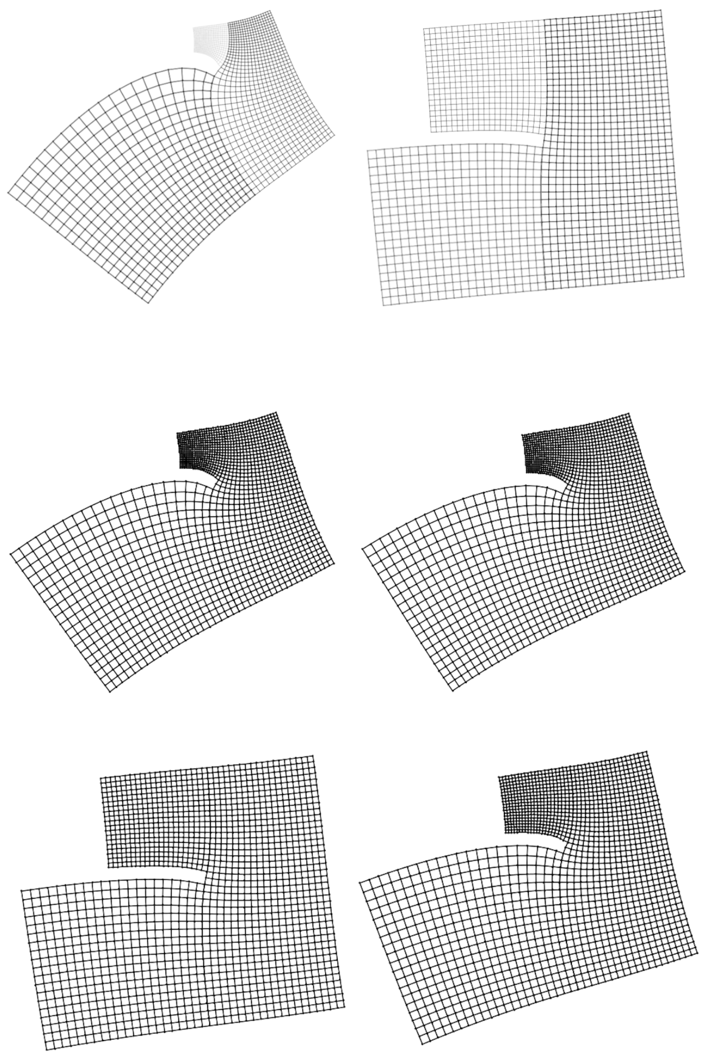

Figure 10: Top left: $x = 16$, Top right: $x = (1+\sqrt{5})/2$, Middle left: $x = 8$, Middle right: $x = ((1+\sqrt{5})/2)^4$, Bottom left: $x = 2$, Bottom right: $x = 4$

6 Further Research

Below is a list of points that will be addressed during the next research period:

- The number of supertile lines is not fully arbitrary. This number is dependent on boundary lines of the tiles in the tiling (which can vary). The researcher will investigate a division of a self-similar substitution tiling with different numbers of supertile lines.
- The researcher will observe how these supertile lines are mapped under the function h_x^{-1} and subsequently, create tilings which contain different hierarchy levels and map these tilings under the map h_x^{-1} to produce a visual representation of a Strange Loop on a variation of Escher's Grid. This will be achieved by developing the Mathematica code the researcher was provided to create the distorted grids given in Section 5.

References

[1] B. de Smit and H.W. Lenstra Jr., The Mathematical Structure of Escher's Print Gallery, Notices of the AMS, (50), 2003, 446-451

[2] M.A. Armstrong, Basic Topology, McGraw-Hill, Berkshire, 1979, 87-96

[3] B. de Smit and H.W. Lenstra Jr., 2003, Escher and the Droste effect [online]. Available from: `http://escherdroste.math.leidenuniv.nl./index.php?menu=im&sub=grid` [Accessed 10.09.19]

[4] Lorenzo Sadun, Topology of Tiling Spaces (University Lecture Series Volume 46), American Mathematical Society, Rhode Island, 2008

[5] Natalie Priebe Frank, A primer of substitution tilings of the Euclidean plane, Expositiones Mathematicae, Volume 26, Issue 4, 2008, 295-326, [online]. Available from: `https://www.sciencedirect.com/science/article/pii/S0723086908000042` [Accessed 12.09.19]

[6] Douglas R. Hofstadter, Gödel, Escher, Bach: an Eternal Golden Braid, Basic Books Inc., 1979

[7] G. McCoyd, Tiles and Tilings with the Visual Arts, MMath Dissertation, University of Leicester, 2005

Appendix A - The Mathematical Structure of Escher's Print Gallery

B. de Smit and H.W. Lenstra [1] (Edited by Ben Sinclair)

April 2003 (Edited August 2019)

1 Escher's Method

The approach Escher used to produce the print *Pretententoonstilling* was to take an undistorted drawing that is invariant under scalar multiplication by 256 and map this drawing to a grid that the anticlockwise closed loop ABCDA in the new distorted grid (shown in Figure 1) corresponds to an invariance of the undistorted drawing under increased scaling of 256. The undistorted drawing can be represented by the function:

$$f : \mathbb{C} \to \{Black, White\} \quad s.t. \quad f(z) = f(256z) \tag{1}$$

where $f(z)$ represents the colour of the grid. However, on further observation of Figure 1, we see that if the square ABCD is rotated clockwise over an angle of about 160 degrees and shrunk by a factor of almost 20, it will coincide with the small central square [1]. Hence we would require that colouring of the square ABCD would match with the colouring of the small central square. Define $\gamma = P/P'$ where $P \in$ Square ABCD and $P' \in$ Small Central Square, and define:

$$g_0 : Small\ Central\ Square \to \{Black, White\} \quad s.t. \quad g_0(\gamma P') = g_0(P') \tag{2}$$

where $g_0(w)$ is the colour of the distorted (Escher's) grid. Note the condition $g_0(\gamma P') = g_0(P') \iff Small\ Central\ Square\ is\ the\ same\ colouring\ as\ Square\ ABCD$

2 A Mathematical formalisation of Escher's Method

On Escher's print *Pretententoonstilling*, there is a hole in the centre of the print. To create an idealised version of Escher's grid, we can extend the function (2) to the set $\mathbb{C}^*$ giving the function:

$$g : \mathbb{C}^* \to \{Black, White\} \quad s.t. \quad g(\gamma w) = g(w) \tag{3}$$

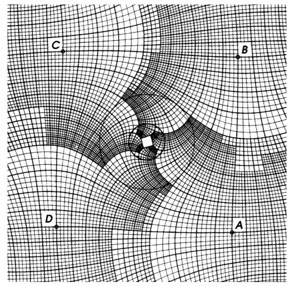

Figure 1: Escher's Grid

f is periodic under scalar multiplication of 256, so if we consider how f acts on the annulus $\{z \in \mathbb{C}^* \mid 1 \leq |z| \leq 256\}$ then we know how f acts on all of $\mathbb{C}^*$. So f can be replaced with the induced function:

$$\bar{f} : \mathbb{C}^*/{<}256{>} \quad \rightarrow \quad \{Black, White\} \tag{4}$$

where ${<}256{>}$ is the normal subgroup generated by 256 and $\mathbb{C}^*/{<}256{>}$ is the set of equivalence classes w.r.t. to the coset ${<}256{>}$. By a similar justification, g can be replaced by the induced function:

$$\bar{g} : \mathbb{C}^*/{<}\gamma{>} \quad \rightarrow \quad \{Black, White\} \tag{5}$$

So we can represent $\mathbb{C}^*/{<}256{>}$ as the undistorted picture and $\mathbb{C}^*/{<}\gamma{>}$ as the distorted picture.

Aside for later: We can visualise the set $\mathbb{C}^*/{<}\delta{>}$ *where* $|\delta| = 256$ *or* $|\gamma|$ *as a torus since the outer circle of the annulus* $\{z \in \mathbb{C}^* \mid 1 \leq |z| \leq |\delta|\}$ *is just a scaled and rotated version of the inner circle under* f *and* g *and therefore these circles can be 'glued' together to form a torus.*

Now we must ask how are f and g related to each other? Let's say we are given a colouring of $\mathbb{C}^*/{<}256{>}$, (i.e we know how $\mathbb{C}^*/{<}256{>}$ is mapped under $\bar{f}$) so if we consider a bijection:

$$h : \mathbb{C}^*/{<}\gamma{>} \quad \rightarrow \quad \mathbb{C}^*/{<}256{>} \tag{6}$$

then we know how to assign a colour to all elements of Escher's grid by means of the composition $\bar{g} = \bar{f} \circ h$ (all elements in Escher's grid are mapped to some point in the undistorted grid by h and since we know how to colour the undistorted grid, we can map these points to a colour by $\bar{f}$).

When Escher was constructing his grid, he wanted the original squares to retain the square appearence when constructing his distorted grid. Formally speaking, Escher wished for h to be a *conformal* map.

2.1 The Fundamental Group

To further understand the mathematics behind Escher's grid, we need to know what we mean by the *fundamental group* of a topological space. *(Note: a lot of the definitions in this section are informal, for more formal definitions see Chapter 5 of [2]).*

Let X be a topological space, let $p \in X$ and let $\alpha, \beta : [0,1] \to X$ be two loops based at p, that is; $\alpha(0) = \alpha(1) = \beta(0) = \beta(1) = p$. The multiplication of these two loops $\alpha \cdot \beta$ is defined as first traversing β followed by traversing α starting and ending at $p \in X$.

We say that α and β are *homotopic* if either loop can be continuously deformed into the other loop.

We know from *Lemma 5.3* of [2] that if the two loops α and β are homotopic, then $\alpha \sim \beta$ (defines an equivalence relation) and we can define the *homotopy class based at* p denoted $<\alpha>$ as the set of all loops in X that are equivalent to α (defined by the above equivalence relation). For example, we have $\beta \in <\alpha>$.

The *Fundamental Group* of a topological space X based at $p \in X$ is defined as the set of all homotopy classes based at p and this forms a group under the operation $< \alpha > \cdot < \beta > = < \alpha \cdot \beta >$ (*Theorem 5.5* of [2]) and the fundamental group of X based at $p \in X$ is denoted: $\pi_1(X, p)$

Fact: If X is path-connected, then $\pi_1(X,p) \cong \pi_1(X,q) \quad \forall\ p, q \in X$ so the fundamental group is independent (up to isomorphism) of the base point in X so we instead write the fundamental group of a path-connected set X as $\pi_1(X)$.

For the rest of this article, X will either be $\mathbb{C}^*$ or some set of equivalence classes of $\mathbb{C}^*$ which are path-connected sets, so we no longer refer to the base point of X when defining its fundamental group.

2.2 Determining the map h

Consider the set $\mathbb{C}^*/<\delta>$ where $\delta \in \mathbb{C}^*$ *and* $|\delta| \neq 1$. The exponential map $\mathbb{C} \to \mathbb{C}^*$, $z \mapsto e^z$ induces a surjective conformal map $\mathbb{C} \to \mathbb{C}^*/<\delta>$ that identifies $\mathbb{C}$ with the *universal covering space* of $\mathbb{C}^*/<\delta>$ (conformal since the derivative of e^z is not equal to 0 $\quad \forall z \in \mathbb{C}$).

Aside (Loose definition): The universal covering space of X is Y if $\exists\ f : Y \to X$ s.t. f is a surjective map and Y is simply connected.

We have that the kernel of this induced exponential map is $L_\delta = \mathbb{Z}2\pi i + \mathbb{Z}log\delta$. We know that the fundamental group of a torus is $\mathbb{Z}$ x $\mathbb{Z}$ and hence can produce group isomorphism from $\mathbb{Z}$ x $\mathbb{Z}$ to L_δ (*Since* $|\delta| \neq 1$ *this implies that the* *real part of* $log\delta = log|\delta| + iarg\delta$ *is not equal to 0. Hence this means that the generators* $2\pi i$ *and* $log\delta$ *form a basis over* $\mathbb{C}$ *which means that* $\mathbb{Z}$ x $\mathbb{Z}$ *and* L_δ *are isomorphic*). Using this and the fact we can visualise $\mathbb{C}^*/<\delta>$ as a torus, this means we can identify L_δ as the fundamental group of $\mathbb{C}^*/<\delta>$. We also recognize that $\mathbb{C}^*/<\delta>$ is a thinly disguised version of the elliptic curve $\mathbb{C}/L_\delta$.

With this information, we can decide what the map h can be. Choosing coordinates properly, we may assume that $h(1) = 1$ (*that is* $h(<\gamma>) = <256>$). By algebraic topology, the map $h : \mathbb{C}^*/<\gamma> \to \mathbb{C}^*/<256>$ lifts to a unique conformal isomorphism $\varphi : \mathbb{C} \to \mathbb{C}$ (a map of universal covers) that maps 0 to 0 and induces an isomorphism $\mathbb{C}/L_\gamma \to \mathbb{C}/L_{256}$. Since we now have:

$$\mathbb{C} \xrightarrow{exp} \mathbb{C}^*/<\gamma> \xrightarrow{h} \mathbb{C}^*/<256> \xrightarrow{log} \mathbb{C} \qquad (7)$$

and since $\varphi : \mathbb{C} \to \mathbb{C}$, we can see that φ can be rewritten as: $\varphi = log \circ h \circ exp$. Let $z \in L_\gamma \subset \mathbb{C}$. Since L_γ is the kernel of the induced exponential map, we have: $\varphi(z) = (log \circ h \circ exp)(z) = (log \circ h)(<\gamma>) = log(<256>) \in L_{256} \subset \mathbb{C}$.

Hence we have that $\varphi(L_\gamma) = L_{256}$.

Since h induces an isomorphism $\mathbb{C}/L_\gamma \to \mathbb{C}/L_{256}$ using *Theorem 4.1 Chapter 6* of [3], this implies that the map φ is just a multiplication by some complex number $\alpha \in \mathbb{C}^*$. Hence $\alpha L_\gamma = L_{256}$ for some $\alpha \in \mathbb{C}^*$.

Now we attempt to determine α. We use the result that L_γ and L_{256} are the fundamental groups of $\mathbb{C}^*/<\gamma>$ and $\mathbb{C}^*/<256>$ respectively and that $\alpha L_\gamma = L_{256}$. The element $2\pi i$ in the fundamental group L_γ corresponds to one anticlockwise loop around in the origin in $\mathbb{C}^*/<\gamma>$ and the path ABCDA (see figure 1) corresponds to the same element $2\pi i$ in the fundamental group (since up to homotopy, the path ABCDA is equivalent to this anticlockwise loop in L_γ).

Escher (in theory) wanted to transform this single anticlockwise loop to correspond to a path in the undistorted picture that goes once around the origin and scaled up by a factor 256. Hence this implies that the map α maps $2\pi i \mapsto 2\pi i + log256$ and since $\alpha L_\gamma = L_{256}$, we have that $\alpha \cdot (2\pi i) = 2\pi i + log256 \iff \alpha = (2\pi i + log256)/(2\pi i)$.

Since $L_\gamma = \alpha^{-1} L_{256}$ and $|\gamma| > 1$, we deduce that $\gamma = e^{2\pi i log256/(2\pi i + log256)} = e^{\alpha^{-1} log256}$.

Since $\varphi = log \circ h \circ exp$, this implies that $h(w) = (exp \circ \varphi \circ log)(w) = exp(\alpha log w) = w^\alpha$.

References

[1] B.de Smit and H.W. Lenstra Jr, *The Mathematical Structure of Escher's Print Gallery*, Notices of the AMS, (50), 2003, 446-451

[2] M.A. Armstrong, *Basic Topology*, McGraw-Hill, Berkshire, 1979

[3] Joseph H. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1986

Appendix B - Proof of h_x^{-1}

Remark. *For the purposes of this appendix, the functions* h_x *and* h_x^{-1} *will be written instead as* h *and* h^{-1} *respectively. In addition,* α_x *and* γ_x *will be written instead as* α *and* γ *respectively.*

Claim 1. $log(e^z) = z + 2\pi ik$ *for some* $k \in \mathbb{Z},\ z \in \mathbb{C}$

Proof. The definition of the complex logarithm is: $log(e^z) = log|e^z| + iarg(e^z)$ where arg is an arbitrary choice of argument.

Since $|e^z| = e^{Re(z)}$ and $arg(e^z) = arg(e^{Re(z)}e^{iIm(z)}) = arg(e^{iIm(z)}) = Im(z) + 2\pi k$ (where $k \in \mathbb{Z}$ is dependent on the choice of argument), $log(e^z)$ is now given by:

$log(e^z) = log(e^{Re(z)}) + i(Im(z) + 2\pi k) = Re(z) + iIm(z) + 2\pi ik = z + 2\pi ik$ □

Claim 2. $\alpha = 2\pi i/(2\pi i(n+1) - log\gamma)$ *for some* $n \in \mathbb{Z}$

Proof. Since $\gamma = e^{a^{-1}logx}$ and by using Claim 1 gives:

$$log\gamma = \alpha^{-1}logx + 2\pi in \text{ for some } n \in \mathbb{Z}$$

Rearranging this equation for $logx$ gives:

$$logx = \alpha(log\gamma - 2\pi in) \tag{1}$$

Since $\alpha = (2\pi i + logx)/2\pi i$, by substituting 1 into this expression gives:

$$\alpha = (2\pi i + \alpha(log\gamma - 2\pi in))/2\pi i$$

By rearranging the above equation for α, this gives the result stated in Claim 2. □

So now there are two expressions for α, namely $\alpha = (2\pi i + logx)/2\pi i$ and $\alpha = 2\pi i/(2\pi i(n+1) - log\gamma)$ for some $n \in \mathbb{Z}$.

Claim 3. $h^{-1}(w) = e^{\alpha^{-1} log w}$ *is the inverse function of* $h(w) = e^{\alpha log w}$.

Proof. The inverse of h is a bijection from $\mathbb{C}^*/<x> \to \mathbb{C}^*/<\gamma>$, which means that equivalence classes of complex numbers must be considered under the maps $h \circ h^{-1}$ and $h^{-1} \circ h$.

Consider the complex number $w \cdot x^m$ for some $w \in \mathbb{C}^*$ and $m \in \mathbb{Z}$. This complex number belongs to the equivalence class $w<x>$. Hence:

$$h \circ h^{-1}(w<x>) = h \circ h^{-1}(w \cdot x^m) = h(e^{\alpha^{-1} log(w \cdot x^m)}) = e^{\alpha log(e^{\alpha^{-1} log(w \cdot x^m)})}$$

Using Claim 1, the above expression for $h \circ h^{-1}$ becomes:

$$h \circ h^{-1}(w<x>) = e^{\alpha(\alpha^{-1} log(w \cdot x^m) + 2\pi i k)} = e^{log(w \cdot x^m)} e^{2\pi \alpha k i} = w \cdot x^m e^{2\pi \alpha k i}$$

for some $k \in \mathbb{Z}$.

Substituting in $\alpha = (2\pi i + log x)/2\pi i$ gives:

$$h \circ h^{-1}(w<x>) = w \cdot x^m e^{2\pi k i((2\pi i + log x)/2\pi i)} = w \cdot x^m e^{k(2\pi i + log x)}$$

Since $k \in \mathbb{Z}$, this implies that $e^{2\pi i k} = 1$ and the above becomes:

$$h \circ h^{-1}(w<x>) = w \cdot x^m e^{2\pi i k} e^{k log x} = w \cdot x^m x^k = w \cdot x^{m+k}$$

The complex number $w \cdot x^{m+k}$ belongs to the equivalence class $w<x>$. Hence, the above shows that:

$$h \circ h^{-1}(w<x>) = w<x> \iff h \circ h^{-1} = Id \qquad (\dagger)$$

which means that h^{-1} is the right inverse to h.

Now consider the complex number $w\gamma^m$ for some $w \in \mathbb{C}^*$ and $m \in \mathbb{Z}$. This complex number belongs to the equivalence class $w<\gamma>$. Hence:

$$h^{-1} \circ h(w<\gamma>) = h^{-1} \circ h(w\gamma^m) = h^{-1}(e^{\alpha log(w\gamma^m)}) = e^{\alpha^{-1} log(e^{\alpha log(w\gamma^m)})}$$

Using Claim 1, the above expression for $h^{-1} \circ h$ becomes:

$$h^{-1} \circ h(w<\gamma>) = e^{\alpha^{-1}(\alpha log(w\gamma^m) + 2\pi i k)} = e^{log(w\gamma^m)} e^{2\pi \alpha^{-1} k i} = w\gamma^m e^{2\pi \alpha^{-1} k i}$$

for some $k \in \mathbb{Z}$.

Claim 2 implies that α^{-1} can be represented as $\alpha^{-1} = (2\pi i(n+1) - log\gamma)/2\pi i$ for some $n \in \mathbb{Z}$ and substituting α^{-1} into the above gives:

$$h^{-1} \circ h(w<\gamma>) = w\gamma^m e^{2\pi k i((2\pi i(n+1) - log\gamma)/2\pi i)} = w\gamma^m e^{k(2\pi i(n+1) - log\gamma)}$$

Since $k(n+1) \in \mathbb{Z}$, this implies that $e^{2\pi ik(n+1)} = 1$ and the above becomes:

$$h^{-1} \circ h(w\!<\!\gamma\!>) = w\gamma^m e^{2\pi ik(n+1)} e^{-klog\gamma} = w\gamma^m e^{-klog\gamma}$$

The definition of a complex power implies that $\gamma^m = e^{mlog\gamma}$ and substituting this into the above gives:

$$h^{-1} \circ h(w\!<\!\gamma\!>) = we^{mlog\gamma} e^{-klog\gamma} = we^{(m-k)log\gamma} = w\gamma^{m-k}$$

The complex number $w\gamma^{m-k}$ belongs to the equivalence class $w\!<\!\gamma\!>$. Hence, the above shows that:

$$h^{-1} \circ h(w\!<\!\gamma\!>) = w\!<\!\gamma\!> \Longleftrightarrow h^{-1} \circ h = Id \qquad (\dagger\dagger)$$

which means that h^{-1} is the left inverse to h.

Combining $\dagger$ and $\dagger\dagger$ shows that h^{-1} is both the left and right inverse of h which implies that h^{-1} is the inverse function of h. $\square$