Research Experience for Undergraduates
Research Reports

Indiana University, Bloomington

Summer 2014

Contents

Notes on Quantum Mechanics over a Finite Field **5**
John Gardiner

Conformal Compactification of $\mathbb{R}^{p,q}$ **19**
Lyndee M Govert

Removing noise from a coarse-grain signal **28**
Zachary Helbert

Configuration Space of Eight-Bar Linkages **43**
Kathryn Marsh

Constructing a Dynamical Boundary at Infinity for Hyperbolic Toral Automorphisms **60**
Aleksandra Niepla

Ends of algebraic surfaces **80**
Eamon Quinlan Gallego

Geometric Properties of Conformal Transformations on $\mathbb{R}^{p,q}$ **95**
Surya Raghavendran

Reaction Time in General Recognition Theory **97**
Aleina Wachtel

The CW-Complex of Translation Surfaces of Genus 2 with One Singularity **114**
Elizabeth Winkelman

Homotopy Type Theory and $\pi_4(S^3)$ **149**
Richard Wong

Preface

During the summer of 2014 ten students participated in the Research Experiences for Undergraduates program in Mathematics at Indiana University. The program ran for eight weeks, from June 7 through July 31, 2014. Eight faculty served as research advisers:

- Chris Connell worked with Eamon Quinlan Gallego (U. Glasgow)
- Michael Jolly worked with Zachary Helbert (East Tennessee State U.)
- Chris Judge worked with Kathryn Marsh (Purdue U.) and Elizabeth Winkelman (U. of Rochester)
- Matvei Libine worked with Lyndee Govert (Indiana U. North) and Surya Raghavendran (U. Texas)
- Gerardo Ortiz (Physics) worked with John Gardiner (Brigham Young U.)
- Amr Sabry (Informatics and Computing) worked with Richard Wong (Rutgers U.)
- James Townsend (Psychological and Brain Sciences) worked with Aleina Wachtel (Harvey Mudd College).
- Bryce Weaver worked with Aleksandra Niepla (Rutgers U.)

Following the introductory pizza party, students began meeting with their faculty mentors and continued to do so throughout the next eight weeks. The students also participated in a number of social events and educational opportunities and field trips.

Individual faculty gave talks throughout the program on their research, about two a week. Students also received LaTeX training from grad student Anne Carter in a series of workshops. Students also were given the opportunity to opt into a GRE and subject test preparation seminar. Other educational activities inclued tours of the library, Slocum puzzle collection at the Lilly Library and the IU cyclotron facility, and self guided tours of the art museum. Students presented their work to faculty mentors and their peers at various times. This culminated in their presentations both in poster form and in talks at the Summer Undergraduate Research conference which took place at the IUPUI campus in Indianapolis.

On the lighter side, students were treated to a reception by the graduate school as well as a fun filled trip to a local amusement park. They were also given the opportunity to enjoy a night of "laser tag" courtesy of Prof. Elizabeth Housworth.

The summer REU program required the help and support of many different groups and individuals to make it a success. We foremost thank the National Science Foundation for major financial support through the REU program through NSF grant DMS-1156515. Additional logistical support was provided

by the Department of Mathematics and our chair, Elizabeth Housworth. Kevin Pilgrim's guidance as the previous REU coordinator and PI was invaluable as was the assistance of co-PI Larry Moss. We thank the staff of the Department of Mathematics for their assistance, including Jeff Taylor and Zeke Henline for computer support, and most especially Mandie McCarty for coordinating the complex logistical arrangments (housing, paychecks, information packets, meal plans, frequent shopping for snacks). We thank Indiana graduate student Anne Carter for serving as our LaTeX consultant and for compiling this volume. Thanks to those faculty who served as mentors and those who gave lectures. Thanks to David Baxter of the Center for Exploration of Energy and Matter (nee IU cyclotron facility) for his personal tour of the cyclotron facility and lecture. Thanks to Andrew Rhoda for his tour of the Slocum puzzle collection.

Chris Connell
September, 2014

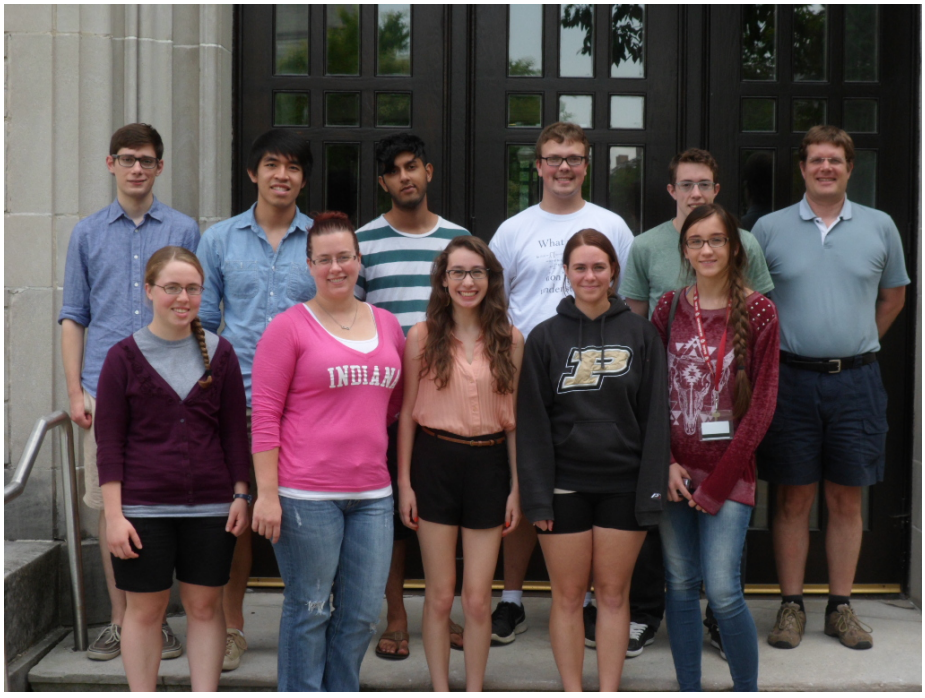


Figure 1: REU Participants, Row one (from left to right): Elizabeth Winkelman, Lyndee Govert, Aleina Wachtel, Kathryn Marsh, Aleksandra Niepla. Row two: John Gardiner, Richard Wong, Surya Raghavendran, Zachary Helbert, Eamon Quinlan Gallego, Chris Connell (coordinator).

Notes on Quantum Mechanics over a Finite Field

John Gardiner

Abstract

In the usual formulation of quantum mechanics the state of a system is described as a vector over the complex numbers. Replacing the field of complex numbers with a field of positive characteristic results in a family of toy theories with novel properties, which we call finite field quantum mechanics (FFQ). This work contains a discussion of some salient features of FFQ in particular with a focus on the meaning and structure of entanglement. On the study the structure of entanglement in FFQ we provide some preliminary results and highlight questions for further study.

1 Introduction

In the usual formulation of quantum mechanics the state of a system is described as a vector over the complex numbers. Various toy theories have been proposed that replace the field of complex numbers with other algebraic structures including the real numbers [10], the quaternions [3], and the subject of this work finite fields [9, 5].

One motivation for studying quantum mechanics over a finite field is related to some potential conceptual difficulties raised by the complex numbers [5]. Not all complex numbers are computable with finite resources. For nature to make fundamental use of complex numbers might seem incompatible with the premise that the universe is in some way a large computer. On the other hand, fields of positive characteristic are finite, making them attractive from a computational standpoint, yet they still have the same kind of algebraic structure as the complex numbers. Simply replacing complex numbers with finite field elements results in a toy theory with properties analogous to ordinary quantum theory. Different choices of finite field result in different theories. We call this family of toy theories finite field quantum mechanics (FFQ).

In this work only finite fields $\mathbb{F}_{p^2}$ where $p \equiv 3 \bmod 4$ are considered. The fields $\mathbb{F}_p$ where $p \equiv 3 \bmod 4$ are exactly those finite fields where $x^2 + 1$ has no root. Where i is a root of $x^2 + 1$ we can adjoin i to the field $\mathbb{F}_p$. The resulting field is isomorphic to $\mathbb{F}_{p^2}$ and has a sense of complex number. The map $\alpha \mapsto \alpha^p$ is analogous to complex conjugation and we write $\alpha^* = \alpha^p$. For vectors of finite dimension d, the conjugate transpose is defined in the natural way giving us the notion of dual vector $\langle\psi|$ for every vector $|\psi\rangle \in \mathbb{F}_{p^2}^d$ as well as unitary operators. For a review of the arithmetic of vectors over "complexified" finite fields see [5].

Among the first immediately noticeable features is that the Hermition dot product, the map from $\mathbb{F}_{p^2}^d \times \mathbb{F}_{p^2}^d$ to $\mathbb{F}_{p^2}$ given by $(|\psi\rangle, |\phi\rangle) \longmapsto \langle\phi|\psi\rangle$, is no longer an inner product as there are zero norm vectors other than the zero vector, i.e. vectors $|\psi\rangle \neq 0$ such that $\langle\psi|\psi\rangle = 0$. Also, without an ordering on the field elements the map from vectors to $\mathbb{F}_p$ given by $|\psi\rangle \longmapsto \langle\psi|\psi\rangle$ is neither a norm nor a seminorm. Many other familiar properties such as $\langle\psi|\phi\rangle = 1 \Longrightarrow |\psi\rangle = |\phi\rangle$ no longer hold. From the outset FFQ is seen to be very different from ordinary quantum mechanics.

So it is an abuse of terminology and notation, but throughout this work $\alpha^*\alpha = \alpha^{p+1}$ will be referred to as the norm of α and will be notated $|\alpha|^2$. In a similar abuse of terminology $\langle\phi|\psi\rangle$ will be referred to as the inner product of $|\psi\rangle$ and $|\phi\rangle$, and $\langle\psi|\psi\rangle^{p+1}$ will be called the norm of $|\psi\rangle$.

In ordinary quantum theory the Born rule gives a physical meaning to the state vectors. Given a state $|\psi\rangle$, a measurement of state $|\psi\rangle$ will result in one of some number of mutually exclusive outcomes. The probability that an outcome corresponding to a (normalized) state $|\phi\rangle$ will occur is given by $|\langle\phi|\psi\rangle|^2$. Via the Born rule the state vectors of ordinary quantum theory can be seen as holding the information regarding outcomes of measurements. It is worth asking whether the state vectors in FFQ contain information about measurement outcomes. When the Born rule is translated directly into the finite field framework the quantity $|\langle\phi|\psi\rangle|^2$ can no longer be interpreted as a probability as it is an element of the field $\mathbb{F}_p$ rather than a real number from 0 and 1. So to get a rule for extracting probabilities from the state vectors the Born rule must be either modified or replaced.

To be precise, we need a function that takes as input a state and an outcome then gives as an output a probability, namely the probability of a measurement on the input state resulting in that outcome. That is we seek a function $f\colon \mathbb{F}_{p^2}^d \times \mathbb{F}_{p^2}^d \longrightarrow [0,1]$. For the output of this new rule to truly be interpreted as probabilities the rule f must satisfy certain properties. First, as the outputs are probabilities they must satisfy a normalization requirement:

i.) For any orthonormal basis $\{|i\rangle\}$ and any state $|\psi\rangle$, $\sum_i f(|\psi\rangle, |i\rangle) = 1$.

That is, the sum of probabilities for a complete set of mutual exclusive outcomes must be 1. To motivate the next requirement consider the state $\alpha|0\rangle + \beta|1\rangle$ in the Hilbert space spanned by $|0\rangle$ and $|1\rangle$. We can embed this Hilbert space in the larger space spanned by $|0\rangle, |1\rangle, \ldots, |n\rangle$ and view our state in the larger Hilbert space, i.e. view $\alpha|0\rangle + \beta|1\rangle$ as $\alpha|0\rangle + \beta|1\rangle + 0|2\rangle + \cdots + 0|n\rangle$. The probabilities corresponding to outcomes $|0\rangle$ and $|1\rangle$ should not depend on whether we are viewing the state in the original two-dimensional Hilbert space or the larger one. So, in light of property (i), the probabilities corresponding to outcomes $|2\rangle, \ldots, |n\rangle$ should all be zero. This suggests the following necessary condition:

ii.) If $\langle\phi|\psi\rangle = 0$ then $f(|\psi\rangle, |\phi\rangle) = 0$.

The final property we require of a rule for deriving probabilities is that it be basis independent:

iii.) $f(U\left|\phi\right\rangle, U\left|\psi\right\rangle) = f(\left|\phi\right\rangle, \left|\psi\right\rangle)$ where U is any unitary.

In ordinary quantum theory, Gleason's theorem shows that properties similar to these are sufficient to uniquely identify the Born rule as the only consistent rule for extracting probabilities from states. Considering this it should not surprise us if properties (i), (ii), and (iii) constitute a serious constraint on the form of f. In fact, we have the following no-go theorem:

Theorem 1.1 *There is no map* $f\colon \mathbb{F}_{p^2}^d \times \mathbb{F}_{p^2}^d \longrightarrow [0,1]$ *satisfying properties (i), (ii), and (iii).*

Proof Assume such a map f satisfying all three properties exists. Consider the state

$$\left|\psi\right\rangle = \alpha\big(\left|0\right\rangle + \left|1\right\rangle + \left|2\right\rangle + \beta\left|3\right\rangle\big)$$

where $\beta \in \mathbb{F}_{p^2}$ is chosen so that $\left|\beta\right|^2 = p-1$ and $\alpha \in \mathbb{F}_{p^2}$ is a constant such that $\left|\psi\right\rangle$ is normalized. Let $\left|\phi\right\rangle = \left|1\right\rangle + \left|2\right\rangle + \beta\left|3\right\rangle$. Note that $\left|\phi\right\rangle$ is normalized. We can write our original state $\left|\psi\right\rangle$ as $\left|\psi\right\rangle = \alpha\big(\left|0\right\rangle + \left|\phi\right\rangle\big)$ The vectors $\left|0\right\rangle$ and $\left|\phi\right\rangle$ are orthonormal and so are part of some larger orthonormal basis. Thus by properties (i) and (ii) we have

$$1 = f\big(\left|\psi\right\rangle, \left|0\right\rangle\big) + f\big(\left|\psi\right\rangle, \left|\phi\right\rangle\big). \tag{1}$$

Let U be the unitary that simply permutes the orthonormal vectors $\left|0\right\rangle$ and $\left|\phi\right\rangle$. Note that $\left|\psi\right\rangle$ is invariant under U, so invoking property (iii) we have

$$f\big(\left|\psi\right\rangle, \left|0\right\rangle\big) = f\big(U\left|\psi\right\rangle, U\left|0\right\rangle\big) = f\big(\left|\psi\right\rangle, \left|\phi\right\rangle\big)$$

This along with equation (1) gives

$$1 = f\big(\left|\psi\right\rangle, \left|0\right\rangle\big) + f\big(\left|\psi\right\rangle, \left|0\right\rangle\big),$$

so

$$f\big(\left|\psi\right\rangle, \left|0\right\rangle\big) = f\big(\left|\psi\right\rangle, \left|\phi\right\rangle\big) = \frac{1}{2}. \tag{2}$$

For $i = 0, 1, 2$ let V_i be the unitary that simply permutes the basis vectors $\left|0\right\rangle$ and $\left|i\right\rangle$. Note that $\left|\psi\right\rangle$ is invariant under V_i. So again by property (iii) we have

$$f\big(\left|\psi\right\rangle, \left|i\right\rangle\big) = f\big(V_i\left|\psi\right\rangle, V_i\left|0\right\rangle\big) = f\big(\left|\psi\right\rangle, \left|0\right\rangle\big) = \frac{1}{2}$$

for $i = 0, 1, 2$. These probability assignments are inconsistent with the requirement that the probabilities for orthogonal outcomes add up to 1. By property (i)

$$\begin{aligned} 1 &= f\big(\left|\psi\right\rangle, \left|0\right\rangle\big) + f\big(\left|\psi\right\rangle, \left|1\right\rangle\big) + f\big(\left|\psi\right\rangle, \left|2\right\rangle\big) + f\big(\left|\psi\right\rangle, \left|3\right\rangle\big) \\ &= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + f\big(\left|\psi\right\rangle, \left|3\right\rangle\big) \\ &\geq \frac{3}{2}, \end{aligned}$$

a clear contradiction. The assumption of a satisfactory map f is thus false. $\square$ (It should be pointed out that the argument above explicitly uses states from a Hilbert space of four dimensions or larger so does not hold if the largest Hilbert space one can consider is only three-dimensionsal or smaller. But it does hold in any universe that contains more than a single qubit.)

What are the implications of this theorem? Given a state, there is no way of extracting probabilities from it that is independent of how we have written the state, i.e. independent of our choice of basis in which to write the state and independent of our choice of ambient Hilbert space in which to view the state. This justifies ruling out FFQ as a quantum theory that describes reality.

Intuitively one might think that the lack of an order on the field elements is the main obstacle to extracting probabiliities from FFQ, but the above result shows that this is not the case. Even supplementing the field elements with an ordering does not allow for a rule that satisfies the three properties above. And any other scheme for extracting probabilities from quantum mechanics over a finite field will violate one of the three properties above. In particular the cardinal probability scheme of [5] first maps $\mathbb{F}_{p^2}$ to $\mathbb{Z} + i\mathbb{Z}$, forgetting the field structure of the coefficients and thus disallowing most unitary transformations (changes of basis) other than permutations.

2 Bell's Theorem

The lack of a satisfactory way of assigning probabilities in FFQ has implications for many other results from ordinary quantum theory. The violation of Bell's inequalities in ordinary quantum theory is an important example of the failure of local realism as well as an example of entanglement as a resource. Bell's inequalities are, however, inequalities involving expectation values, and thus probabilities. In FFQ, where there is no concept of probability, it is worth asking whether Bell's theorem holds and whether entanglement can be viewed as a resource.

Of course, there are proofs of Bell's theorem that do not involve inequalities and we will consider this option in section 2.2. But it is instructive to first consider the status of Bell's inequalities in FFQ, as doing so will involve a discussion of the concepts of measurement and observable that will establish a way of talking about performing experiments in the FFQ framework and that may also shed light on some assumptions of ordinary quantum theory.

2.1 Observables in FFQ

In ordinary quantum mechanics Hermitian operators are identified with observables. In FFQ this is problematic. Hermitian operators in FFQ are not in general diagonalizable, and the eigenvalues of a Hermitian operator, when they do exist, are necessarily elements of $\mathbb{F}_{p^2}$. As measurement results need not have any algebraic structure, specifying that the results of measurements be eigenvalues of Hermitian operators, and thus field elements, does not make sense.

So in FFQ the usual identification of Hermitian operators and observables will not work, and whatever principle replaces it should be one that refrains from imposing an algebraic structure on the results of measurements. To elaborate, in ordinary quantum theory measurable properties such as spin, momentum, or position are always labeled with real numbers (sometimes with units) even when the property has an algebraic structure different from the reals, or perhaps no algebraic structure at all.

For instance position on a lattice with periodic boundary conditions has a structure like that of the integers modulo some number of lattice points. But in ordinary quantum theory these positions will necessarily be labeled with real numbers (the eigenvalues of the observable corresponding to position) which "forget" this periodic structure. Another example would be a quantum system that could be in one of two configurations, say "on" or "off". In ordinary quantum theory, an observable corresponding to the measurement of on-ness or off-ness would have two real numbers as eigenvalues, one corresponding to on and the other corresponding to off, even though the measurement results themselves, on and off, can't be added or multiplied like the real number eigenvalues which label them. So identifying observables with Hermitian operators and measurement values with eigenvalues of Hermitian operators imposes superfluous structure on the set {on,off}.

Similarly, in FFQ if observables are to be identified with Hermitian operators then the algebraic structure of the eigenvalues will be imposed on the measurement values. In this case the eigenvalues are field elements so the field structure will be imposed on any measurable property. This causes serious difficulties. Take again the example of position on a lattice with periodic boundary conditions. Suppose the number of points along a dimension of the lattice happens to be relatively prime to the field characteristic p. Then arithmetic involving position will lead to inconsistencies. So again, the usual identification of Hermitian operators and observables leads to difficulties in FFQ.

What then should observables in FFQ look like? One path to avoiding the problems outlined above would be to replace the eigenvalues of Hermitian operators with formal values that have no algebraic structure. The significance of the stipulation that the results of measurements in ordinary quantum theory be real numbers is that real numbers are invariant under conjugation. One can generalize this picture by replacing the real number eigenvalues of observables with formal objects that have the single property that they are invariant under conjugation. We then write observables as formal sums of progection operators. For example consider an observable with orthonormal eigenstates $|\phi_1\rangle$, $|\phi_2\rangle$,..., $|\phi_n\rangle$ and the formal values x_1, x_2,..., x_n as the corresponding measurement values. This observable would be written as the formal sum

$$O = \sum_i x_i \left|\phi_i\right\rangle \left\langle\phi_i\right| .$$

As long as we stipulate $x_i^* = x_i$ this operator is Hermitian. From this form of

the observable one can calculate formal expectation values:

$$\langle\psi|O|\psi\rangle = \langle\psi| \left(\sum_i x_i \, |\phi_i\rangle \, \langle\phi_i| \right) |\psi\rangle = \sum_i x_i \, |\langle\phi_i|\psi\rangle|^2 \, .$$

Considering observables this way allows us to remain agnostic as to the algebraic structure of measurement values, i.e. whether the results of measurements are field elements, real numbers, or anything else. And this in turn allows us to create observables for any property we can measure. We can translate observables from ordinary quantum mechanics, like spin operators, into FFQ while avoiding the question of how to label properties, like spins, with field elements.

Equipped with a way of constructing observables and expectation values for spin one can translate the experiments like the thought experiments behind Bell's theorem into FFQ. In ordinary quantum theory, when the usual real values (for instance $\pm\frac{1}{2}\hbar$ for spin $\frac{1}{2}$ measurements) are plugged into the formal variables we can compare the expectation values with inequalities. Bell's inequalities can be seen as a way of expressing certain constraints on the formal expectation values. In the case of FFQ constraints on the formal expectation values can't take the form of inequalities, and so must be expressed in some other way.

2.2 GHZ experiment

As mentioned there are well-known proofs of Bell's theorem that do not rely on inequalities. The Greenberger–Horne–Zeilinger (GHZ) experiment, for instance, provides just such a set up [6]. Briefly, the experiment is as follows. Three experimenters share a specific three-qubit quantum state with one qubit given to each experimenter. Each experimenter chooses one of two measurements to perform on her qubit: measurement of spin along the x-axis or measurement of spin along the y-axis. After each experimenter makes the measurement they come together and compare results. This process of sharing a particular state, making measurements, and comparing the results is repeated many times so that a statistical analysis can be made.

Using the idea of observables as formal sums discussed above we can translate the GHZ experiment into the FFQ framework. Let m_{up} and m_{down} be formal values representing the two possible outcomes of a spin measurement. Spin along the x-axis is associated with the observable $m_{\text{up}} \, |\uparrow_x\rangle \, \langle\uparrow_x| + m_{\text{down}} \, |\downarrow_x\rangle \, \langle\downarrow_x|$ where $|\uparrow_x\rangle$ is defined as $\alpha\big(\, |0\rangle + |1\rangle \, \big)$ and $|\downarrow_x\rangle$ as $\alpha\big(\, |0\rangle - |1\rangle \, \big)$, α being a constant for normalization. Spin along the y-axis is associated with the observable $m_{\text{up}} \, |\uparrow_y\rangle \, \langle\uparrow_y| + m_{\text{down}} \, |\downarrow_y\rangle \, \langle\downarrow_y|$ where $|\uparrow_y\rangle$ is defined as $\alpha\big(\, |0\rangle + i \, |1\rangle \, \big)$ and $|\downarrow_y\rangle$ as $\alpha\big(\, |0\rangle - i \, |1\rangle \, \big)$. After each of the three parties has made its measurement they come together and compare results. They count how many of the three measurements resulted in up. In particular they note whether the number of spin up outcomes was odd (1 or 3) or even (0 or 2).

We can view this observation of either "odd" or "even" as a measurement and can construct a corresponding observable. This observable will have the formal eigenvalues m_{odd} and m_{even}. Let a, b, and c be the axes (x or y) along

which parties one, two, and three make their measurements respectively. The vectors $|{\downarrow_a}{\downarrow_b}{\uparrow_c}\rangle$, $|{\downarrow_a}{\uparrow_b}{\downarrow_c}\rangle$, $|{\uparrow_a}{\downarrow_b}{\downarrow_c}\rangle$, and $|{\uparrow_a}{\uparrow_b}{\uparrow_c}\rangle$ form an orthonormal basis for the eigenspace associated with m_{odd}, while $|{\uparrow_a}{\uparrow_b}{\downarrow_c}\rangle$, $|{\uparrow_a}{\downarrow_b}{\uparrow_c}\rangle$, $|{\downarrow_a}{\uparrow_b}{\uparrow_c}\rangle$, and $|{\downarrow_a}{\downarrow_b}{\downarrow_c}\rangle$ form an orthonormal basis for the eigenspace associated with m_{even}. Our observable for the choice of axes a, b, and c is

$$\begin{aligned} O_{abc} =& m_{\text{odd}}\big(|{\downarrow_a}{\downarrow_b}{\uparrow_c}\rangle\langle{\downarrow_a}{\downarrow_b}{\uparrow_c}| + |{\downarrow_a}{\uparrow_b}{\downarrow_c}\rangle\langle{\downarrow_a}{\uparrow_b}{\downarrow_c}| \\ &+ |{\uparrow_a}{\downarrow_b}{\downarrow_c}\rangle\langle{\uparrow_a}{\downarrow_b}{\downarrow_c}| + |{\uparrow_a}{\uparrow_b}{\uparrow_c}\rangle\langle{\uparrow_a}{\uparrow_b}{\uparrow_c}|\big) \\ &+ m_{\text{even}}\big(|{\uparrow_a}{\uparrow_b}{\downarrow_c}\rangle\langle{\uparrow_a}{\uparrow_b}{\downarrow_c}| + |{\uparrow_a}{\downarrow_b}{\uparrow_c}\rangle\langle{\uparrow_a}{\downarrow_b}{\uparrow_c}| \\ &+ |{\downarrow_a}{\uparrow_b}{\uparrow_c}\rangle\langle{\downarrow_a}{\uparrow_b}{\uparrow_c}| + |{\downarrow_a}{\downarrow_b}{\downarrow_c}\rangle\langle{\downarrow_a}{\downarrow_b}{\downarrow_c}|\big). \end{aligned}$$

We can use these observables to calculate (formal) expectation values for the measurements in the experiment. In the version of the GHZ experiment we are considering the shared state is $|\text{GHZ}\rangle = \alpha\big(|000\rangle - |111\rangle\big)$. The experimenters do measurements for each of the following four choices of axes: $(a,b,c) = (x,y,y)$, (y,x,y), (y,y,x), and (x,x,x). These choices of axes correspond to the observables O_{xyy}, O_{yxy}, O_{yyx}, and O_{xxx} respectively. Straightforward calculations give the expectation values as

$$\begin{aligned} \langle\text{GHZ}|O_{xyy}|\text{GHZ}\rangle &= m_{\text{odd}} \\ \langle\text{GHZ}|O_{yxy}|\text{GHZ}\rangle &= m_{\text{odd}} \\ \langle\text{GHZ}|O_{yyx}|\text{GHZ}\rangle &= m_{\text{odd}} \\ \langle\text{GHZ}|O_{xxx}|\text{GHZ}\rangle &= m_{\text{even}} \end{aligned}$$

Theorem 1.1 suggests that there is no way to interpret these expectation values as giving probabilities. But the four expectation values above are all rather special expectation values; they all contain a single term. All we need to make sense of the GHZ experiment is to be able to interpret these one-term expectation values. If we were using real number coefficients then calculating an expectation value of $1m_{\text{odd}} + 0m_{\text{even}}$ would mean that the outcome of the measurement was always m_{odd} and never m_{even}. But we are using finite fields where we have shown that $\beta m_{\text{odd}} + (1-\beta)m_{\text{even}}$ cannot in general be interpreted as giving the probability of m_{odd} or of m_{even}. So is there a scheme for interpreting these formal expectation values that does not involve probability but that does allow us to assign meaning like "always m_{odd} and never m_{even}" to expectation values like $1m_{\text{odd}} + 0m_{\text{even}}$.

This question suggests a modal interpretation of the coefficients. Rather than interpret coefficients as probabilities we can look for a different meaning for them, one that tells about possibilities, not probabilities. A state may not provide information on the probability of any measurement outcome, but perhaps it can provide information on the possibility of any measurement outcome. That is, while there is no equivalent of the Born rule in FFQ, there may be some function that takes as inputs a state and a measurement outcome and outputs either "impossible" or "possible" rather than a probability. More precisely we seek a function $g\colon \mathbb{F}_{p^2}^d \times \mathbb{F}_{p^2}^d \longrightarrow \{\text{impossible}, \text{possible}\}$ which must satisfy certain properties, in order for the output of this new rule to have the correct modal

interpretation. It should satisfy at least the following requirements which are analogous to properties (i), (ii), and (iii), listed in section 1. First g must satisfy a normalization requirement:

1.) For any orthonormal basis $\{|i\rangle\}$ and any state $|\psi\rangle$, for at least one choice of i we must have $g(|\psi\rangle, |i\rangle) =$ possible.

That is, at least one outcome from a complete set of mutually exclusive outcomes must be possible. Reasoning similar to that behind property (ii) suggests the following condition:

2.) If $\langle\phi|\psi\rangle = 0$ then $g(|\psi\rangle, |\phi\rangle) =$ impossible.

The final property is that of basis independence:

3.) $f(U|\phi\rangle, U|\psi\rangle) = f(|\phi\rangle, |\psi\rangle)$ where U is any unitary.

These three properties are satisfied by the obvious choice of g:

$$g\big(|\psi\rangle, |\phi\rangle\big) = \begin{cases} \text{impossible} & \text{if } \langle\psi|\phi\rangle = 0 \\ \text{possible} & \text{otherwise} \end{cases}.$$

Equipped with this "modal Born rule" the GHZ experiment in FFQ can be interpreted. The expectation value $1m_{\text{odd}} + 0m_{\text{even}}$ can be interpreted as saying that the outcome m_{odd} is possible while the outcome m_{even} is impossible. In other words an expectation value of m_{odd} means that upon measurement there will be an odd number of spin up results. And similarly an expectation value of m_{even} means that there will be an even number of spin measurements resulting in up. From here the arguments establishing Bell's theorem are the same as those used in the case of ordinary quantum theory. See [6] for the full argument.

Thus in FFQ, as in ordinary quantum theory, local realism is violated. Moreover, it is true in FFQ, as it is in ordinary quantum theory, that any variation of the GHZ experiment where the state shared by the researchers is separable (the tensor product of three qubits) will not lead to a violation of local realism.

3 Entanglement

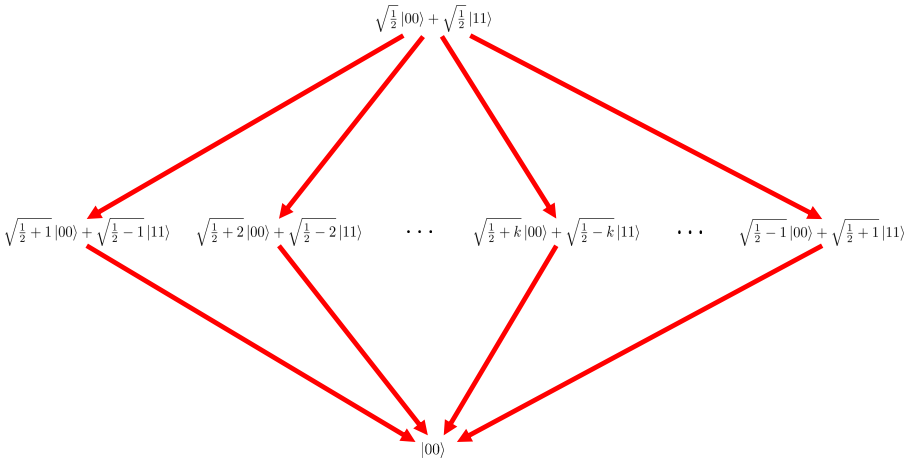
That local realism can be violated only with a shared entangled state suggests the view of entanglement as a resource. The GHZ experiment involves a rather specific task and a rather specific entangled state, however, so it by itself does not give the full picture of entanglement as a resource. An important part of any such "full picture" would be a way of quantifying entanglement. Entanglement, viewed as a nonlocal resource should be a quantity that does not increase under local operations or classical communication (LOCC) [8]. We can associate the amount and kind of entanglement in a state to its LOCC relationships to other states. When two states can be reached from each other via local operations and classical communication they must be equally entangled. The equivalence

relation of being reachable from each other via LOCC divides the space of states into equivalence classes. The LOCC relationships between these classes give a partial ordering on the classes where $|\psi\rangle \succeq |\phi\rangle$ when one can transform $|\psi\rangle$ into $|\phi\rangle$ via local operations and classical communication. This partial order provides a complete description of entanglement.

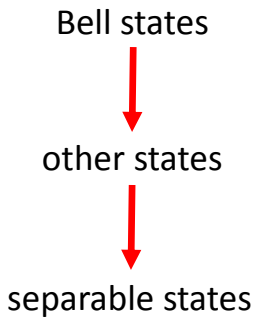
One of the few desirable properties of FFQ is that all the states as well as operations between them can be enumerated. Thus, in principle, we can explicitly chart the partial ordering imposed by LOCC on the states in any given vector space $\mathbb{F}_{p^2}^{d_1} \otimes \mathbb{F}_{p^2}^{d_2} \otimes \cdots \otimes \mathbb{F}_{p^2}^{d_n}$. Let the states be the vertices of a direted graph. For any pair of states we can ask if they are related by LOCC and if so draw an arrow between them. This directed graph would completely describe the structure of entanglement.

We here present some partial results on characterizing the LOCC structure of states. For simplicity the discussion is restricted to pure states. First it is clear that any separable state can be obtained from any other state via LOCC. This can be done via the local operation that simply replaces each subsystem with the desired pure state. This is same as in ordinary quantum theory and is expected. A more surprising fact is that it is not immediately obvious, as it is in ordinary quantum mechanics, that a nonseparable state can never be obtained from a separable one via LOCC. If a LOCC existed that took a separable state to a non-separable one that would imply that non-separability per se is not a resource, and that non-separability is distinct from entanglement. This would be in contrast to ordinary quantum mechanics where any non-separable state gives an advantage in some tasks [8]. A cursory search for a LOCC from a separable to a non-separable state was insufficient to find any, however.

In ordinary quantum mechanics entanglement between two qubits is completely characterized by the Schmidt decomposition of a state [7], a result known as Nielsen's majorization theorem. However, similarly to the case of diagonalization of matrices, the Schmidt decomposition of a state in FFQ does not usually exist. We have not found a result in FFQ comparable to the majorization theorem that would completely characterize entanglement. Instead, in numerical searches we have found that many if not all two-qubit states can be reached via LOCC from a Bell state. By Bell state we mean a two-qubit pure state whose two qubits viewed individually are both the maximally mixed state, the mixed state whose density matrix is proportional to the identity. All Bell states are LOCC equivalent in FFQ, i.e. any Bell state can be reached by any other Bell state via LOCC. This picture is illustrated in the following figure:



In the above figure, in an abuse of notation we mean by $\sqrt{\beta}$ a choice of element of $\mathbb{F}_{p^2}$ such that $\sqrt{\beta}^*\sqrt{\beta} = \left|\sqrt{\beta}\right|^2 = \beta$. The states in the figure are representative from their LOCC class. An arrow between from one state to another means that via LOCC one can transform anything in the LOCC class of the first state into anything in the LOCC class of the second. The above chart is very incomplete. There are other LOCC classes whose position in the chart is not known and there may possibly be additional arrows between the classes shown. For instance investigations for small characteristic p have found paths between some of the states in the middle row of the figure above. This observation leads to a conjecture that for all p there are three LOCC classes, Bell states, separable states, and everything else. Then as there are LOCCs from Bell states to the other two as well as LOCCs to separable states from the other two, the complete picture of entanglement for two qubits would be the following: Bell states are the most entangled, separable states are not entangled at all, and everything else has the same amount of entanglement, more than separable states but less than Bell states. As a picture the conjecture is the following:



For bipartite entanglement in more dimensions and for multipartite entanglement we expect the picture to be more complex.

3.1 Counting entangled states

In the partial ordering of LOCC classes some LOCC classes are maximal, i.e. states in a maximal class cannot be obtained via LOCC from any other state outside the class. This defines a concept of maximal entanglement. The states in maximal LOCC classes are in some sense the most entangled states. There may be many distinct maximal LOCC classes, as in the case of ordinary quantum mechanics, where for instance the states $|\mathrm{GHZ}\rangle = |000\rangle + |111\rangle$ and $|W\rangle = |001\rangle + |010\rangle + |100\rangle$ are both maximally entangled but are not related via LOCC [2].

As discussed above, in the case of two qubits the Bell states likely form a maximal class. The Bell states are the two-qubit pure states whose individual qubit density matrices are proportional to the identity. Generalizing this we call a pure state of n qubits maximally entangled if its qubits viewed individually each have density matrices proportional to the identity. For example, in the state $\alpha\big(\,|000\rangle + |111\rangle\,\big)$ (where α is for normalization) each qubit has density matrix $\left(\begin{smallmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{smallmatrix}\right)$, so by the above definition $\alpha\big(\,|000\rangle + |111\rangle\,\big)$ is maximally entangled. We conjecture that these maximally engtangled states, like the Bell states in the case of two qubits, form maximal LOCC classes.

This definition of maximally entangled differs from one proposed previously in [4]. There, a maximally entangled state is defined as a state $|\psi\rangle$ such that

$$0 = \sum_{j=1}^{n} \sum_{\mu=x,y,z} \langle\psi|\sigma_\mu^j|\psi\rangle^2 \tag{3}$$

where σ_μ^j is the μ Pauli operator acting on the jth qubit:

$$\sigma_\mu^j = \mathbb{1} \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1} \otimes \sigma_\mu \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1}.$$

This definition is unsatisfactory. Given a state, simply adding an extra pure qubit will increase the sum in (3) by 1. So adding some multiple of p pure qubits to a state does not change the sum at all. This allows for some states that are mostly made up of separable qubits to be counted among the maximally entangled states. This is unacceptable for the view of entanglement as a resource. Intuitively, adding a separable qubits to a system makes it no more useful than before as pure qubits can be created by any party locally and without communication.

Fortunately this definition can be modified into a satisfactory one. Our differing definition of maximally entangled, phrased above as the density matrix of each qubit being proportional to the identity, is equivalent to the condition

$$\forall j, \forall \mu \quad 0 = \langle\psi|\sigma_\mu^j|\psi\rangle\,.$$

This is a modification of (3). Whereas our definition requires that each of several expectation values equal zero, the condition (3) requires only that their sum equal zero. Note that in the case of ordinary quantum mechanics these two definitions coincide, because for real numbers a sum of square terms is zero if and only if each term is zero. In FFQ, however, because of the modular arithmetic of finite field elements the definitions are very different.

Because the state space in FFQ is finite the maximally entangled states in any number of qubits n can be enumerated. The problem of counting the number of maximally entangled states was attempted in [4]. We again attempt this problem but with the improved definition of maximal entanglement. For small p and n the number of maximally entangled states can be computed by explicitly listing all states, testing them for maximal entanglement, and counting the results. Results of such computations are summarized in the following table:

(n, p)	total states $= \frac{p^{2^n-1}(p^{2^n}-1)}{(p+1)}$ [4]	maximally entangled states
$(2, 3)$	540	24
$(2, 7)$	102900	336
$(2, 11)$	1623820	1320
$(2, 19)$	44693244	6840
$(3, 3)$	3586680	2160
$(3, 7)$	593445085800	823200

These values suggest the formula $p(p^2-1)$ for two-qubit states, while the formula for three-qubit states remains unclear. This brute-force method quickly becomes intractable as n and p increase, and in any case it cannot prove a formula for arbitrary n and p, the eventual goal.

The problem of counting the number of maximally entangled states can be reformulated in terms of orbits under a group action. The local unitary transformations $U(d_1)\otimes U(d_2)\otimes\cdots\otimes U(d_n)$ form a group that acts on the space of states. As local unitary transformations are reversible, two states related by a local unitary are in the same LOCC class. Thus entanglement is unchanged by local unitary transformations. What this means is that given a state $|\psi\rangle$, all states in its orbit O_ψ have the same entanglement properties. In particular the orbit of a maximally entangled state should contain only maximally entangled states, and in fact it does. Under the action of the group of local unitaries the space of states splits into several different orbits a certain number of which will contain maximally entangled states. Counting the size of each of these orbits gives the desired number, the number of maximally entangled states. With the problem formulated in terms of orbits results like the well-known orbit-stabilizer theorem can be applied to the problem. This may allow the derivation of a general formula in p and n for the number of maximally entangled states.

4 Conclusion

The incompatibility of FFQ with a probabilistic interpretation suggests it is not a quantum theory describing reality. At the very least it is cannot be a complete

description of reality. The state vector does not contain the full information on the probabilities of various outcomes. For nature to use FFQ it must keep track of the probabilities through some other object or hidden variable.

FFQ is useful, on the other hand, as a foil theory. Certain properties that may be taken for granted in ordinary quantum theory fail in FFQ. Thus a study of FFQ can clarify the conceptual structure of ordinary quantum mechanics. The role of Hermitian operators as observables in ordinary quantum mechanics, for example, is clarified through the lense of FFQ, as is discussed above. A look at FFQ can also clarify other concepts that are not discussed in this work, including quantum dynamics.

In the search for quantum theories amenable to computation there are likely other choices besides quantum theory over finite fields. There are many ways to generalize quantum theory including such natural ways as generalizing the space of states to other convex spaces [1]. It is conceivable that among these sorts of generalizations are theories that are amenable to computation and that simultaneously have some bearing on reality.

References

[1] H. Barnum and A. Wilce. Post-Classical Probability Theory. *ArXiv e-prints*, May 2012.

[2] W. Dur, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62:062314, Nov 2000.

[3] D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser. Foundations of quaternion quantum mechanics. *Journal of Mathematical Physics*, 3(2):207–220, 1962.

[4] A. J. Hanson, G. Ortiz, A. Sabry, and Y.-T. Tai. Geometry of discrete quantum computing. *Journal of Physics A: Mathematical and Theoretical*, 46(18):185301, 2013.

[5] A. J. Hanson, G. Ortiz, A. Sabry, and Y.-T. Tai. Discrete quantum theories. *Journal of Physics A: Mathematical and Theoretical*, 47(11):115305, 2014.

[6] N. D. Mermin. Quantum mysteries revisited. *American Journal of Physics*, 58(8):731–734, 1990.

[7] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83:436–439, Jul 1999.

[8] M. B. Plenio and S. Virmani. An Introduction to entanglement measures. *Quant.Inf.Comput.*, 7:1–51, 2007.

[9] B. Schumacher and M. Westmoreland. Modal quantum theory. *Foundations of Physics*, 42(7):918–925, 2012.

[10] E. C. G. Stueckelberg. Quantum theory in real hilbert space. *Helvetica Physica Acta*, 33:727, 1960.

Conformal Compactification of $\mathbb{R}^{p,q}$

Lyndee M Govert *

Abstract

In this paper we are concerned with the conformal transformations on the general $\mathbb{R}^{p,q}$ setting. We will explore and establish the conformal compactification of the space. Further, we will generalize the property of mapping circles and lines into circles and lines that exists on the extended complex plane.

1 Introduction

We can describe $\mathbb{R}^{p,q}$ as the generalized Minkowski space. This is the vector space $\mathbb{R}^n$ where $n = p + q$ together with the indefinite quadratic form $Q(x) = (x_1)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q})^2$ with p pluses and q minuses. Related to this quadratic form we have the bilinear form given by the polarization identity $B(x, y) = \frac{1}{4}[Q(x+y) - Q(x-y)] = x_1y_1 + ... + x_py_p - x_{p+1}y_{p+1} - ... - x_{p+q}y_{p+q}$. Also related to this quadratic form we have the indefinite orthogonal group of linear transformations $O(p, q) = \{ M \in GL_{p+q}(\mathbb{R}), Q(Mx) = Q(x) \forall x \in \mathbb{R}^{p+q}\}$. The group $SO(3, 1)$ gives precisely the group of Mobius transformations over the extended complex plane which will be discussed in the next section.

2 Fractional Linear Transformations

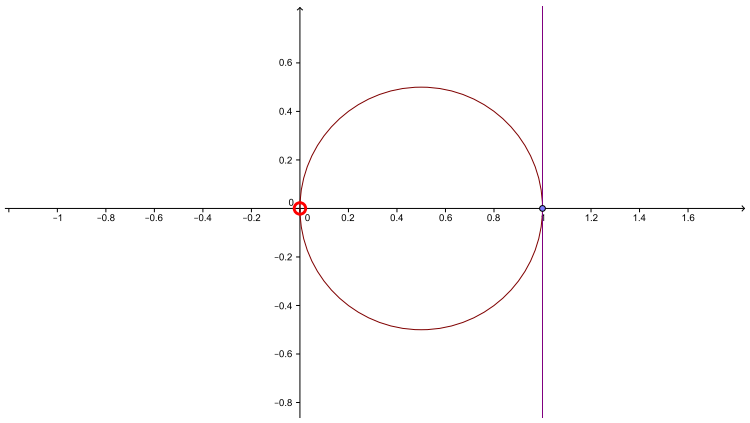
Let us consider the action of the general linear group of invertible 2×2 matrices with complex entries over the complex plane. That is $GL(2, \mathbb{C})$ acting on $\mathbb{C}$. So for $z \in \mathbb{C}$, $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2, \mathbb{C})$ takes $z \to \frac{az+b}{cz+d}$. These are the orientation preserving transformations. We also have the orientation reversing transformations which will take $z \to \frac{a\overline{z}+b}{c\overline{z}+d}$ where $\overline{z}$ is the complex conjugate of z. In the specific instance that $cz + d = 0$ we can consider $z \to \infty$. These transformations also have the property that for M_1 and M_2 we have $M_1z \to \frac{a_1z+b_1}{c_1z+d_1}$ and $M_2z \to \frac{a_2z+b_2}{c_2z+d_2}$. The composition of these M_1 and M_2 gives the transformation associated with $M_3 = M_2M_1$.

Three types of transformations described by this group are the parallel translations, dilations, and rotations. Parallel translations will take $z \to z + b$ and are given by the matrices of the form $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$. The dilations act by a factor

*Indiana University Northwest, lgovert@iun.edu

of λ on z such that $z \to \lambda z$ by the matrices of the form $\begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}$. Similarly the rotations act by a factor of θ with the identity $a = e^{i\theta} = \cos(\theta) + i\ \sin(\theta)$ such that $z \to az$ given by the matrices of the form $\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$.

Finally, in this group we also have the transformations described as inversions. These take $z \to \frac{1}{z}$ given by the matrices $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. This group produces the transformations we are particularly interested in this paper in that they have the property that they will take circles and lines into circles and lines. One inversion is shown here where we map the vertical line at 1 into the circle centered at $\frac{1}{2}$. This example leaves the point at 1 fixed and maps points on the line that approach ∞ to points on the circle that approach the origin.



This transformation gives us a slight problem in the situation where we wish to examine maps including 0 or ∞. We would like to say that $0 \to \infty$, however this is not defined under this mapping. This problem can be eradicated by considering, instead of $\mathbb{C}$, the extended complex plane $\mathbb{C} \cup \{\infty\}$.

3 Projective Plane and Compactification

As we have just seen, when considering transformations we sometimes run into a problem at infinity. For the complex plane we have an established solution to this by use of the extended complex plane. This is one example of a one point compactification. In general however this is not enough. It is possible that we have more than just one point hiding at infinity and this will require a more involved compactification.

For purposes of completeness we will recall the definition of the projective plane using the notion of representative vectors. Given that V is an n-dimensional vector space, then any 1-dimensional subspace of V is the set of

multiples of a non-zero vector $v \in V$ [Hitch03]. So we say v is a representative vector for the point $[v] \in P_{n-1}(V)$. It is clear from this definition that $[\lambda v] = [v]$ for $\lambda \in \mathbb{R}\backslash\{0\}$. Another way to say this for $\mathbb{R}$ would be $P_{n-1}(V) \cong V\backslash\{0\}$ with respect to the equivalence of proportionality.

Consider now the following map:

$$\iota : \mathbb{R}^{p,q} \to \mathbb{R}^{p+1,q+1} \to \mathbb{P}_{n+1}(\mathbb{R})$$

$$x = (x_1, ..., x_n) \to (\frac{1 - Q(x)}{2}, x_1, ..., x_n, \frac{1 + Q(x)}{2}) \to [\frac{1 - Q(x)}{2}, x_1, ..., x_n, \frac{1 + Q(x)}{2}]$$

The conformal compactification of $\mathbb{R}^{p,q}$ is defined as the closure of the image from this map.

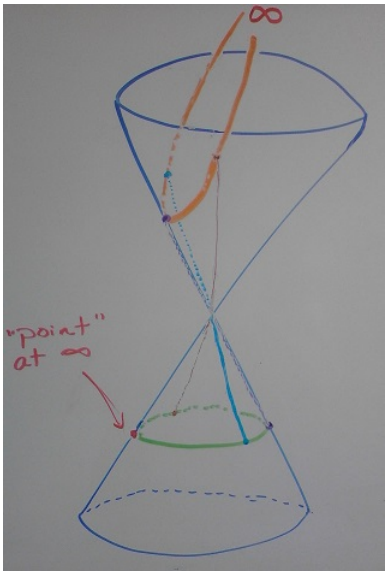
$$\overline{\iota(\mathbb{R}^{p,q})} = N^{p,q}$$

$$N^{p,q} = \{(x_0, x_1, ..., x_n, x_{n+1}) \in \mathbb{P}_{n+1}(\mathbb{R}) \text{ such that}$$

$$\hat{Q}(x) = (x_0)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q+1})^2 = 0\}$$

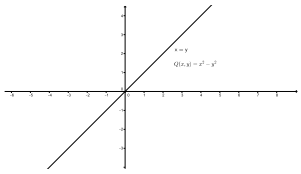
[Schot08]

Example 3.1 $\mathbb{R}^{1,0}$



For an example that we can visualize we look at $\mathbb{R}^{1,0}$. This is the real line. The first part of the map takes the real line to the cone $\mathbb{R}^{2,1}$ as shown by the orange parabola in the graphic. Next we project the parabola to the circle shown in green in the graphic. Every part of the line maps except for the "point" at infinity. This is included as we take the closure of the image. Thus we have mapped the real line into a closed circle by this map.

Example 3.2 (A Line at Infinity) *In the case where $p \neq 0$ and $q \neq 0$ we have to consider there may be more than just one point at infinity. In the case of $\mathbb{R}^{1,1}$ when we apply the inversion map $x \to \frac{x}{Q(x)}$ consider what happens with the lines $x = y$ or $x = -y$.*



The quadratic form associated with $\mathbb{R}^{1,1}$ will give us $Q(x, y) = x^2 - y^2$ and thus this inversion mapping will send the entire $x = y$ line and $x = -y$ line to infinity. This idea motivates us to reconsider how we define spheres and hyperplanes in the conformal compactification $N^{p,q}$.

Just as we have the group $O(p, q)$ acting on $\mathbb{R}^{p,q}$ as described earlier now we have $O(p+1, q+1)$ acting on $N^{p,q}$. This group $O(p+1, q+1)$ acts on $N^{p,q}$ by linear transformation and has the property of preserving the quadratic form.

We look at the intersection of $N^{p,q}$ with a hyperplane in $\mathbb{R}^{p+1,q+1}$ and we obtain the hypersurfaces given by $\alpha Q(x) + B(a, x) + \gamma = 0$. We claim that our hyperplanes in $\mathbb{R}^{p+1,q+1}$ are given by a linear equation of the form $a_0x_0 + ... + a_px_p - a_{p+1}x_{p+1} - ... - a_{p+q+1}x_{p+q+1} = 0$. To prove this is the case we can pull back the image here by the map above.

From the image of x given above we have $x_0 = \frac{1-Q(x)}{2}$ and $x_{p+q+1} = \frac{1+Q(x)}{2}$.

$$a_0\frac{1-Q(x)}{2} + a_1x_1 + ... + a_px_p - a_{p+1}x_{p+1} - ... - a_{p+q}x_{p+q} - a_{p+q+1}\frac{1+Q(x)}{2} = 0$$

$$\frac{a_0}{2} - \frac{a_0}{2}Q(x) + B(a, x) - \frac{a_{p+q+1}}{2} - \frac{a_{p+q+1}}{2}Q(x) = 0.$$

$$-\frac{a_0 + a_{p+q+1}}{2}Q(x) + B(a, x) + \frac{a_0 - a_{p+q+1}}{2} = 0$$

Let $\alpha = -\frac{a_0+a_{p+q+1}}{2}$ and $\gamma = \frac{a_0-a_{p+q+1}}{2}$ then we have $\alpha Q(x) + B(a,x) + \gamma = 0$, which shows we have the correct definition for a hypersurfaces in $\mathbb{R}^{p+1,q+1}$. In the lower dimensional cases like the extended complex plane as mentioned earlier we have the concept of the mapping circles and lines into circles and lines. Now we have generalized this idea to the case where we replace circles and lines with quadratic hypersurfaces, spheres, hyperboloids, cones, and hyperplanes.

4 Hypersurfaces, $p = 0$ or $q = 0$

Let us first recall that $N^{p,q} = \{(x_0, x_1, ..., x_n, x_{n+1}) \in \mathbb{P}_{n+1}(\mathbb{R})$ such that $\hat{Q}(x) = (x_0)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q+1})^2 = 0\}$. And our hyperplanes in $\mathbb{R}^{p+1,q+1}$ are now given in the form $a_0x_0 + ... + a_px_p - a_{p+1}x_{p+1} - ... - a_{p+q+1}x_{p+q+1} = 0$. Where vector $A = (a_0, ..., a_p, a_{p+1}, ..., a_{p+q+1})$ is the vector perpendicular to the plane. With these definitions at hand we can define a hyperboloid as the intersection of a hyperplane with $N^{p,q}$.

In the case where either $p = 0$ or $q = 0$ it may be possible that this intersection is empty or a single point. We want to identify these situations since we would not want to call these particular intersections spheres. We will examine here the case where $q = 0$.

When $q = 0$ we have:

$$a_0x_0 + ... + a_px_p - a_{p+1}x_{p+1} = 0 \text{ and } (x_0)^2 + ... + (x_p)^2 - (x_{p+1})^2 = 0$$

Acting on the hyperplanes described here we have the indefinite orthogonal group of $n \times n$ matrices $O(p+1) \subset O(p+1,1)$. With the appropriate selection of $M \in O(p+1)$ we obtain a hyperplane with all but possibly the first and last entry of the A vector being zero. We allow for the first or the last entry to be zero but not both. Thus we are looking at the intersection of $a_0x_0 - a_{p+1}x_{p+1} = 0$ and $(x_0)^2 + ... + (x_p)^2 - (x_{p+1})^2 = 0$. We have three cases to consider.

4.1 Case 1

$$\text{Let } a_0 = 0,\ a_{p+1} \neq 0$$

$$a_{p+1}x_{p+1} = 0 \implies x_{p+1} = 0$$

$$(x_0)^2 + ... + (x_p)^2 = 0$$

Thus we obtain only the trivial solution $[x] = [0]$ and since $N^(p,q)$ does not contain 0 this trivial solution corresponds to $\emptyset$, and we have $\hat{Q}(A) < 0$.

4.2 Case 2

$$\text{Let } a_0 \neq 0,\ a_{p+1} = 0$$

$$a_0x_0 = 0 \implies x_0 = 0$$

$$(x_1)^2 + ... + (x_p)^2 - (x_{p+1})^2 = 0$$

Thus we obtain hyper surfaces, in the case where $p = 0$ or $q = 0$ these are hypersurfaces, and we have $\hat{Q}(A) > 0$.

4.3 Case 3

$$\text{Let } a_0 \neq 0,\ a_{p+1} \neq 0, \text{ So } x_0 = \tfrac{a_{p+1}}{a_0}(x_{p+1})$$

$$(\frac{a_{p+1}}{a_0})^2(x_{p+1})^2 + (x_1)^2 + ... + (x_p)^2 - (x_{p+1})^2 = 0$$

$$(x_1)^2 + ... + (x_p)^2 = (1 - (\tfrac{a_{p+1}}{a_0})^2)(x_{p+1})^2$$

We have here three possibilities for $1 - (\frac{a_{p+1}}{a_0})^2$.

4.3.1

$$1 - (\frac{a_{p+1}}{a_0})^2 = 0$$

So $|a_0| = \pm|a_{p+1}|$, solutions are of the form $(0, 0, ..., 0, \lambda)$, $\lambda \in \mathbb{R}$, where $x_0 = x_1 = ... = x_p = 0$, and we have $\hat{Q}(A) = 0$.

4.3.2

$$1 - (\frac{a_{p+1}}{a_0})^2 > 0$$

So $|a_0| > |a_{p+1}|$, all solutions give hypersurfaces, and $\hat{Q}(A) > 0$.

4.3.3

$$1 - (\frac{a_{p+1}}{a_0})^2 < 0$$

So $|a_0| < |a_{p+1}|$, only solution is trivial, and $\hat{Q}(A) < 0$.

Since we originally used the action of invertible $O(p + 1) \subset O(p + 1, 1)$ to simplify this intersection we can apply the inverse of the necessary $M \in O(p + 1, 1)$ to generalize these results. The important notion here is that all elements of $O(p + 1, q + 1)$ preserve quadratic form so the properties of $Q(A)$ will not be changed.

We have established three distinct families of hypersurfaces given by the intersections of hyperplanes in $\mathbb{R}^{p+1,q+1}$ and $N^{p,q}$. When $p = 0$ these families are classified by their representative quadratic form. $\hat{Q}(A) = 0$ are the one point intersections in the projective space. $\hat{Q}(A) > 0$ give the hyper surfaces we wish to call hyperboloids. Finally $\hat{Q}(A) < 0$ represent the trivial intersections which in the projective space give exactly $\emptyset$. When $p = 0$ our $\hat{Q}(A) < 0$ and $\hat{Q}(A) > 0$ cases are reversed.

5 Hypersurfaces, $p \neq 0$ and $q \neq 0$

The case where both $p \neq 0$ and $q \neq 0$ is similar in approach to the previous section. We have hyperplanes in $\mathbb{R}^{p+1,q+1}$ given as $a_0x_0+...+a_px_p-a_{p+1}x_{p+1}-...-a_{p+q+1}x_{p+q+1}=0$. We use an element of $O(p+1)\times O(q+1) \subset O(p+1,q+1)$ which acts as a rotation bringing all but possibly the first and last entries of the A vector to be zero. Again we allow for either $a_0 = 0$ or $a_{p+q+1} = 0$ but not for both. Now we are considering the intersection of the plane $a_0x_0 - a_{p+1}x_{p+1} = 0$ with $(x_0)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q+1})^2 = 0$. We again have three cases to consider.

5.1 Case 1

$$\text{Let } a_0 = 0,\ a_{p+q+1} \neq 0$$

$$a_{p+q+1}x_{p+q+1} = 0 \implies x_{p+q+1} = 0$$

$$(x_0)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q})^2 = 0$$

Here we obtain a hypersurface and we have $\hat{Q}(A) < 0$.

5.2 Case 2

$$\text{Let } a_0 \neq 0,\ a_{p+q+1} = 0$$

$$a_0x_0 = 0 \implies x_0 = 0$$

$$(x_1)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q+1})^2 = 0$$

Again we obtain a hypersurface but here we have $\hat{Q}(A) > 0$.

5.3 Case 3

$$\text{Let } a_0 \neq 0,\ a_{p+q+1} \neq 0,\ \text{So } x_0 = \tfrac{a_{p+q+1}}{a_0}x_{p+q+1}$$

$$(\frac{a_{p+q+1}}{a_0})^2(x_{p+q+1})^2 + (x_1)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q+1})^2 = 0$$

$$(x_1)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q})^2 = (1 - (\tfrac{a_{p+q+1}}{a_0})^2)(x_{p+q+1})^2$$

Now again we have three possibilities for $1 - (\frac{a_{p+q+1}}{a_0})^2$.

5.3.1

$$1 - (\frac{a_{p+q+1}}{a_0})^2 = 0$$

Here we obtain hypersurfaces, and we have $|a_0| = \pm|a_{p+q+1}|$ so $\hat{Q}(A) = 0$.

5.3.2

$$1 - (\frac{a_{p+q+1}}{a_0})^2 > 0$$

Here we obtain hypersurfaces and we have $|a_0| > |a_{p+q+1}|$ so $\hat{Q}(A) > 0$.

5.3.3

$$1 - (\frac{a_{p+q+1}}{a_0})^2 < 0$$

Finally we again obtain hypersurfaces and we have $|a_0| < |a_{p+q+1}|$ so $\hat{Q}(A) < 0$.

As we have now seen it is not possible to obtain a surface of co-dimension greater than 1 when both $p \neq 0$ and $q \neq 0$. We do still however end up with the families of hypersurfaces characterized by their quadratic form being $\hat{Q}(A) = 0$, $\hat{Q}(A) > 0$, or $\hat{Q}(A) < 0$.

6 Conclusion

We have established a conformal compactification of $\mathbb{R}^{p,q}$ which we have called $N^{p,q}$. We have also considered a new way to define hypersurfaces in this conformal compactification.

Definition 6.1 *Quadratic Hypersurfaces in $N^{p,q}$ are given by the intersection of hyperplanes in $\mathbb{R}^{p+1,q+1}$, $a_0x_0 + ... + a_px_p - a_{p+1}x_{p+1} - ... - a_{p+q+1}x_{p+q+1} = 0$ and $(x_0)^2 + ... + (x_p)^2 - (x_{p+1})^2 - ... - (x_{p+q+1})^2 = 0$.*
With the exception of the following cases:

1. *When $q = 0$ we must exclude $\hat{Q}(A) = 0$ which gives a one point intersection and $\hat{Q}(A) < 0$ which gives $\emptyset$.*

2. *When $p = 0$ we must exclude $\hat{Q}(A) = 0$ which gives a one point intersection and $\hat{Q}(A) > 0$ which gives $\emptyset$.*

With this definition in mind we can now establish a generalization to the property of mapping circles and lines into circles and lines where we replace circles and lines with quadratic hypersurfaces.

Theorem 6.2 *The indefinite orthogonal group $O(p+1, q+1)$ acts conformaly on $\mathbb{R}^{p+1,q+1}$ in such a way that it maps quadratic hypersurfaces onto quadratic hypersurfaces of the same dimension.*

Proof Let a quadratic hypersurface be defined as the intersection of $N^{p,q}$ and a hyperplane in $\mathbb{R}^{p+1,q+1}$ as in Def 6.1. We can then say the action of $O(p+1, q+1)$ gives another intersection of a hyperplane in $\mathbb{R}^{p+1,q+1}$ with $N^{p,q}$ which is another hypersurface. So $O(p+1, q+1)$ maps hypersurfaces onto hypersurfaces. □

References

[DCar67] Do Carmo, M. P. (1967). Differential geometry of curves and surfaces. Englewood Cliffs, NJ:Prentice-Hall. 214-227.

[Hitch03] Hitchin,N. (2003). *Projective geometry* [Lecture notes]. Retrieved from Lecture notes by Nigel Hitchin website: http://people.maths.ox.ac.uk/hitchin/hitchinnotes/hitchinnotes.html

[Schot08] M.Schottenloher. *A mathematical introduction to conformal field theory*, volume 759 of Lecture Notes in Physics. Springer-Verlag, 2nd edition, 2008.

Removing noise from a coarse-grain signal

Zachary Helbert[*], *Michael Jolly*[†]

Abstract

The long term behavior of a dissipative differential equation in a possibly infinite-dimensional phase space can be captured by a determining form which is an ODE in $C_b(\mathbb{R}, \mathbb{R}^n)$, a space of projected trajectories. Here n is the number of determining parameters which could be nodal values in a coarse-grain, a finite difference method, or Fourier modes in a low-dimensional Galerkin approximation, just to name two examples. We seek to remove noise in a coarse-grain signal (time series in t), which acts as an initial condition for the determining form. As the solution of the form evolves, it converges to a steady state, which is a portion of a trajectory in t on the global attractor. We apply this technique to both the Lorenz and Kuramoto-Sivashinsky equations, showing the effects of changing damping parameters in the determining form. All results show the reduction of noise and convergence toward a solution on the global attractor.

1 Introduction

Determing parameters were first introduced as a way to gauge the number of degrees of freedom for the Navier-Stokes equations (NSE) [1]. Determining parameters can be described by a projector P (onto, for example nodal values in a coarse-grain, finite difference method) such that any complete solution $u(\cdot) = \{u(t)\colon t \in \mathbb{R}\}$ on the global attractor $\mathcal{A}$ is uniquely determined by its projection $p(\cdot) = Pu(\cdot)$. This defines a map W from the set $\mathcal{S}$ of projected trajectories in $\mathcal{A}$ into the space $C_b(\mathbb{R}, H)$, where H is the phase space of the 2D NSE. The extension of this map from the set $\mathcal{S}$ to the Banach space $X = C_b(\mathbb{R}, PH)$ leads to a *determining form*, which is an ordinary differential equation in X. This was first done in [2] for the particular case where P is the projection onto a finite number of Fourier modes. In that case trajectories in the global attractor of the NSE correspond to traveling waves of the determining form. Later, a different extension and determining form were derived for the 2D NSE in [3], which allows for the use of a variety of determining parameters. In this more general approach solutions in $\mathcal{A}$ are identified with the steady states of the determining form.

The extension in [3] follows a method of data assimilation. This refers to a process for completing, or improving an inadequate initial condition. For

[*]East Tennessee State University, helbertz@goldmail.etsu.edu
[†]Indiana University, msjolly@imap.iu.edu

most applications the initial data which should be defined on the entire physical domain, can be measured only discretely, often without the sufficient resolution. The classic example is the way weather data is collected at a limited number of stations. In [4], the observational measurements are used in a feedback control term on the large scales of the model, enabling the recovery of the small scales. To extend the map W, the measurements $p(\cdot) = Pu(\cdot)$ for $u(\cdot) \subset \mathcal{A}$ are replaced by $v \in X = C_b(\mathbb{R}, PH)$. The determining form in [3] is then

$$\frac{dv}{d\tau} = -\gamma \sup_t \|v - PW(v)\|_H^2 (v - Pu^*), \tag{1}$$

where γ is a damping parameter and u^* is any steady state of the NSE.

It is worthwhile to compare determining forms with inertial forms. An inertial form is the restriction of an evolutionary equation to a finite dimensional, invariant, Lipschitz manifold (called an inertial manifold) [5, 6]. It follows that an inertial form is an ODE in a finite dimensional space. The catch is that some of the most important physical systems are not known to have inertial manifolds. In particular, this is still open for the 2D NSE. The theoretical significance of the determining forms described above is that even though each is a differential equation in an infinite dimensional space, it is an ODE in the true sense, as it can be shown to satisfy a global Lipschitz property. Often partial differential equations are formally written as ODEs where the phase space is a function space for the spatial variable. Due to differential operators, however, such equations do not enjoy the Lipschitz condition.

In this paper we exhibit numerical evidence that determining forms can remove noise from a coarse-grain signal. We start with a time series of a high-resolution solution on $\mathcal{A}$, take a low-resolution projection and then add a random perturbation at each time step. From this initial condition the solution to the determining form settles rather quickly to a steady state that is the original low-resolution projection. In the process, the high-resolution is recovered as well. This is illustrated first for the Lorenz system, and then demonstrated on the Kuramoto-Sivashinsky equation.

2 Preliminaries

Consider an evolution equation

$$\frac{du}{dt} = F(u), \qquad u \in H \tag{2}$$

where the initial data $u(0) = u_0$ is missing. The phase space H may be $\mathbb{R}^n$ for a finite system of coupled ODEs or, in the case of a PDE, an infinite dimensional space such as $L^2(\Omega)$, the Hilbert space of Lebesgue square-integrable functions, or a subspace thereof consistent with boundary conditions. Equation (2) is dissipative if there exists an absorbing ball. This means that that for any initial condition u_0 there exists $T = T(\|u_0\|)$ such that $u(t, u_0) \in \mathcal{B}$, for all $t \geq T$,

where $\mathcal{B}$ is an absorbing ball. Here $u(t, u_0)$ is the solution at time t starting at an initial condition u_0. The global attractor $\mathcal{A}$ can be defined as

$$\mathcal{A} = \bigcap_{t \geq 0} u(t, \mathcal{B}),$$

or equivalently as,

$$\mathcal{A} = \{u_0 |\ u(t, u_0) \text{ is bounded } \forall\ t \in \mathbb{R}\}.$$

As a consequence, $\mathcal{A}$ contains all steady states, limit cycles, and unstable manifolds of such, as well as chaotic trajectories.

Suppose $u(\cdot) \subset \mathcal{A}$ and let $Pu(\cdot)$ represent the observations of the reference solution at a coarse spatial resolution. The algorithm in [7] constructs a solution $w(\cdot)$ that is governed by

$$\frac{dw}{dt} = F(w) - \mu P(w - u), \tag{3a}$$

$$w(0) = w_0, \tag{3b}$$

where $\mu > 0$ is a relaxation parameter and w_0 can be taken to be arbitrary initial data. Under certain conditions on both μ and the spatial resolution, one has that $\|w(t) - u(t)\|_H \to 0$ at an exponential rate as $t \to \infty$ [7]. The idea is to evolve (3) over a relaxation time period $[0, T_0]$ after which $w(T_0)$ is within some tolerance of $u(T_0)$. In data assimilation the state $w(T_0)$ would then be used as an initial condition for a high resolution simulation. To compute the determining form, however, u is replaced by v, and $W(v)(t) = w(t)$ for $t \geq T_0$, i.e., $w(t)$ for $t \in [0, T_0]$ is discarded.

In this paper we carry out numerical computations using the determining form

$$\frac{dv}{d\tau} = -\gamma \sup_t \|v - PW(v)\|_H^2 (v - PW(v)). \tag{4}$$

which is chosen so that small perturbations of $u(\cdot) \in \mathcal{A}$ evolve toward $u(\cdot)$ rather u^*, as in (1). Note that there are two time variables: the time t in the original equation (2), and τ, the evolutionary variable in the determining form. For this reason, it is at times convenient to write $v = v(\tau, t)$.

3 Lorenz equation

The Lorenz equations can be written as

$$\begin{aligned} \frac{du_1}{dt} &= -u_1 - u_2 u_3 \\ \frac{du_2}{dt} &= -\sigma u_2 + \sigma u_1 \\ \frac{du_3}{dt} &= -b u_3 - br + u_2 u_3. \end{aligned} \tag{5}$$

We use the classic parameter values $\sigma = 10, b = 8/3$, and $r = 28$. After a sufficient transient period just about every solution is essentially on the well known chaotic trajectory shown in Figure 1(a). The role of the low-resolution observables is played by the single component u_1. It is easy to see that u_1 by itself is a determining mode. Indeed we can solve for u_2 in terms of $u_1(\cdot)$

$$u_2(t) = e^{\sigma(t_0-t)}u_2(t_0) + \int_{t_0}^{t} e^{\sigma(s-t)}u_1(s)ds.$$

Taking $t_0 \to -\infty$ and using the fact that $u(\cdot) \subset \mathcal{A}$ is bounded for all time, we have

$$u_2(t) = \int_{-\infty}^{t} e^{\sigma(s-t)}u_1(s)ds.$$

Similarly, u_3 can be expressed in terms $u_1(\cdot)$ and $u_2(\cdot)$, and since u_2 is determined by u_1, so is u_3.

To calibrate the relaxation time, we solve (5) simultaneously with

$$\begin{aligned} \frac{dw_1}{dt} &= -w_1 - w_2w_3 - \mu(w_1 - u_1) \\ \frac{dw_2}{dt} &= -\sigma w_2 + \sigma w_1 \\ \frac{dw_3}{dt} &= -bw_3 - br + w_2w_3 \end{aligned} \tag{6}$$

until $\|w-u\|$ is less than some acceptable tolerance. For simplicity time stepping is done by the explicit Euler method. Once the relaxation time is determined, we evolve the determining form (4) by solving (6) at each time step in τ with u_1 replaced by v. In order to optimize our results for this equation we test various values of μ. We then evolve the equation over time τ in order to see the convergence to a solution on the global attractor.

3.1 Numerical Results

We seek an optimal μ value that produces the smallest relaxation time. Experimenting with $\mu = 30, 40, 50, 100$, we found that $\mu = 40$ gave the least relaxation time (see Figure 1 (b)).

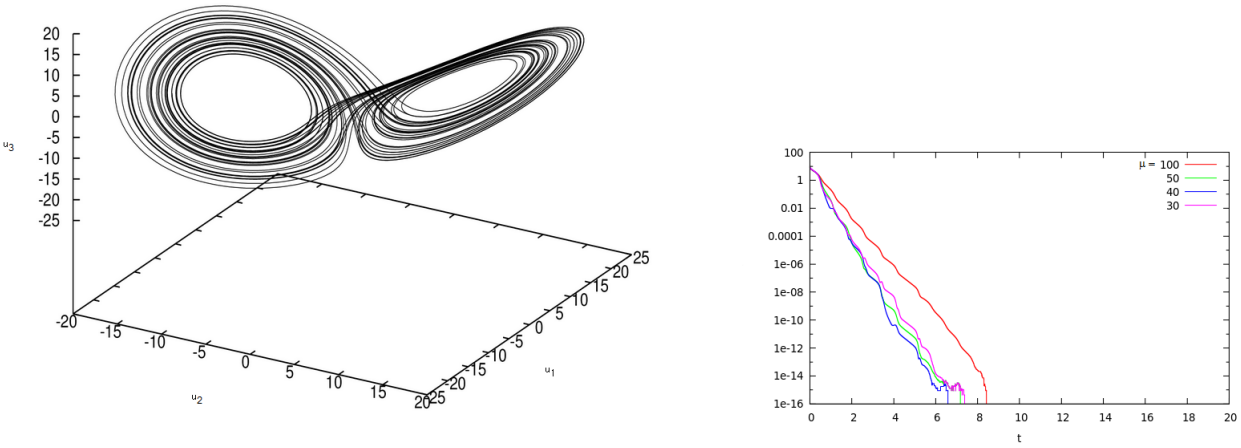


Figure 1: (a) The chaotic reference trajectory $u(\cdot)$. (b) The discrepancy $\|u-w\|$ between solutions to (5) and (6).

Thus, after a relaxation time of $T_0 = 1$ we have computed $W(p)$ to within 0.01, where $p(\cdot) = Pu(\cdot)$, and $u(\cdot)$ is the reference chaotic attractor. We assume that this same relaxation time is sufficient when we replace $p(\cdot)$ for an arbitrary $v \in X$.

We take $v(0, t_k) = u_1(t_k) + r_k$ as an initial condition for the determining form (4), for randomly generated $r_k \in (-\epsilon, \epsilon]$, and $t_k = 0, \Delta t, 2\Delta t, \ldots 10$, where Δt is the explicit Euler time step used to solve (6). This represents a noisy, coarse-grain observable. We then take explicit Euler steps for (4) with $\Delta\tau = 0.0005$. To account for the relaxation time for the $W - map$, we truncate the time interval at each step. At the first step we keep $v(\Delta\tau, t_k)$ only for $t_k \in [T_0, 10]$, at the second only for $t_k \in [2T_0, 10]$, and so on. Thus, as the solution to (4) approaches a steady state, we see in Figure 2 a progressively shorter portion of the trajectory.

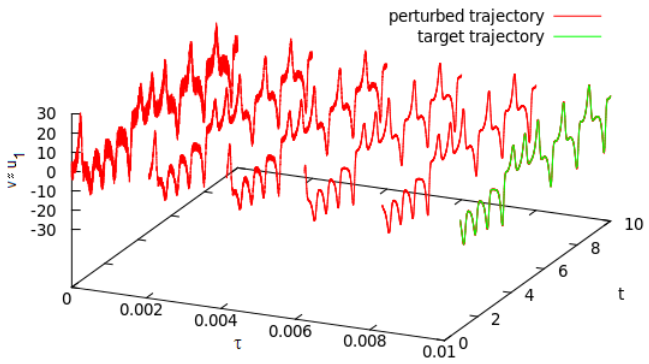


Figure 2: Evolution of the determining form of the Lorenz equation for $\gamma = 20$. The final value $\tau = 0.01$ is superposed over $v(0.01, \cdot)$.

After observing that there is indeed reduction of the noise, we wish to optimize the damping parameter γ. In Figure 3 we plot the error and note that increasing γ too much causes a deterioration in the error. The relatively large value of the error measured in this way is probably unavoidable due to the highly oscilliatory nature of the chaotic target solution.

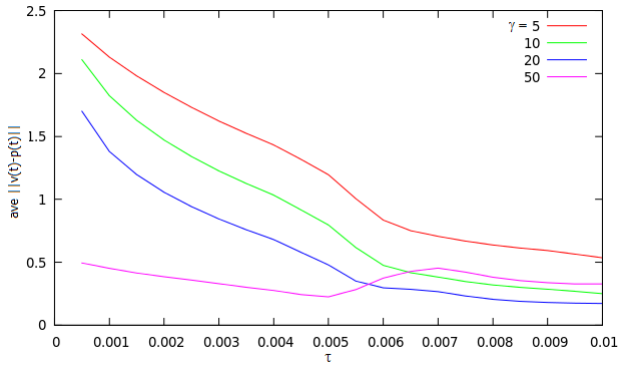


Figure 3: Average pointwise error as τ advances.

4 Kuramoto-Sivashinsky Equation

We write the Kuramoto-Sivanshinsky equation (KSE) with periodic and odd boundary conditions as

$$\frac{\partial u}{\partial t} + 4\frac{\partial^4 u}{\partial x} + \alpha\left[\frac{\partial^2 u}{\partial x} + u\frac{\partial u}{\partial x}\right] = 0 \tag{7}$$
$$u(t,x) = u(t,x+2\pi), \qquad u(t,-x) = -u(t,x).$$

This system has been derived independently to model both combustion and fluid flow down an inclined plane. Though it is known to have an inertial form [8], it is not evident how to exploit this to remove noise from a coarse-grain signal. We discretise the KSE in space using both the Galerkin method and finite difference methods. Two reference solutions in the global attractor, a limit cycle, and a chaotic orbit are first found by the Galerkin method, and then translated to nodal values. We then apply the determining form to the finite differences method.

Let P be the projector from the space of odd, periodic functions onto the span of $\sin(x), \ldots, \sin(nx)$. To obtain the Galerkin approximation we substitute

$$p = Pu = \sum_{j=1}^{n} a_j(t)\sin(jx) \tag{8}$$

into the KSE and apply the projector P.

$$\frac{dp}{dt} + Ap + PF(p) = 0$$

where

$$A = \frac{\partial^4}{\partial x^4} \qquad F(u) = \alpha\frac{\partial^2 u}{\partial x^2} + \alpha u\frac{\partial u}{\partial x}.$$

The Galerkin approximation is thus a system of ODEs for $a_1, a_2, \ldots, a_n$ with $n = 32$. These equations are used along with Euler's method to find reference solutions which are then converted to its nodal representation using (8).

This nodal representation is used as an initial condition in a finite difference scheme specified by

$$u\frac{\partial u}{\partial x}(x_j) \approx \frac{u_j(-u_{j+1} - u_{j-1}) + u_{j+1}u_{j+1} - u_{j-1}u_{j-1}}{6\Delta x}$$
$$\frac{\partial^4 u}{\partial x^4}(x_j) \approx \frac{-(u_{j+2} - 4u_{j+1} + 6u_j - 4u_{j-1} + u_{j-2})}{\Delta x^4}$$
$$\frac{\partial^2 u}{\partial x^2}(x_j) \approx -\frac{u_{j+1} - 2u_j + u_{j-1}}{\Delta x^2}.$$

where $u_j \approx u(x_j)$, $x_j = j\Delta x$. The nonlinear term is treated in this way in order to make the scheme dissipative [9]. After a sufficient transient period the

solution of the finite difference approximation, combined with Euler's method settles onto an analogue of the reference solution. The cartoon in Figure 4 depicts the nodal positions in a coarse-grain sample. In our implementation $n = 32$, and m is varied somewhat.

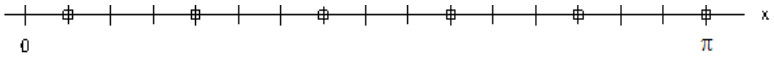


Figure 4: Here $m = \dim PH$ = the number of □'s, and $n = \dim H$ = the number of hash marks.

4.1 Numerical Results

The first reference solution for the KSE is a periodic solution at $\alpha = 32$ shown in Figure 5.

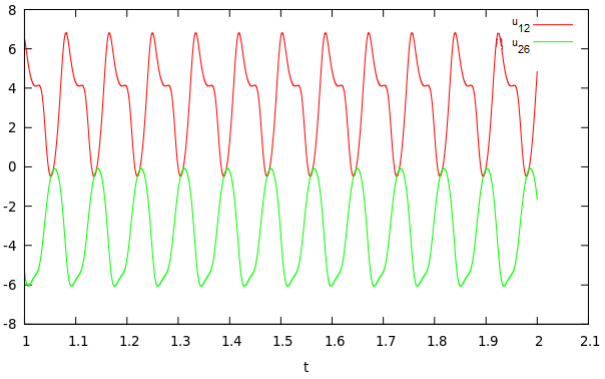


Figure 5: The periodic solution to the KSE at the 12^{th} and 26^{th} nodal positions.

We first kept μ fixed at 4000, and varied m by equally distributing $4, 8$, and 16 coarse mesh points with the result shown in Figure 6.

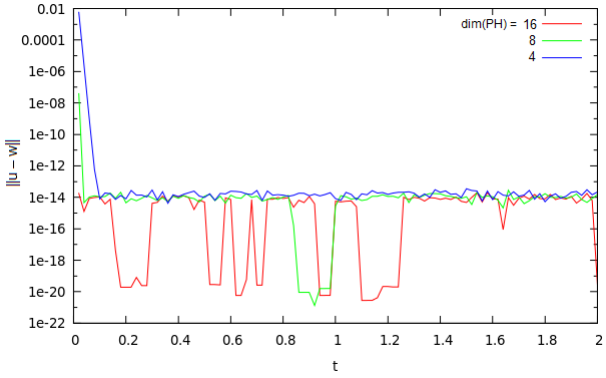


Figure 6: The discrepency $\|w - u\|$ where $u(\cdot)$ is the periodic solution to the finite difference approximation at 32 nodes, $\mu = 4000, \Delta t = .0000001$

We see that for the periodic reference solution we can use just 4 out of 32 mesh points, and achieve close to double precision accuracy after a short relaxation time. In Figure 7 we fix $m = 4$, and vary μ to find that $\mu = 4000$ works best.

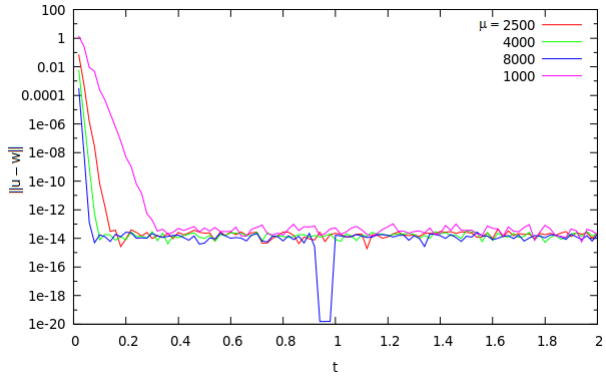


Figure 7: The discrepency $\|w - u\|$ where $u(\cdot)$ is the periodic solution to the finite difference approximation at 32 nodes, $m = 4, \Delta t = .0000001$

With the relaxation time calibrated for $m = 4$ and $\mu = 4000$, we evolve the determining form to remove noise added to the periodic reference solution (see Figure 8).

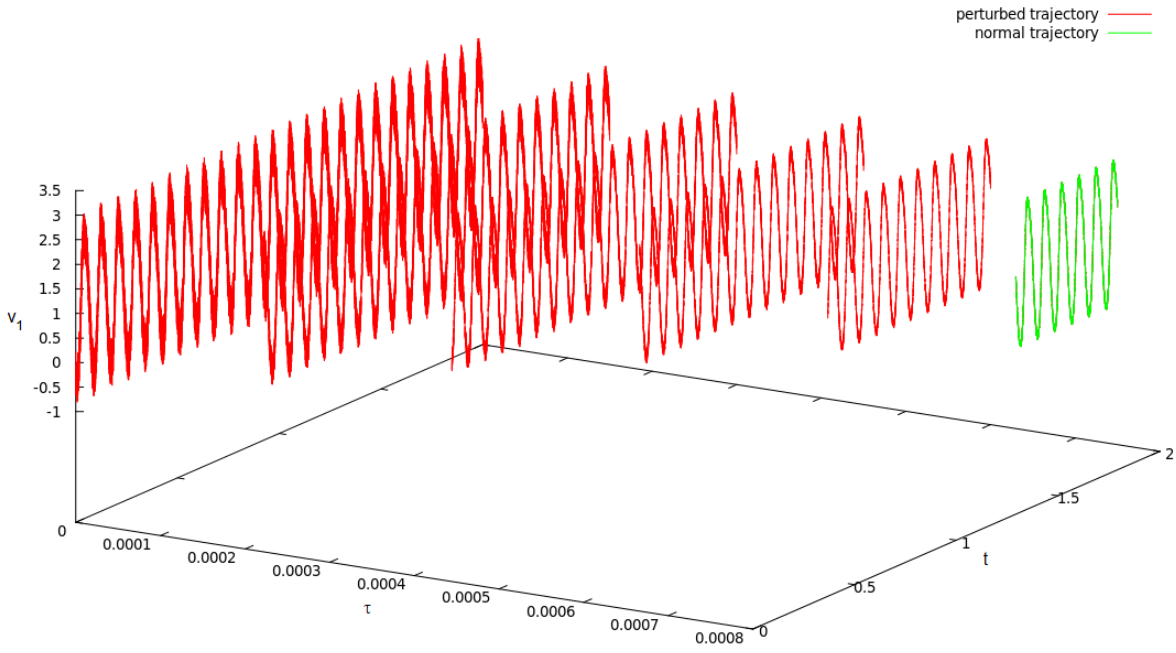


Figure 8: Evolution of the determining form solution from the perturbed initial condition. The target trajectory is superposed at the final value of $\tau = 0.0005$. Here $v_1 \to u_1 \approx u(\Delta x)$.

We then vary in Figure 9 the damping parameter γ for the KSE periodic reference solution. This leads to a γ value of 500 for our determining form.

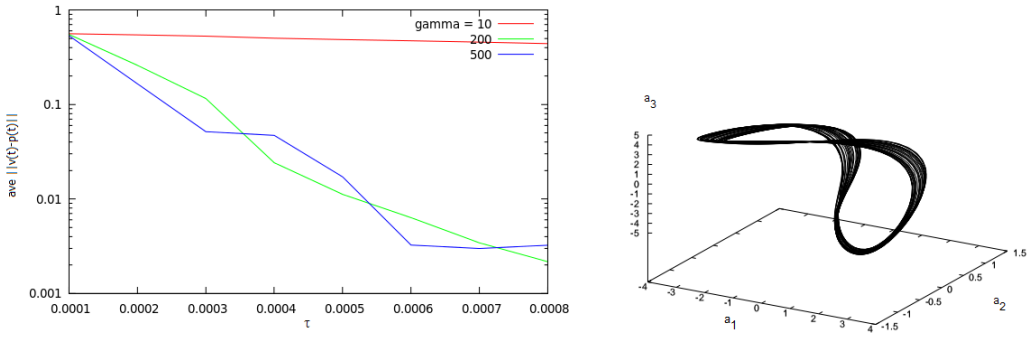


Figure 9: (a) Average pointwise error as τ advances with the logscale of the y-axis. (b) Chaotic reference solution for Galerkin approximation.

Next we consider a chaotic reference solution of the KSE. It has been shown that there is such a solution at $\alpha = 133.73454$ in [10]. This is shown for the first few Galerkin modes in Figure 9.

We then proceed as for the periodic reference solution, translating to nodal values, running through a transient period to settle on an initial chaotic trajectory for the finite difference approximation (see Figure 10).

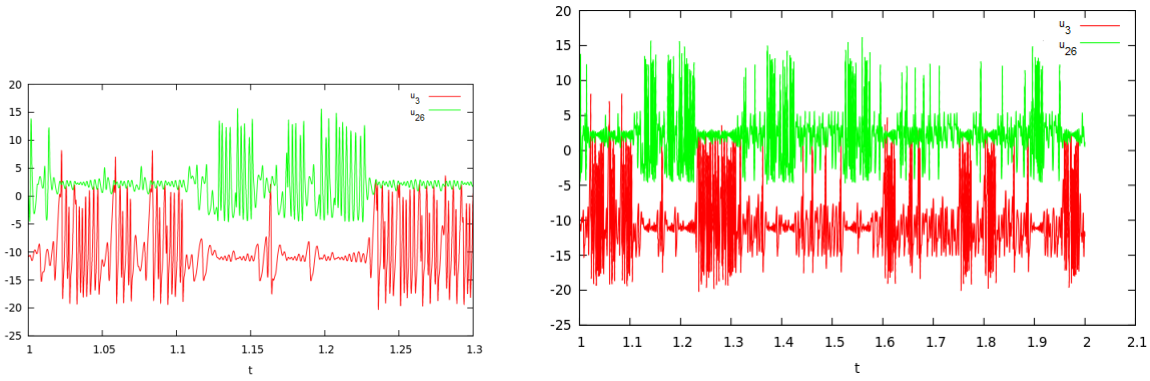


Figure 10: (a) The chaotic solution to the KSE at the 3^{rd} and 26^{th} nodal positions on a decreased interval to show the occilatory nature of the solution. (b) The same chaotic solution on the entire interval from $t = [1, 2]$.

We then fixed μ at 4000, and varied m by equally distributing 8 and 16 coarse mesh point with the result shown in Figure 11

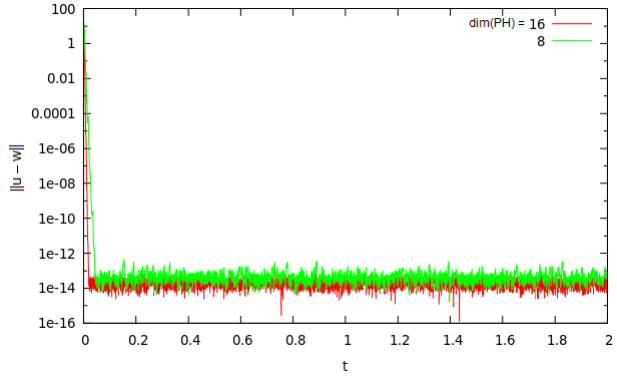


Figure 11: The discrepency $\|w - u\|$ where $u(\cdot)$ is the chaotic solution to the finite difference approximation at 32 nodes, $\mu = 4000, \Delta t = .000001$

We find that for the chaotic reference solution we must use 8 out of 32 mesh points, rather than 4, and achieve close to double precision accuracy after a short relaxation time. Using 8 mesh points seems reasonable due to the excessive number of oscillations. In Figure 12 we fix $m = 8$ and vary μ to find that $\mu = 4000$ works best.

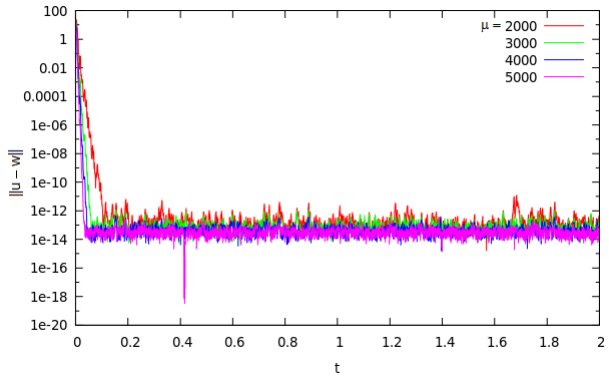


Figure 12: The discrepency $\|w - u\|$ where $u(\cdot)$ is the chaotic solution to the finite difference approximation at 32 nodes, $m = 4, \Delta t = .000001$

With the relaxation time calibrated for $m = 8$ and $\mu = 4000$, we evolve the determining form to remove noise added to the chaotic reference solution (see Figure 13).

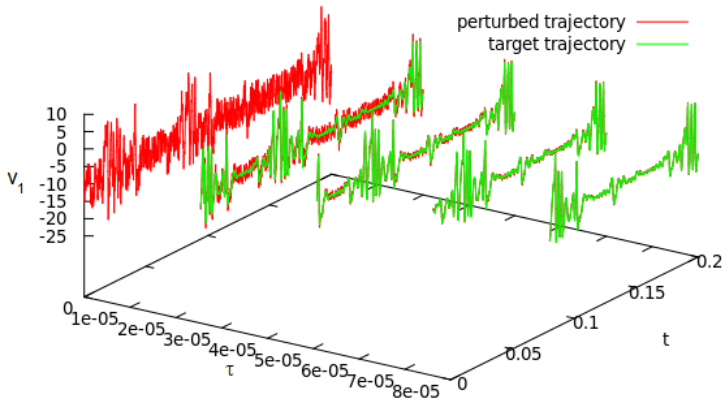


Figure 13: Evolution of the determining form solution from the perturbed initial condition. The target trajectory is superposed. Here $v_1 \to u_1 \approx u(\Delta x)$.

We then vary in Figure 14 the damping parameter γ for the KSE chaotic reference solution. This leads to a γ value of 200 for our determining form.

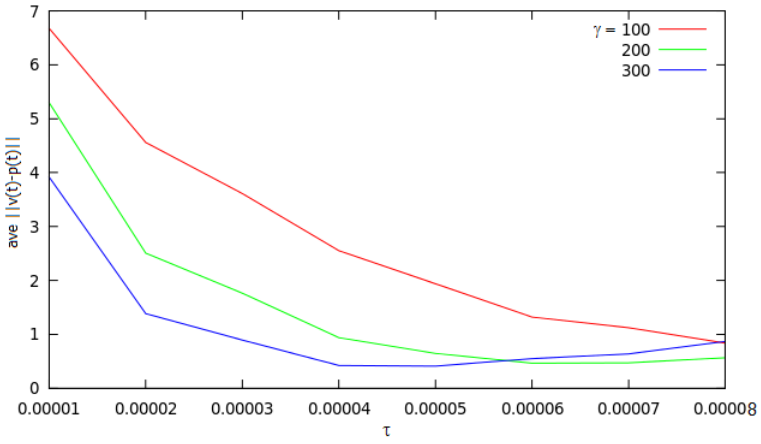


Figure 14: Average pointwise error as τ advances.

5 Discussion & Conclusion

We have shown that the detemining form can reduce noise in a coarse-grain signal. In Figures 13, 8 and 2 we can see the perturbation decrease as we advance

in τ. We are able to optimize somewhat our parameter values to enhance the rate of noise reduction. The proper values of μ and γ are chosen for both the Lorenz and the Kuramoto-Sivashinsky cases. The determining form is a large system of equations, as it is in a space of trajectories. Determining parameters for PDEs can be nodal values at sparse spatial positions, coefficients in Fourier expansions, or others, such as finite elements. One can take a relatively large step, $\Delta\tau$ when evolving the determining form. We continue to work on how to deal with the relaxation time.

Acknowledgments

I would like to thank Professor Mike Jolly of University of Indiana for hhis guidance, knowledge, and support. I would also like to thank Professor Chris Connell for organizing the REU program during which this report was written, as well as all of the other REU students for such a great experience.

References

[1] C. Foias and G. Prodi. Sur le comportement global des solutions non-stationnaires des equations de Navier-Stokes en dimension 2. *Rend. Sem. Mat. Univ. Padova*, 39:1–34, 1967.

[2] C. Foias, M. S. Jolly, R. Kravchenko, and E. S. Titi. A determining form for the two-dimensional Navier-Stokes equations: the Fourier modes case. *J. Math. Phys.*, 53(11):115623, 30, 2012.

[3] C. Foias, M. S. Jolly, R. Kravchenko, and E. S. Titi. A unified approach to determining forms for the 2-D Navier -Stokes equations–the general interpolants case. 69(2):359–381, 2014.

[4] A. Azouani and E. S. Titi. Feedback control of nonlinear dissipative systems by finite determining parameters – a reaction-diffusion paradigm. page 21, 2013.

[5] C. Foias, G. R. Sell, and E. S. Titi. Exponential tracking and approximation of inertial manifolds for dissipative nonlinear equations. *J. Dynam. Differential Equations*, 1(2):199–244, 1989.

[6] R. Temam. *Infinite-dimensional dynamical systems in mechanics and physics*, volume 68 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1997.

[7] A. Azouani, E. Olson, and E. S. Titi. Continuous data assimilation using general interpolant observables. *J. Nonlinear Sci.*, 24(2):277–304, 2014.

[8] C. Foias, B. Nicolaenko, G. R. Sell, and R. Temam. Inertial manifolds for the Kuramoto-Sivashinsky equation and an estimate of their lowest dimension. *J. Math. Pures Appl. (9)*, 67(3):197–226, 1988.

[9] C. Foias, M. S. Jolly, I. G. Kevrekidis, and E. S. Titi. Dissipativity of numerical schemes. *Nonlinearity*, 4(3):591–613, 1991.

[10] F. Christiansen, P. Cvitanovic, and V. Putkaradze. Spatiotemporal chaos in terms of unstable recurrent patterns. *Nonlinearity*, 10(1):55–70, 1997.

[11] M. E. Johnson, M. S. Jolly, and I. G. Kevrekidis. The Oseberg transition: visualization of global bifurcations for the Kuramoto-Sivashinsky equation. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 11(1):1–18, 2001.

[12] C. Foias and E. S. Titi. Determination of the solutions of the navier–stokes equations by a set of nodal values. *Math. Comp.*, 43(167):117–133, 1984.

[13] B. Cockburn, D. Jones, and E. S. Titi. Estimating the number of asymptotic degrees of freedom for nonlinear dissipative systems. *Math. Comp.*, 66(219):1073–1087, 1997.

Configuration Space of Eight-Bar Linkages

Kathryn Marsh

Abstract

A closed planar n-bar linkage can be regarded as a collection of n vectors in the plane whose sum equals zero. The set of all closed n-bar linkages is a subset of $\mathbb{R}^{2n}$ and hence has a natural topology. I will describe the subset of unit-length, 8-bar linkages with a packing radius of $\frac{1}{2}$. A linkage of this type is the minimal vector decomposition of a well-rounded translation surface of genus 2 with one singularity of 6π. The space has a natural decomposition as a 3-dimensional cell complex with relations between the cells described via a cutting and pasting identification. I will discuss the details of this decomposition and the resulting cell complex.

1 Linkages and Translation Surfaces

The set of well-rounded translation surfaces of genus 2 with a 6π singularity, can be thought of in terms of linkages in the following way. Firstly, we call a linkage an n-bar linkage if it has n distinct, rigid parts. An n-bar linkage can be represented by an ordered set of n vectors. An n-bar linkage is closed if $\sum_{i=1}^{n} v_i = 0$. We will consider closed 8-bar linkages together with a matching of the vectors. In particular, each vector v_i is matched with a unique and distinct 'partner' vector v_j. In addition, we assume that each vector is parallel to its partner but with opposite direction, that is $v_i + v_j = 0$. Each such linkage with a matching scheme can be uniquely described by four vectors (v_1, v_2, v_3, v_4) and a skipping pattern between partners. The number of skips is given by $(n_1, n_2, n_3, n_4) \in \mathbb{N}^4$ so that v_i is partnered with v_{i+n_i+1}. For example, the linkage that corresponds to an octagon with opposite sides acting as partners would have a distance pattern of (3,3,3,3) and vectors (v_1, v_2, v_3, v_4). For simplicity, we will call linkages by listing the order of the vectors as they would appear reading counterclockwise around the octagon. Therefore the linkage with a distance pattern of (3,3,3,3) will be represented by the 8 letter "word" 12341234.

Let us add the following restrictions to our set of linkages:

1. Vector partners cannot be adjacent.
2. Pairs of vectors cannot show up in the same word in a different order. This amounts to excluding words such as 12342134 since we have v_1 followed by v_2 and later encounter v_2 followed by v_1.
3. Vertices cannot be closer than distance 1 to each other.

By taking these restrictions into consideration, we end up with a set of linkages which are all the minimal vector decomposition of a well-rounded translation surface of genus 2 with a conical singularity of 6π.

Proposition 1.1 *Let L be the ordered set of unit vectors $(v_1, v_2, ...v_8)$ such that L is the minimal vector decomposition of a well-rounded translation surface of genus 2 with a conical singularity of 6π. Each v_i in L has a partner v_j in L such that $v_i = -v_j$. Then up to relabeling, there exist three distinct arrangements of partner vectors.*

Proof To begin, we will make use of the notation using 'words' and skipping patterns as established above. The minimal distance between any element of $\{1, 2, 3, 4\}$ and an element $\{5, 6, 7, 8\}$ in $\mathbb{Z}/8\mathbb{Z}$ is 3. Hence the set of possible arrangements of skipping patterns is a subset of $\{1, 2, 3\}^4$. Therefore the maximum possible number of distinct distance patterns is $3*3*3*3 = 81$. Since this is a rather large set to check all at once, we will break it up into three subsets of size 27 which are the skipping patterns where $n_1 = 1$, $n_1 = 2$, and $n_1 = 3$.

Let's start with the set of distance patterns that have $n_1 = 1$ first. This corresponds to a word of the form 1*1*****. If $n_2 = 1$ we immediately violate the condition that vertices have distance greater or equal to 1 from each other. Therefore we are down to only 18 possibilities with $n_1 = 1$.

Now examine the first 9 corresponding to $(1, 2, n_3, n_4)$. Immediately, $n_3 = 2$ and $n_4 = 1$ because any other arrangement either forces a distance to be 0 (which violates that vector pairs cannot be adjacent) or cannot be made without calling two vectors by different names. Figure 1 shows how this octagon arises.

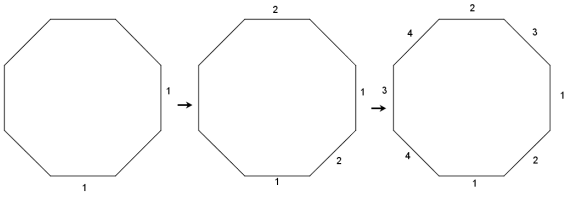


Figure 1: Labeling an octagon

Now we examine what happens when $n_2 = 3$. When $n_2 = 3$, n_3 automatically cannot be 1 because the second 2 is in the way. If $n_3 = 2$ then n_4 and if $n_3 = 3$ then $n_4 = 1$. So we have the two new patterns of $(1, 3, 2, 2)$ and $(1, 3, 3, 1)$. These correspond to the words 12134234 and 12134243 respectively. However, 12134243 violates the restriction that pairs of vectors cannot show up in the same word in a different order because we have a 34 followed by a 43. So $(1, 3, 3, 1)$ is not a valid pattern for our purposes and we have found only two patterns, $(1, 2, 2, 1)$ and $(1, 3, 2, 2)$ thus far.

Now we look at the set of distance patterns where $n_1 = 2$. If $n_1 = 2$ then n_2 cannot be 1. So we only need to examine the cases for $(2, 2, n_3, n_4)$ and $(2, 3, n_3, n_4)$. Looking at the cases for $(2, 2, n_3, n_4)$ we notice that n_3 cannot be equal to 1 because of the placement of the second 2. Similarly, n_3 cannot be equal to 2 because then n_4 would have to equal 0. Therefore the only choice is $(2, 2, 3, n_4)$ which then forces $n_4 = 1$. This corresponds to the word 12312434. However, this is only a relabeling of the word 12134234 we discovered earlier so does not represent a distinct type.

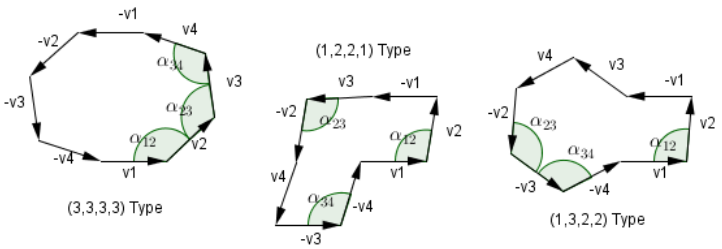
If we investigate the case of $(2, 3, n_3, n_4)$ we note that n_3 cannot be 1 or it would force n_4 to be 0. Likewise, n_3 cannot be 2 because of the placement of the second 2. Therefore, $n_3 = 3$ which forces $n_4 = 2$. This results in the word 12314234 which violates our restriction because we have a 14 followed by a 41.

Moving on to the last category where $n_1 = 3$, we let $n_2 = 1$ and see that the only viable candidate is for $(3, 1, 3, 1)$. This is a relabeling of a previous distance pattern and fails by the same reason. If $n_1 = 3$ then n_2 cannot be 2 because of the placement of the second 1. In the category of $(3, 3, n_3, n_4)$ we see that n_3 cannot be 1 or 2 because of the placement of the second 1 and 2. So $n_3 = 3$ which forces $n_4 = 3$. So we have a new unique type of $(3, 3, 3, 3)$ which corresponds to the word 12341234. Therefore, there are only three distinct arrangements for the four partner vectors corresponding to distance patterns $(1, 2, 2, 1), (1, 3, 2, 2)$, and $(3, 3, 3, 3)$ with respective words 12132434, 12134234, and 12341234.

□

2 Configuration Space

We will now build the configuration space for each type of linkage. As each linkage is comprised of four distinct vectors, each linkage has four degrees of freedom and therefore the configuration space is a subset of $\mathbb{R}^4$. In order to visualize the configuration space in 3-dimensions, we will mod out by rotations by fixing $v_1 = (1, 0)$. Let α_{ij} be the interior angle between vectors v_i and v_j. Each linkage can be uniquely determined (up to rotation) by three interior angles so we set up our coordinate system such that $(x, y, z) = (\alpha_{12}, \alpha_{23}, \alpha_{34})$.

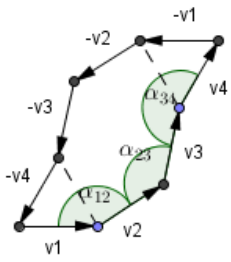
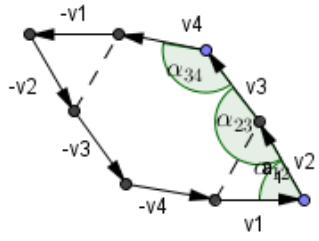


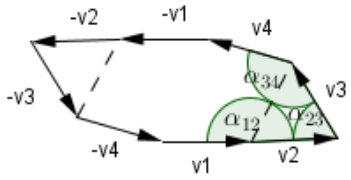
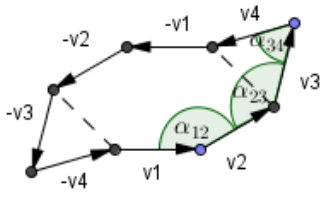
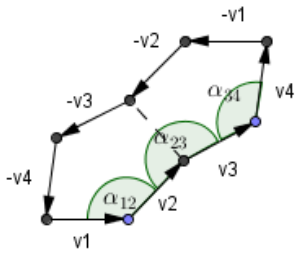
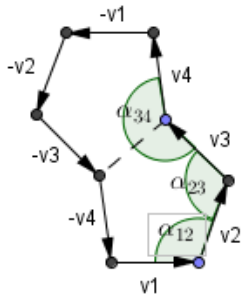
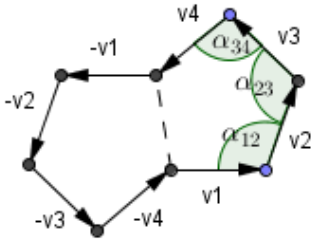
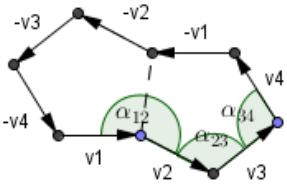
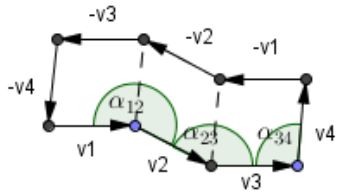
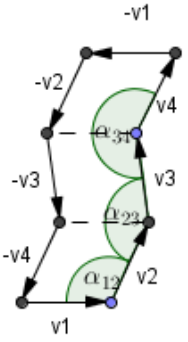
Let our configuration space for linkage type (3,3,3,3) mod rotation be called P_0, type (1,3,2,2) P_1, and type (1,2,2,1) P_2. P_k is the intersection of half spaces given by the boundaries of the linkage and is a subset of $\mathbb{R}^3$. The interior of each configuration space is determined by the restriction that the minimal interior distances between any two vertices be greater than 1.

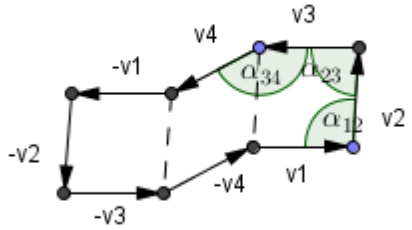
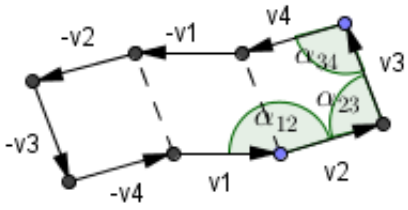
The boundary of a configuration space will be the set of linkages which have some interior distances between vertices equal to 1 such that the degree of freedom of that linkage is less than 3. The boundary can be divided into linkages with 2 degrees of freedom, 1 degree of freedom, and 0 degrees of freedom. These correspond to 2-cells (faces), 1-cells (edges), and 0-cells (vertices). For reference, a pictorial description of each face is provided within each section. In each linkage picture, the dashed lines denote a distance of 1 which is fixed.

2.1 P_0

P_0 consists of 12 faces, 30 edges, and 20 vertices. The faces are shown below in the first table.

Name	Linkage	Name	Linkage
F1		F2	

F3		F4	
F5		F6	
F7		F8	
F9		F10	

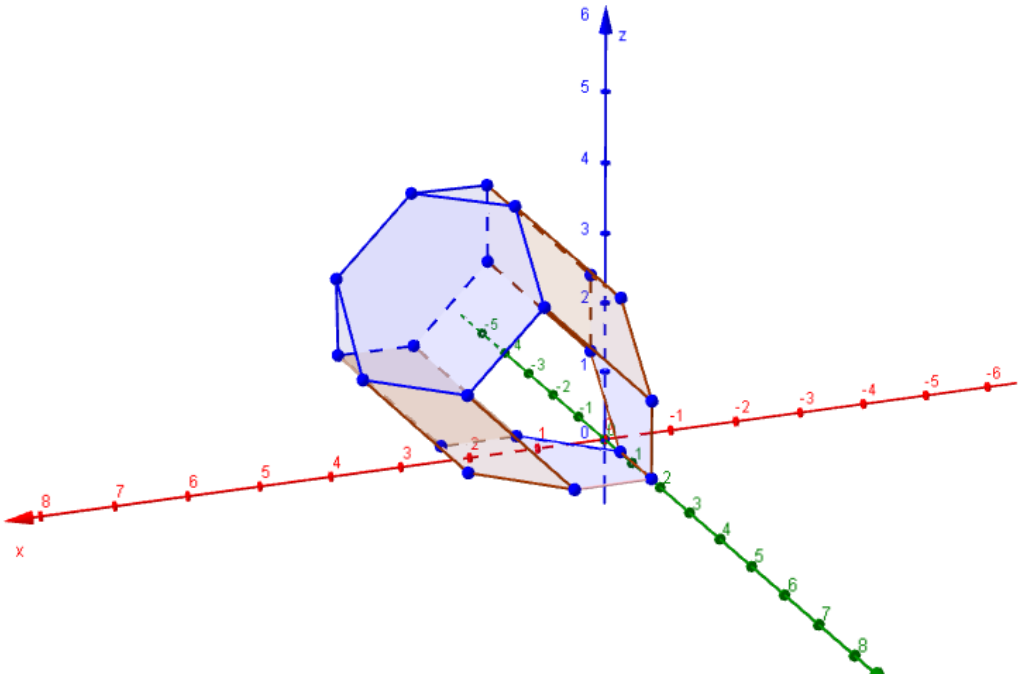
F11		F12	

The table below gives equations for each face and associated edges and vertices.

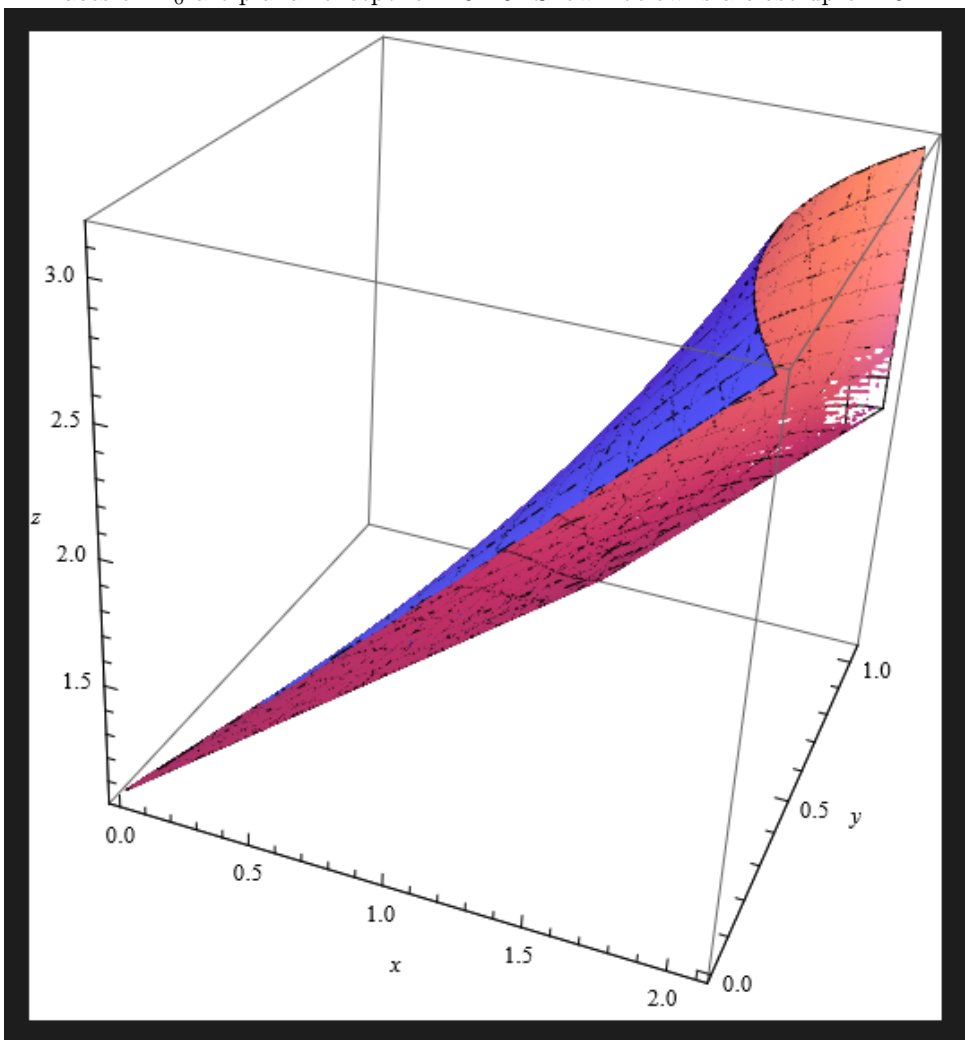
Face	Edges	Vertices	Equation
F1	E10, E14, E15, E19, E23, E27	V2, V4, V13, V8, V12, V17	$\alpha_{12}+\alpha_{23}+\alpha_{34}=8\pi/3$
F2	E13, E16, E24, E9, E28, E20	V11, V7, V14, V3, V1, V18	$\alpha_{12}=\pi/3$
F3	E12, E17, E25, E10, E29, E21	V10, V6, V15, V2, V4, V19	$\alpha_{23}=\pi/3$
F4	E9, E30, E22, E11, E18, E26	V1, V3, V20, V9, V5, V16	$\alpha_{34}=\pi/3$
F5	E8, E28, E26, E1, E14	V1, V16, V12, V8, V18	$\cos(\alpha_{23})-\cos(\alpha_{23}+\alpha_{34})-\cos(\alpha_{34})-\cos(\alpha_{12})-\cos(\alpha_{12}+\alpha_{23})-\cos(\alpha_{12}+\alpha_{23}+\alpha_{34})=-3/2$
F6	E13, E7, E29, E23, E4	V4, V19, V7, V11, V13	$\cos(\alpha_{23})+\cos(\alpha_{23}-\alpha_{34})-\cos(\alpha_{34})+\cos(\alpha_{12})-\cos(\alpha_{12}+\alpha_{23})+\cos(\alpha_{12}+\alpha_{23}+\alpha_{34})=3/2$
F7	E6, E30, E24, E3, E12	V3, V14, V10, V6, V20	$-\cos(\alpha_{23})+\cos(\alpha_{23}-\alpha_{34})-\cos(\alpha_{34})-\cos(\alpha_{12})+\cos(\alpha_{12}+\alpha_{23})+\cos(\alpha_{12}+\alpha_{23}+\alpha_{34})=-3/2$
F8	E5, E11, E2, E25, E27	V5, V9, V15, V2, V17	$\cos(\alpha_{23})-\cos(\alpha_{23}-\alpha_{34})+\cos(\alpha_{34})-\cos(\alpha_{12})+\cos(\alpha_{12}+\alpha_{23})+\cos(\alpha_{12}+\alpha_{23}+\alpha_{34})=3/2$
F9	E1, E18, E5, E19	V12, V16, V5, V17	$\alpha_{12}+\alpha_{23}=2\pi$
F10	E4, E20, E8, E15	V11, V13, V8, V18	$\alpha_{23}+\alpha_{34}=2\pi$
F11	E21, E3, E16, E7	V19, V10, V14, V7	$\alpha_{12}+\alpha_{23}=\pi$
F12	E6, E17, E2, E22	V6, V15, V9, V20	$\alpha_{23}+\alpha_{34}=\pi$

The table below gives α_{ij} coordinates for each labeled vertex. By using this information concatenated with the tables above, one can construct all the appropriately labeled edges and faces and arrive at the same P_0 shown below.

Vertices	Coordinates
V1	$(\pi/3, 4\pi/3, \pi/3)$
V2	$(4\pi/3, \pi/3, \pi)$
V3	$(\pi/3, \pi, \pi/3)$
V4	$(\pi, \pi/3, 4\pi/3)$
V5	$(\pi, \pi, \pi/3)$
V6	$(\pi, \pi/3, 2\pi/3)$
V7	$(\pi/3, 2\pi/3, \pi)$
V8	$(2\pi/3, \pi, \pi)$
V9	$(\pi, 2\pi/3, \pi/3)$
V10	$(2\pi/3, \pi/3, \pi)$
V11	$(\pi/3, \pi, \pi)$
V12	$(\pi, \pi, 2\pi/3)$
V13	$(2\pi/3, 2\pi/3, 4\pi/3)$
V14	$(\pi/3, 2\pi/3, 2\pi/3)$
V15	$(4\pi/3, \pi/3, 2\pi/3)$
V16	$(2\pi/3, 4\pi/3, \pi/3)$
V17	$(4\pi/3, 2\pi/3, 2\pi/3)$
V18	$(\pi/3, 4\pi/3, 2\pi/3)$
V19	$(2\pi/3, \pi/3, 4\pi/3)$
V20	$(2\pi/3, 2\pi/3, \pi/3)$

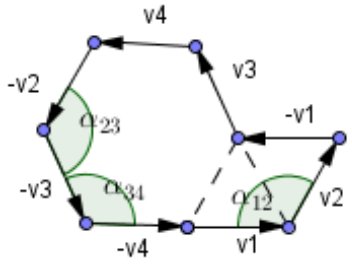
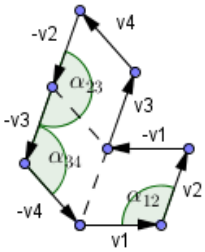
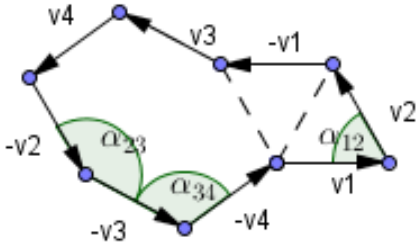
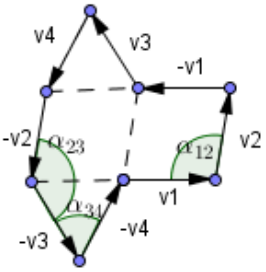
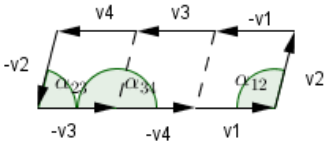
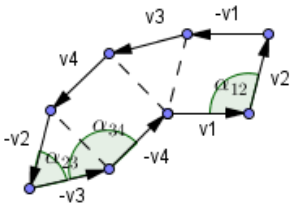
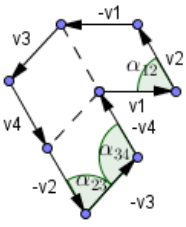
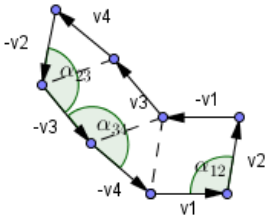


All faces on P_0 are planar except for F5-F9. Shown below is a close-up of F5.



2.2 P_1

P_1 has 8 faces, 18 edges, and 12 vertices. Below is a tabular description of the cell.

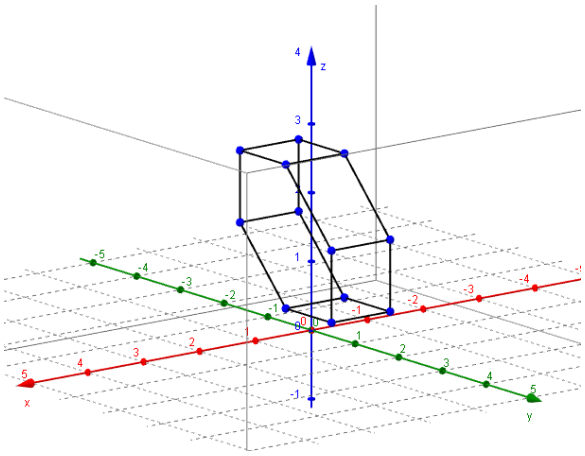
Name	Linkage	Name	Linkage
$F_1 1$		$F_1 5$	
$F_1 2$		$F_1 6$	
$F_1 3$		$F_1 7$	
$F_1 4$		$F_1 8$	

The table below gives equations for each face and associated edges and vertices.

Face	Edges	Vertices	Equation
$F_1 1$	$E_1 3, E_1 5, E_1 7, E_1 8, E_1 15, E_1 17$	$V_1 5, V_1 9, V_1 1, V_1 3, V_1 10, V_1 6$	$\alpha_{12} = 2\pi/3$
$F_1 2$	$E_1 4, E_1 6, E_1 9, E_1 10, E_1 16, E_1 18$	$V_1 2, V_1 4, V_1 12, V_1 7, V_1 8, V_1 11$	$\alpha_{12} = \pi/3$
$F_1 3$	$E_1 1, E_1 2, E_1 3, E_1 4$	$V_1 1, V_1 2, V_1 3, V_1 4$	$\alpha_{34} = \pi$
$F_1 4$	$E_1 7, E_1 9, E_1 11, E_1 13$	$V_1 5, V_1 9, V_1 11, V_1 8$	$\alpha_{23} + \alpha_{34} = \pi$
$F_1 5$	$E_1 8, E_1 10, E_1 12, E_1 14$	$V_1 6, V_1 7, V_1 10, V_1 12$	$\alpha_{23} = \pi$
$F_1 6$	$E_1 5, E_1 6, E_1 11, E_1 12$	$V_1 5, V_1 6, V_1 7, V_1 8$	$\alpha_{34} = \pi/3$
$F_1 7$	$E_1 1, E_1 13, E_1 15, E_1 16$	$V_1 1, V_1 2, V_1 9, V_1 11$	$\alpha_{23} = \pi/3$
$F_1 8$	$E_1 2, E_1 14, E_1 17, E_1 18$	$V_1 3, V_1 4, V_1 10, V_1 12$	$\alpha_{23} + \alpha_{34} = 5\pi/3$

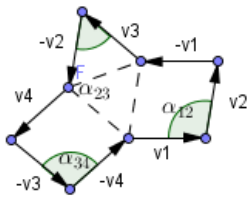
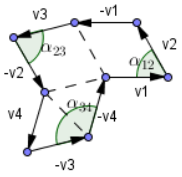
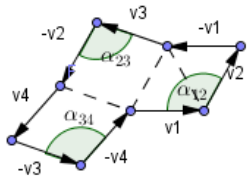
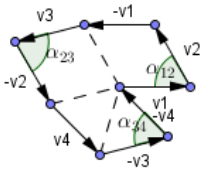
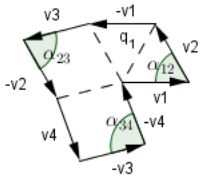
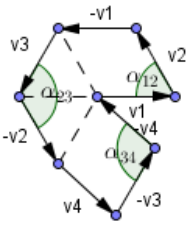
The table below gives α_{ij} coordinates for each labeled vertex. By using this information concatenated with the tables above, one can construct all the appropriately labeled edges and faces and arrive at the same P_1 shown below.

Vertices	Coordinates
$V_1 1$	$(2\pi/3, \pi/3, \pi)$
$V_1 2$	$(\pi/3, \pi/3, \pi)$
$V_1 3$	$(2\pi/3, 2\pi/3, \pi)$
$V_1 4$	$(\pi/3, 2\pi/3, \pi)$
$V_1 5$	$(2\pi/3, 2\pi/3, \pi/3)$
$V_1 6$	$(2\pi/3, \pi, \pi/3)$
$V_1 7$	$(\pi/3, \pi, \pi/3)$
$V_1 8$	$(\pi/3, 2\pi/3, \pi/3)$
$V_1 9$	$(2\pi/3, \pi/3, 2\pi/3)$
$V_1 10$	$(2\pi/3, \pi, 2\pi/3)$
$V_1 11$	$(\pi/3, \pi/3, 2\pi/3)$
$V_1 12$	$(\pi/3, \pi, 2\pi/3)$



2.3 P_2

P_2 has 6 faces, 12 edges, and 8 vertices. Below is a tabular description of the cell.

Name	Linkage	Name	Linkage
$F_2 1$		$F_2 4$	
$F_2 2$		$F_2 5$	
$F_2 3$		$F_2 6$	

The table below gives equations for each face and associated edges and vertices.

Face	Edges	Vertices	Equation
$F_2 1$	$E_2 1, E_2 2, E_2 3, E_2 4$	$V_2 1, V_2 2, V_2 3, V_2 4$	$\alpha_{23} = \pi/3$
$F_2 2$	$E_2 1, E_2 5, E_2 6, E_2 9$	$V_2 1, V_2 2, V_2 5, V_2 6$	$\alpha_{12} = 2\pi/3$
$F_2 3$	$E_2 2, E_2 7, E_2 8, E_2 11$	$V_2 3, V_2 4, V_2 7, V_2 8$	$\alpha_{12} = \pi/3$
$F_2 4$	$E_2 4, E_2 6, E_2 8, E_2 10$	$V_2 1, V_2 4, V_2 6, V_2 8$	$\alpha_{34} = 2\pi/3$
$F_2 5$	$E_2 3, E_2 5, E_2 7, E_2 12$	$V_2 2, V_2 3, V_2 5, V_2 7$	$\alpha_{34} = \pi/3$
$F_2 6$	$E_2 9, E_2 10, E_2 11, E_2 12$	$V_2 5, V_2 6, V_2 7, V_2 8$	$\alpha_{23} = 2\pi/3$

The table below gives α_{ij} coordinates for each labeled vertex. By using

this information concatenated with the tables above, one can construct all the appropriately labeled edges and faces and arrive at the same P_2 shown below.

Vertices	Coordinates
$V_2 1$	$(2\pi/3, \pi/3, 2\pi/3)$
$V_2 2$	$(2\pi/3, \pi/3, \pi/3)$
$V_2 3$	$(\pi/3, \pi/3, \pi/3)$
$V_2 4$	$(\pi/3, \pi/3, 2\pi/3)$
$V_2 5$	$(2\pi/3, 2\pi/3, \pi/3)$
$V_2 6$	$(2\pi/3, 2\pi/3, 2\pi/3)$
$V_2 7$	$(\pi/3, 2\pi/3, \pi/3)$
$V_2 8$	$(\pi/3, 2\pi/3, 2\pi/3)$

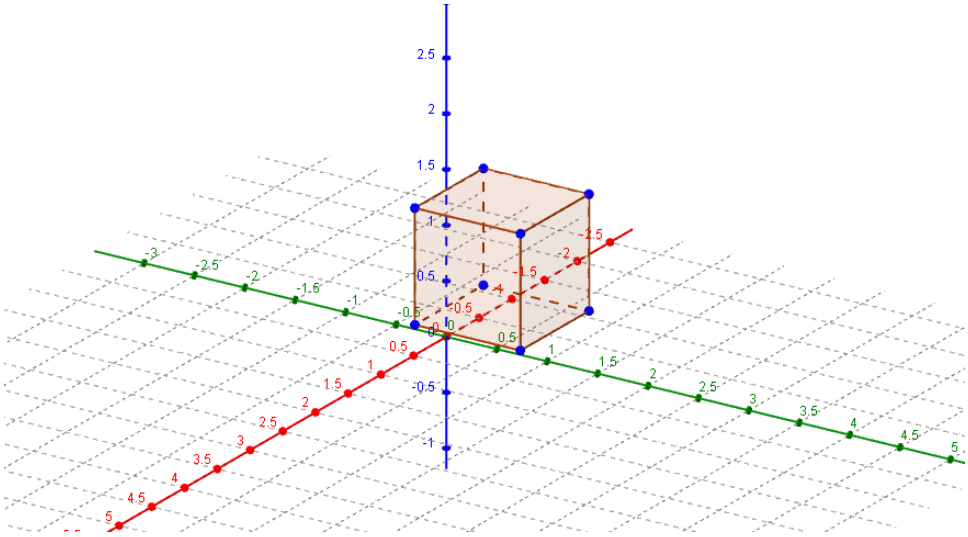
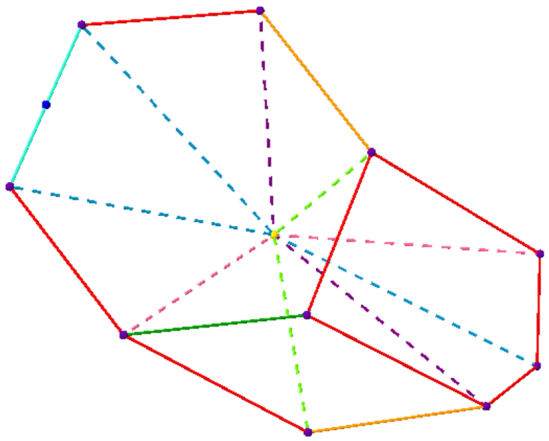


Figure 2: (1,2,2,1) type cell in α_{ij} coordinates

2.4 Fundamental Domain of P_0

Due to the nature of the hyperelliptic symmetry within the (3,3,3,3) type linkages, the configuration space previously describes consists of some overcount. The reason for overcount is because we would like to ignore differences which arise from rotation (hence our reasoning for setting $v_1 = (1, 0)$) and as can be seen, faces F1-F4 are merely rotations and relabellings of each other and similarly for F5-F8 and F9-F12. Therefore we arrive at a fundamental domain for the P_0 which is the intersection of the equations given below.

Face	Edges	Vertices	Equation
F1	E10, E14, E15, E19, E23, E27	V2, V4, V13, V8, V12, V17	$\alpha_{12}+\alpha_{23}+\alpha_{34} \leq 8\pi/3$
F5	E8, E28, E26, E1, E14	V1, V16, V12, V8, V18	$\cos(\alpha_{23})-\cos(\alpha_{23}+\alpha_{34})-\cos(\alpha_{34})-\cos(\alpha_{12})-\cos(\alpha_{12}+\alpha_{23})-\cos(\alpha_{12}+\alpha_{23}+\alpha_{34}) \geq -3/2$
F9	E1, E18, E5, E19	V12, V16, V5, V17	$\alpha_{12}+\alpha_{23} \leq 2\pi$
I1	E31, E32	V8, V13, V21	$6\alpha_{12}+\alpha_{23}+\alpha_{34} > 6\pi$
I2	E32, E33	V8, V18, V21	$3\alpha_{12}+2\alpha_{23}-\alpha_{34} > 3\pi$
I3	E33, E34	V18, V1, V21	$7\alpha_{12}+5\alpha_{23} > 9\pi$
I4	E34, E35	V1, V16, V21	$5\alpha_{23}+7\alpha_{34} > 9\pi$
I5	E35, E36	V16, V5, V21	$5\alpha_{12}+5\alpha_{23}+6\alpha_{34} > 12\pi$
I6	E36, E37	V5, V17, V21	$\alpha_{12}+4\alpha_{23}+3\alpha_{34} > 6\pi$
I7	E37, E38	V2, V17, V21	$2\alpha_{12}+7\alpha_{23}+7\alpha_{34} > 12\pi$
I8	E38, E39	V2, V4, V21	$\alpha_{12}+2\alpha_{23}+\alpha_{34} > 3\pi$
I9	E39, E32	V4, V13, V21	$7\alpha_{12}+7\alpha_{23}+2\alpha_{34} > 12\pi$



Vertices	Coordinates
V1	$(\pi/3, 4\pi/3, \pi/3)$
V2	$(4\pi/3, \pi/3, \pi)$
V4	$(\pi, \pi/3, 4\pi/3)$
V5	$(\pi, \pi, \pi/3)$
V8	$(2\pi/3, \pi, \pi)$
V12	$(\pi, \pi, 2\pi/3)$
V13	$(2\pi/3, 2\pi/3, 4\pi/3)$
V16	$(2\pi/3, 4\pi/3, \pi/3)$
V17	$(4\pi/3, 2\pi/3, 2\pi/3)$
V18	$(\pi/3, 4\pi/3, 2\pi/3)$
V21	$(3\pi/4, 3\pi/4, 3\pi/4)$
V22	$(7\pi/6, \pi/3, 7\pi/3)$

3 Moduli Space

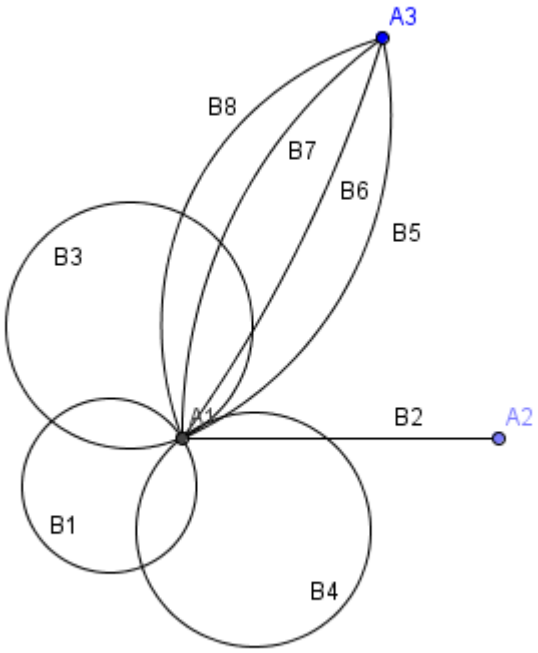
We would now like to examine the moduli space of well-rounded translation surfaces of genus 2 with a conical singularity of 6π. Since we are now considering these linkages as surfaces, we will need to identify linkages which are currently considered separate but which represent the same surface. In order to see which

linkages are equivalent as surfaces, we will make all possible identifications of sides by between any vertices with distance and pasting parallel vector pairs together. By performing this cutting and pasting identification process on all the faces, edges, and vertices, we end up with a space with 3 3-cells, 10 faces, 8 edges, and 3 vertices. We will name this space P.

Face	Equivalences
C1	F5
C2	F1
C3	$F9, F_1 3, F_1 4, F_1 5$
C4	$F_2 2, F_2 3, F_2 3, F_2 4, F_2 5, F_1 6$
C5	$F_2 1, F_2 6, F_1 7, F_1 8$
C6	$I1, I5$
C7	$I3, I7$
C8	$I9, I4$
C9	$I2, I6$
C10	$I8$

Edges	Equivalences
B1	E1-E8,E11-E14,E23-E30, $E_1 3, E_1 4, E_1 7, E_1 8, E_1 9, E_1 10$
B2	E9,E10,$E_1 5, E_1 6, E_2 5, E_2 6, E_2 7, E_2 8$
B3	E15-E18, $E_1 2, E_1 11, E_1 14, E_1 15, E_1 16, E_2 1, E_2 2, E_2 12, E_2 10$
B4	E19-E22,$E_1 1, E_1 12, E_1 13, E_1 18, E_1 17, E_2 3, E_2 4, E_2 9, E_2 11$
B5	E31,E35
B6	E32,E36
B7	E33,E37
B8	E34,E38,E39

Vertices	Equivalences
A1	V1-V20, $V_1 1 - V_1 12, V_2 1 - V_2 8$
A2	V22
A3	V21



This picture is the 1-skeleton of the moduli space.

3.1 Euler Characteristic

We will now compute the euler characteristic for our cell complex.

Definition 3.1 [Euler Characteristic] For a finite CW complex X, the **Euler characteristic** $\chi(X)$ is defined to be the alternating sum $\sum_n (-1)^n c_n$ where c_n is the number of n-cells of X. [1]

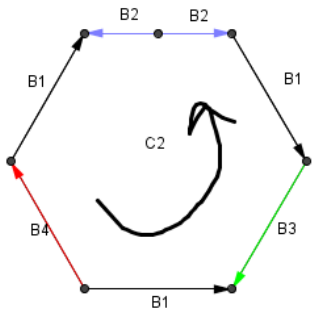
$$\chi(P) = \sum_{n=0}^{3} (-1)^n c_n = 3 - 8 + 10 - 3 = 2.$$

3.2 Homology Groups

We would now like to compute the dimensions of the homology groups $\dim H_k(P) = \dim \ker(\partial_{k+1}) - \dim Im(\partial_k)$ where ∂_k is a homomorphism from the group of k-chains (finite linear combinations of oriented k-cells) to the group of (k-1)-chains which represents a boundary operation.

Let $C_n(P)$=\{group of n-chains over $\mathbb{R}$\}. Then $\partial_n : C_n \to C_{n-1}$.

For example, lets look at the 3-cell C2. C2 is just a new name for F1 which was bounded by edges E10,E14,E15,E19,E23,E27. These edges however correspond to the 2-cells B2,B1,B3,B4,B1,B1. They are arranged and oriented as shown in the picture.



Therefore $\partial_2 C2 = B1 - B3 - B1 - B2 + B2 - B1 - B4 = -B3 - B1 - B4$

The choice of orientation is arbitrary but once a choice is made it must be held consistently throughout the process. Therefore, once I have decided the orientation of B1, it is fixed for all other B1s.

Once we have painstakingly oriented all of our n-cells correctly we arrive at the following matrices representations for ∂_1, ∂_2, and ∂_3.

$$\partial_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{pmatrix}$$

$$\partial_2 = \begin{pmatrix} 5 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{pmatrix}$$

$$\partial_3 = \begin{pmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \\ -2 & 0 & 0 \\ 2 & 0 & 0 \\ 2 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

Then we have the dimensions of the homology groups over $\mathbb{R}$ $H_k(P)$ are:

$$\dim H_0(P) = 1$$

$$\dim H_1(P) = 0$$

$$\dim H_2(P) = 1$$

We can use this to check our Euler Characteristic calculation because for a finite CW-complex, $\chi(X) = \dim H_0(X) - \dim H_1(X) + \dim H_2(x) - ...$ Therefore we have

$$\chi(P) = 1 - 0 + 1 = 2$$

which reaffirms our previous calculation.

Aknowledgements

I'd like to thank my fellow researcher Elizabeth Winkelman and Professor Chris Judge for their assistance in this project. I'd also like to thank Indiana University's Math Department and the NSF for providing this REU opportunity.

References

[1] Allen Hatcher: Algebraic Topology.Cambridge University Press, Cambridge, United Kingdom, 2002.

Constructing a Dynamical Boundary at Infinity for Hyperbolic Toral Automorphisms

Aleksandra Niepla[*]
Advisor: Bryce Weaver[†]

Abstract

We construct a dynamical boundary at infinity for hyperbolic toral automorphisms. Our construction consists of a backward boundary, (∂^-), and a forward boundary, (∂^+). With some limitations, we develop a correspondence between $(\partial^- \times \partial^+) \times \mathbb{Z}^2$ and the behavior of orbits in the system. This is a generalization of the boundary at infinity naturally obtained for geodesic flows defined on the unit tangent bundle of hyperbolic surfaces.

1 Introduction

We are interested in classifying the orbits of a dynamical system based on their asymptotic behavior. The general question inspiring this projects is: for which dynamical systems can we develop a notion of a "dynamical boundary at infinity"? This boundary is given by a set whose elements distinguish the asymptotic behaviors of the orbits up to a time shift. In particular, it is desirable to create a boundary with certain topological or measure theoretical properties. Before discussing the problem in question, we recall the basic components of a dynamical system. Below, is a short summary of the more detailed introductions to dynamical systems found in the first chapter of [BS] and in p.$1-12$ of [HK].

In the general setting, a dynamical system is composed of the following three components:

(i) A non-empty set X called the *phase space*. The elements of X correspond to the allowed states of the system. In most cases, X has a specific structure. For example, X may be a measure space, a topological space, or a smooth manifold.

(ii) A notion of "time": $\mathbf{T} = \mathbb{Z}, \mathbb{Z}_0^+, \mathbb{R}, \mathbb{R}_0^+$. It can be *discrete* or *continuous* and *reversible* or *irreversible*. In the case of a discrete-reversible dynamical system,

[*]Rutgers University, aleksandra.niepla@rutgers.edu. Research conducted during the math REU program at Indiana University Bloomington.
[†]Indiana University Bloomington, weavbryc@indiana.edu.

time can be represented by the integers. For a discrete-irreversible dynamical system time is not extended into the past and can be represented by the non-negative integers. Similarly, for continuous-reversible dynamical systems and continuous-irreversible dynamical systems, time is represented by the real numbers and the non-negative real numbers, respectively.

(iii) Lastly, a law governing the evolution of states as time progresses is required. Generally, this law is given in the form of a function $F : X \times \mathbf{T} \to X$

$$F(x, s + t) = F(F(x, t), s)$$

Given an initial state x, the future of the system after a time t can determine by $F(x, t)$. In the discrete-time case, this requirement can also be satisfied by defining a map $f : X \to X$ and considering its iterations (i.e. $F(x, n) = f^n(x)$ for $n \in \mathbb{Z}$). The *orbit* of a point $x \in X$ by $\mathcal{O}(x) = \bigcup_t F(x, t)$.

In most of the interesting examples of dynamical systems, this law preserves the structure of X. For instance, it may be measure preserving if X is a measure space, a continuous map or a homeomorphism if X is a topological space, or a differential map if X is a smooth manifold. To gain more insight into these various cases see p. $2 - 7$ of [HK] where a brief discussion and history of ergodic theory, topological dynamics, differentiable dynamics, and Hamiltonian dynamics can be found.

1.1 The dynamical boundary at infinity

The asymptotic behaviors of a dynamical system are described in terms of stable and unstable manifolds. Let (X, d) be a metric space and $F : X \times \mathbf{T} \to X$ be reversible. Given an $x \in X$, define the *stable manifold* of x as the set

$$W^s(x) = \left\{ y \in X : \lim_{t \to +\infty} d(F(x, t), F(y, t)) = 0 \right\} \tag{1}$$

and define the *unstable manifold* of x as the set

$$W^u(x) = \left\{ y \in X : \lim_{t \to -\infty} d(F(x, t), F(y, t)) = 0 \right\}. \tag{2}$$

For $x \in X$, $W^s(x)$ can be understood as the set of elements in X that have the same future as x and $W^u(x)$ can be understood as the set of elements in X with the same past as x.

We refer to the set distinguishing the past behavior of orbits as the backward boundary and denote it (∂^-). Similarly, we call the set distinguishing future behaviors of orbits the forward boundary and denote it (∂^+). We would like to find a backward boundary, a forward boundary, and an addition set Y such that elements of $(\partial^- \times \partial^+) \times Y$ distinguish a unique orbit in X and location on that orbit.

Note that a naive approach to constructing such a boundary is to let $(\partial^+) = X/\sim$ where: for $x, y \in X$ $x \sim y$ if there exists a $t_{x,y} \in \mathbf{T}$ such that

$$\lim_{t\to\infty} d(F(x,t), F(y, t + t_{x,y})) = 0. \tag{3}$$

Unfortunately, there are often technical problems with this approach. This is related to properties of systems such as ergodicity and mixing that rely on Hopf-type arguments (see for example p.17 of [HK] and Appendices A,C of [CM]).

2 Motivating example

As an example for the desired construction, we study the geodesic flow on hyperbolic surfaces. Sections $2.1 - 2.2$ introduce the hyperbolic plane and discuss its connection with hyperbolic surfaces. Sections 2.3 - 2.5 discuss the geodesic flow of hyperbolic surfaces and its relation to the orientation-preserving isometries of the hyperbolic plane. Unless otherwise specified, all constructions, lemmas, and theorems are taken directly from [Ka, Li] where a more detailed development of these ideas can be found. Lastly, the boundary at infinity is explained in Section 2.6.

2.1 The hyperbolic plane

Denote the upper-half plane by $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$. For $z \in \mathbb{H}$, the *tangent space* of $\mathbb{H}$ at z is the set of all vectors tangent to $\mathbb{H}$ at the point z and it is denoted $T_z\mathbb{H}$. The identity map $z \in \mathbb{R} \times \mathbb{R}_{>0} \to z \in \mathbb{H}$ is a parametrization covering $\mathbb{H}$. It follows that $T_z\mathbb{H} \cong \mathbb{R}^2 \cong \mathbb{C}$. For $v = v_r + iv_i$, $w = w_r + iw_i \in T_z\mathbb{H}$ define an inner product $< \cdot, \cdot >_z$ by

$$< v, w >_z = \frac{v_r w_r + v_i w_i}{\mathrm{Im(z)}^2}. \tag{4}$$

The inner product $< \cdot, \cdot >_z$ induces a Riemannian metric on $\mathbb{H}$ called the *hyperbolic metric*. The Riemannian manifold arising from $\mathbb{H}$ and $< \cdot, \cdot >_z$ is called the *hyperbolic plane.*

The angle between $v, w \in T_z\mathbb{H}$ is given by:

$$\cos\phi = \frac{< v, w >_z}{||v||_z ||w||_z}. \tag{5}$$

By direct computation, it can be seem that this definition of angle agrees with the Euclidean angle measure. Hence, the hyperbolic metric is conformal to the Euclidean metric. For the remainder of this paper, we denote the standard Euclidean inner product by $(\cdot, \cdot)$.

2.2 Hyperbolic surfaces

Definition 2.2.1 A hyperbolic surface is a complete, orientable, smooth surface with constant curvature equal to -1 (see [Li]).

The focus of this section is to present two main theorems, which demonstrate the connection between the hyperbolic plane and all other hyperbolic surfaces. The next two definitions are required for the understanding of Theorem 2.2.4.

Let X be a locally compact metric space, and let G be a group of isometries of X.

Definition 2.2.2 A family $\{M_\alpha | \alpha \in A\}$ of subsets of $\mathbb{H}$ is called locally finite if, for any compact subset $K \subset \mathbb{H}$, $M_\alpha \bigcap K \neq \emptyset$ for only finitely many $\alpha \in A$ (see [Ka]).

Definition 2.2.3 We say a group G acts properly discontinuously on $\mathbb{H}$ if the G-orbit of any point $x \in \mathbb{H}$ is locally finite (see [Ka]).

Denote the group of isometries of $\mathbb{H}$ by Isom($\mathbb{H}$).

Theorem 2.2.4 *Let M be a complete Riemannian surface with constant curvature $K = -1$. Then, M is isometric to $\Gamma\backslash\mathbb{H}$, where Γ is a subgroup of Isom($\mathbb{H}$) acting in a properly discontinuous manner on $\mathbb{H}$. The metric on $\Gamma\backslash\mathbb{H}$ is induced from the covering $\pi : \mathbb{H} \to \Gamma\backslash\mathbb{H}$.*

See p. $165 - 166$ of [HK] for an extended version of this theorem along with a proof. As a direct application of Theorem 2.2.4, we see that any complete hyperbolic surface, X, can be written $X = \Gamma\backslash\mathbb{H}$. Therefore, it is helpful to classify the possible subgroups $\Gamma \subset$ Isom($\mathbb{H}$) satisfying the conditions of the theorem.

Definition 2.2.5 A subgroup Γ of Isom($\mathbb{H}$) is called discrete if the induced topology on Γ is a discrete topology, i.e. if Γ is a discrete set in the topological space Isom($\mathbb{H}$) (see [Ka]).

Theorem 2.2.6 *Let Γ be a subgroup of Isom($\mathbb{H}$). Then, Γ acts properly discontinuously on $\mathbb{H}$ if and only if Γ is a discrete subgroup of orientation preserving isometries of $\mathbb{H}$.*

This results is obtained by combining Definition 8.1 and Theorem 8.6 of [Ka].

2.3 Geodesic flow

The geodesics of $\mathbb{H}$ are smooth, constant speed, and length minimizing curves between any two points in $\mathbb{H}$. For any $z, w \in \mathbb{H}$, there is a unique geodesic joining them. By the standard theory for Riemannian manifolds, for $(z, v) \in T\mathbb{H}$, there exists a unique geodesic γ such that $\gamma(0) = z$ and $\gamma'(0) = v$. For an in depth discussion on geodesics and the geodesic flow see Chapter 3 of [DC].

Theorem 2.3.1 *The geodesics in* $\mathbb{H}$ *are the semicircles and the rays orthogonal to the real axis* $\mathbb{R}$ *(see Figure 2.3.1).*

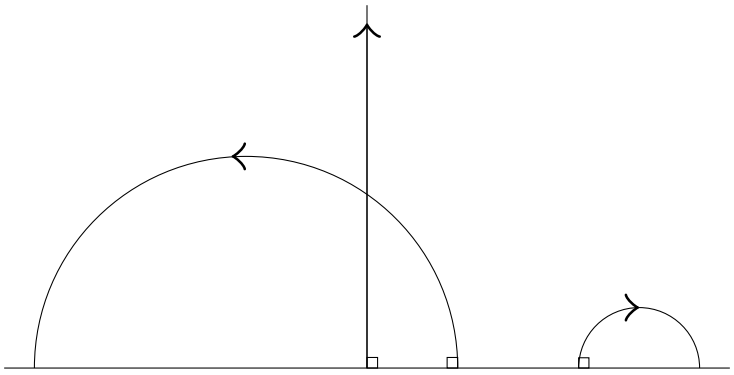


Figure 2.3.1: Geodesics in the hyperbolic plane

Now, we look at $\mathbb{H}$ from a dynamical perspective by studying its geodesic flow. Since geodesics are constant speed, it makes sense to re-parameterize them by arc-length. Let $T^1\mathbb{H} = \{(p, v) : p \in \mathbb{H}, v \in T_p\mathbb{H}, ||v|| = 1\}$ denote the unit tangent bundle of $\mathbb{H}$. The geodesic flow is an $\mathbb{R}$-action on $T^1\mathbb{H}$ defined as follows: for $v \in T^1_z\mathbb{H}$ let γ be the unique geodesic such that $(\gamma(0), \gamma'(0)) = (z, v)$. The geodesic flow $g = \{g_t\}_{t\in\mathbb{R}}$ of (z, v) is given by $g_t(z, v) = \gamma'(t)$ (see p. 10 of [Li]).

Note that $g_t(z, v) = (\gamma(t), \gamma'(t)) \in T\mathbb{H}$. Since γ is parametrized by arc-length, $||\gamma'(0)|| = ||v|| = 1$. Therefore, $g_t(z, v) \in T^1\mathbb{H}$. So, g_t maps $T^1\mathbb{H}$ into itself. It can be shown that g_t is invertible and hence a homeomorphism. The geodesic flow defined on $T^1\mathbb{H}$ gives a dynamical system $(T^1\mathbb{H}, g_t)$, where the unit tangent bundle, $T^1\mathbb{H}$, is the phase space.

Let $(z_o = x_o + iy_o, v_o = i/y_o) \in T^1\mathbb{H}$. Then, v_o is a vertical unit vector pointing upward and based at z_o. In this case, $W^s(z_o, v_o) = \{(z, v) \in T^1\mathbb{H} : y = y_o \text{ and } v = v_o\}$. The stable manifold of (z_o, v_o) consists of the elements in $T^1\mathbb{H}$ with base point on the horizontal line passing through y_o and vertical vectors pointing upward. Figures 2.3.2 shows the stable and unstable manifolds for (z_o, v_o).

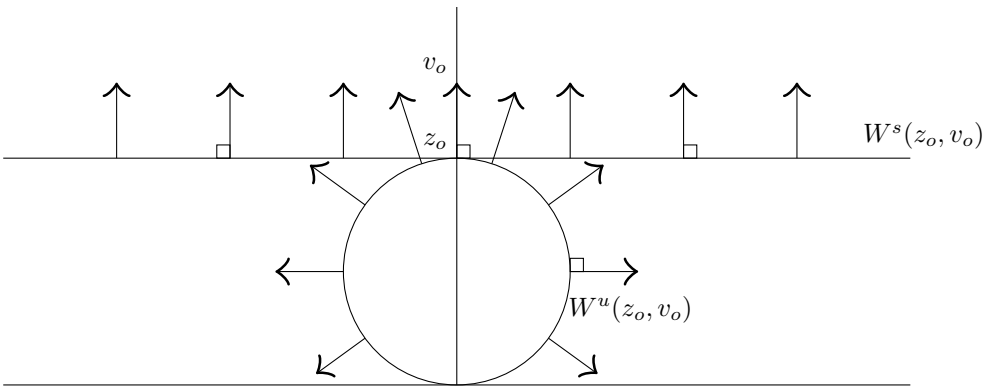


Figure 2.3.2: The stable and unstable manifolds of (z_o, v_o).

In the case when (z, v) is not a vertical unit vector pointing upward, $W^s(z, v)$ and $W^u(z, v)$ can be visualized as shown in Figure 2.3.3.

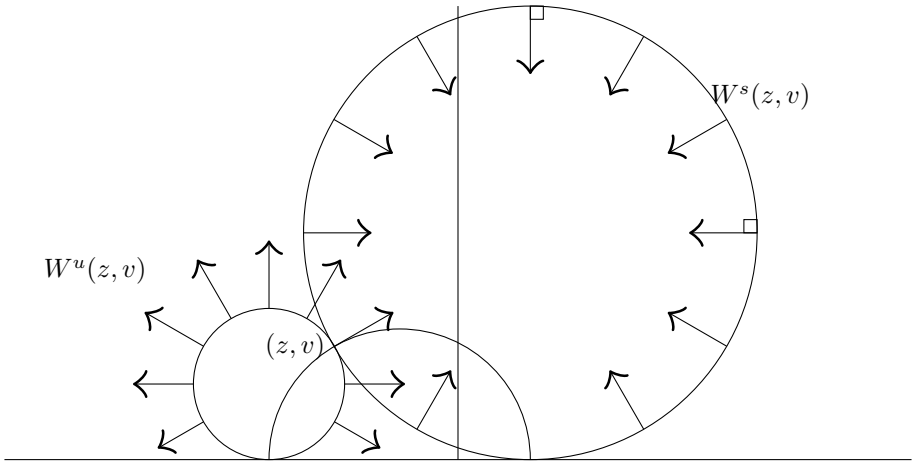


Figure 2.3.3: Stable and unstable manifolds for a generic (z, v) (see [Li]).

In the context of hyperbolic surfaces, the stable manifolds and unstable manifolds are often referred to as stable and unstable horospheres, respectively.

2.4 Möbius transformations

Let Isom($\mathbb{H}$) denote the set of isometries of the hyperbolic plane $\mathbb{H}$ and consider the group $SL(2,\mathbb{R})$ of real 2×2 matrices with determinant one. Define an action of $SL(2,\mathbb{R})$ on $\mathbb{H}$ as follows: for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{R})$, a Möbius transformation is given by

$$T_g(z) = \frac{az+b}{cz+d}. \tag{6}$$

Theorem 2.4.1 *Let $g \in SL(2,\mathbb{R})$. Then, the Möbius transformaion T_g is an orientation-preserving isometry of $\mathbb{H}$. Therefore, the set of Möbius transformaion is a subgroup of Isom($\mathbb{H}$) (see [Li]).*

Proof . Fix $z = x+iy \in \mathbb{H}$. Then,

$$T_g(z) = \frac{az+b}{cz+d} = \frac{(az+b)(c\bar{z}+d)}{|cz+d|^2} = \frac{ac(x^2+y^2)+(ad+bc)x+bd}{|cz+d|^2} + i\frac{y}{|cz+d|^2}.$$

As $\frac{y}{|cz+d|^2} > 0$, $T_g(z) \in \mathbb{H}$. Also,

$$dT_g(z) = \frac{a(cz+d)-(az+b)c}{(cz+d)^2} = \frac{\det(g)}{(cz+d)^2} = \frac{1}{(cz+d)^2}. \tag{7}$$

Therefore, given $v, w \in T_z\mathbb{H}$,

$$\langle dT_g v, dT_g w\rangle_{T(z)} = \frac{(dT_g v, dT_g w)}{\mathrm{Im}(T(z))^2} = \frac{|cz+d|^4(v,w)}{y^2|cz+d|^4} = \frac{(v,w)}{y^2} =< v, w >_z .$$

In the above calculation, we use Equation 7 together with the property that $(vz, wz) = |z|^2(v,w)$. Hence, T_g is an isometry of $\mathbb{H}$. □

In addition, note that the two matrices $g, -g \in SL(2,\mathbb{R})$ give the same Möbius transformations; so it is enough to consider the action of the quotient group $SL(2,\mathbb{R})/\pm I_2 = PSL(2,\mathbb{R})$ on $\mathbb{H}$. In fact, $PSL(2,\mathbb{R})$ gives all orientation preserving isometries of $\mathbb{H}$ (see [Li]).

The group $PSL(2,\mathbb{R})$ is also a topological space with a norm induced from $\mathbb{R}^4$ (see p. 27 of [Ka]). Therefore, by Theorem 2.2.4, we have that $X = \Gamma\backslash\mathbb{H}$, where Γ is a discrete subgroup of $PSL(2,\mathbb{R})$.

2.5 Geodesic flow in terms of PSL(2, $\mathbb{R}$)

The geodesic flow of $\mathbb{H}$ can be understood in terms of $PSL(2,\mathbb{R})$. The group $PSL(2,\mathbb{R})$ acts on $T^1\mathbb{H}$ in the following way: For $g \in PSL(2,\mathbb{R})$ and $(z,w) \in T^1\mathbb{H}$,

$$T_g(z,w) = (T_g(z), DT_g(w)),$$

where T_g is the Möbius transformation corresponding to g.

Proposition 2.5.1 *The action of $PSL(2,\mathbb{R})$ on $T^1\mathbb{H}$ is transitive. Therefore, given $(z_1,w_1),(z_2,w_2) \in T^1\mathbb{H}$, there is a $g \in PSL(2,\mathbb{R})$ such that $T_g(z_1) = z_2$ and $Dg(w_1) = w_2$. In addition, such g is unique. (see Section 1.2.4 of [Li] for proof).*

Theorem 2.5.2 *There is a homeomorphism between $PSL(2,\mathbb{R})$ and $T^1\mathbb{H}$ such that the action of $PSL(2,\mathbb{R})$ on itself by left multiplication corresponds to the action of $PSL(2,\mathbb{R})$ on $T^1\mathbb{H}$ described above (see [Ka]).*

As done in [Ka, Li], the homeomorphism can be constructed directly by fixing the element $(i,i) \in T^1\mathbb{H}$. By Proposition 2.4.1, for any $(z,w) \in T^1\mathbb{H}$, there is a unique T_g such that $T_g(i,i) = (z,w)$. Remarkably, the map $(z,w) \to T_g$ is a homeomorphism. See p. 26 of [Ka] for more details.

Next, we express the geodesic flow in terms of $PSL(2,\mathbb{R})$. To start, the geodesic passing through (i,i) is the positive imaginary axis. After parameterizing by arc length it can be calculated that $g_t(i,i) = (ie^t, ie^t), t \in \mathbb{R}$. The Möbius transformation taking (i,i) to $g_t(i,i)$ is $\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}$.

For a given $(z,w) \in T^1\mathbb{H}$, take the $g \in PSL(2,\mathbb{R})$ such that $(T_g(i), DT_g(i)) = (z,w)$. Since T_g is an isometry, we have that T_g sends $g_t(i,i)$ to $g_t(z,w)$ (see Figure 2.5.1). By composition, the element of $PSL(2,\mathbb{R})$ that sends (i,i) to $g_t(z,w)$ is the product $g\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}$. Therefore, in terms of $PSL(2,\mathbb{R})$, $g_t : PSL(2,\mathbb{R}) \to PSL(2,\mathbb{R})$ is right multiplication:

$$g \to g\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}.$$

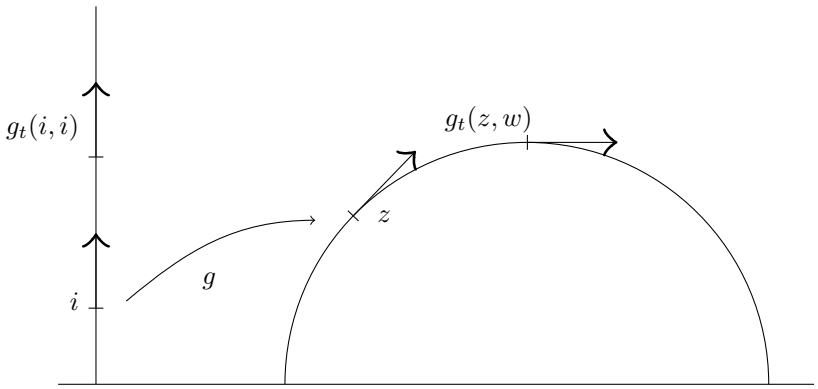


Figure 2.5.1: The geodesic flow of (i,i) can be mapped to the geodesic flow of any $(z,w) \in T^1\mathbb{H}$ via right multiplication in $PSL(2,\mathbb{R})$.

Similarly, since for a general hyperbolic surface X, we have $TX = \Gamma\backslash T\mathbb{H}$, the geodesic flow of X is given by the projection $T^1\mathbb{H} \to \Gamma\backslash T^1\mathbb{H}$ (see [Li]). The geodesic flow of X can be represented by right multiplication, $g_t : \Gamma\backslash PSL(2,\mathbb{R}) \to \Gamma\backslash PSL(2,\mathbb{R})$, as follows:

$$\Gamma g \to \Gamma g \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix} \text{(see [Li])}.$$

2.6 Constructing the boundary at infinity

From Section 2.3, we see that the future and past behaviors of the orbits in the system can be identified with $\mathbb{R} \cup \{\infty\}$ (see Figure 2.3.2 and 2.3.3). In this section, we show that the forward boundary, (∂^+), and backward boundary, (∂^-), can be given the topology of S^1. First, we start by considering the unit disk centered at i. For any non-vertical unit vector $(z,v) \in T^1\mathbb{H}$ the flow moves in a fixed direction along a semicircle orthogonal to the real axis. The asymptotic behavior is identified with the point on the real line intersecting the semicircle in the direction of the flow. In what follows, we consider $\mathbb{R}$ as the subset of $\mathbb{C}$ corresponding to the real axis.

Lemma 2.6.1 *There is a unique semicircle orthogonal to the real axis passing through i and $p \in \mathbb{R}\backslash\{0\}$. It has center $h = \frac{p^2-1}{2p}$ and radius $r = |p-h|$.*

Proof. Let h denote the center of the semicircle. Since we require the semicircle to be orthogonal to the real axis, $h \in \mathbb{R}$. Therefore, any such semicircle satisfies the equation:

$$(x-h)^2 + y^2 = r^2. \tag{8}$$

Also, i and p are points on the circle. Plugging into Equation 8, we have:

$$h^2 + 1 = r^2 \tag{9}$$

and

$$(p - h)^2 = r^2. \tag{10}$$

Setting Equations 9 and 10 equal to one another, gives

$$h^2 + 1 = (p - h)^2 = p^2 - 2ph + h^2.$$

Solving for h yields:

$$h = \frac{p^2 - 1}{2p}.$$

Therefore, the semicircle orthogonal to the real axis, and passing through i and p, is unique and has h, r as given. □

Proposition 2.6.2 *The forward boundary, (∂^+), and by symmetry the backward boundary, (∂^-), can be identified with S^1.*

Proof . Let $x \in \mathbb{R}\backslash\{0\}$ correspond to a future behavior of an orbit. By Lemma 2.6.1., there is a unique semicircle passing through i and x. This semicircle will intersect the unit disk at two points, one point having positive real coordinate and the other having negative real coordinate. Identify x with the point of intersection, z, having the same sign as x.

For a point z in the unit disk, consider the semicircle passing through i and z orthogonal to the real axis. It must hit a unique point, $x \in \mathbb{R}\backslash\{0\}$. So, this identification is one-to-one and onto. Finally, we identify $2i$ with ∞ and $0 \in \mathbb{R}$ with 0. By symmetry, for a given geodesic , $g_t(v, w)$, in $T^1\mathbb{H}$ the backward boundary can also be identified with the unit disk by considering $g_t(v, -w)$ and repeating the identification process above. Therefore, both the forward and backward boundaries, corresponding to the possible pasts and futures of the orbits in the system, can be given the topology of S^1.

□

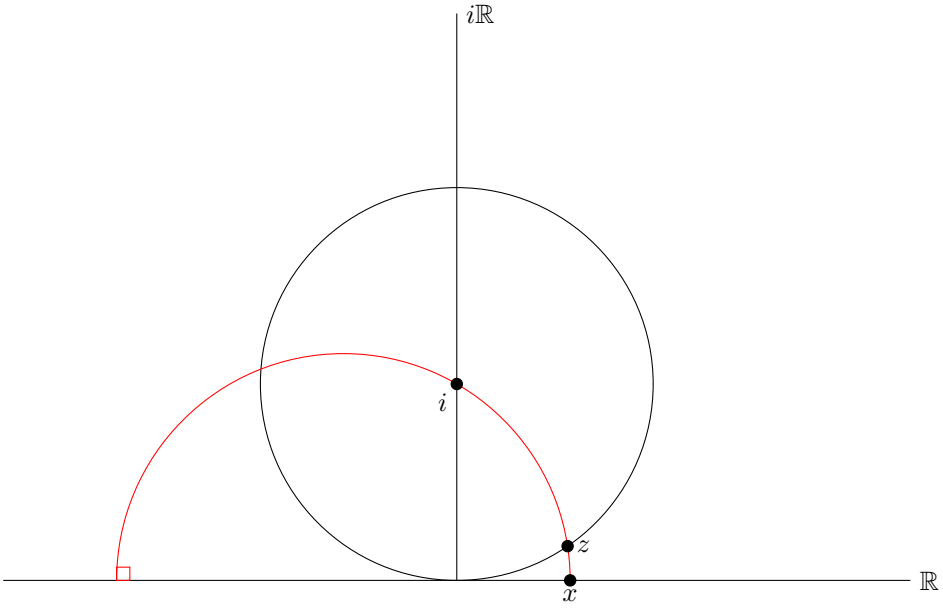


Figure 2.6.1: The correspondence between the geodesic flow of $\mathbb{H}$ and the unit disk is $z \longleftrightarrow x$.

Note that these geodesics intersect the unit disk twice. This is expected since one geodesic gives rise to two distinct asymptotic behaviors depending on the direction of the flow. This corresponds to past and future behaviors of orbits.

2.7 Properties of the boundary

Let $\Delta = \{(z, z) : z \in S^1\}$. Note that given $(z_1, z_2) \in (S^1 \times S^1)\backslash\Delta$, a unique orbit is identified with past corresponding the z_1 and future corresponding to z_2. This is because a past and a distinct future corresponds to a unique geodesic. This can be seen by noting the correspondence of S^1 and $\mathbb{R} \cup \{\infty\}$ (see Figure 2.6.1). There are no semicircles, hence no geodesics, connecting the same point in $\mathbb{R}$. This implies that no geodesic can have he same past and future behavior. This is why we eliminate points in Δ. For the remainder of the section, we fix the notation $(\partial^-) = S^1$ and $(\partial^+) = S^1$.

Next, we see that the set $(\partial^- \times \partial^+)\backslash\Delta \times \mathbb{R}$ gives a bijective correspondence between the orbits of the system and the possible locations on the orbits. Given $(x, y, t) \in (\partial^- \times \partial^+)\backslash\Delta \times \mathbb{R}$, the geodesic flow is first uniquely identified by its past, x, and its future, y. Next, fix the geodesic, γ, passing through the element $(i, v) \in T^1\mathbb{H}$ with the same future y. Traverse a length of time t from i obtaining $(\gamma(t), \gamma'(t))$. Take the stable horosphere containing $(\gamma(t), \gamma'(t))$ and corresponding to the future y. This horosphere has a unique intersection, (z, v),

with the geodesic flow identified with x and y. This intersection point gives the location on the orbit corresponding to (x, y, t).

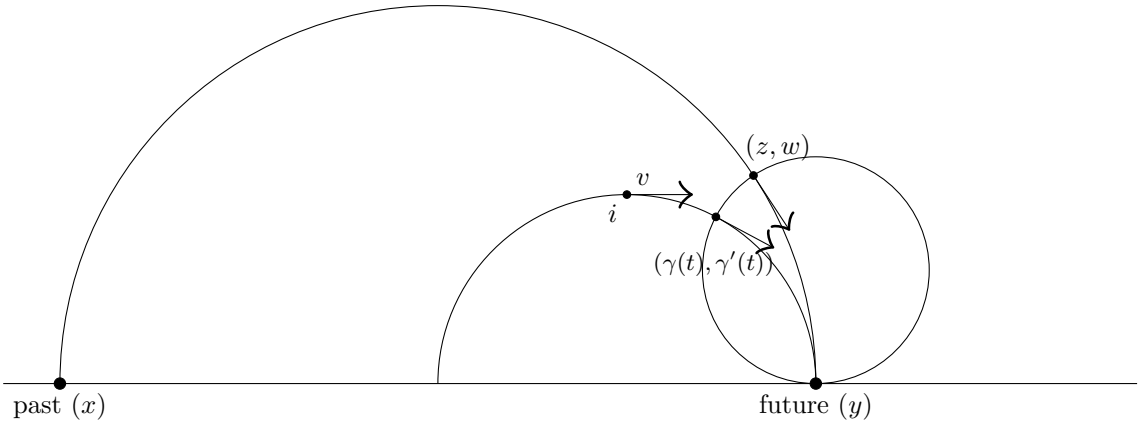


Figure 2.7.1: In this correspondence $(z, w) \longleftrightarrow (x, y, t) \in (\partial^+ \times \partial^-)\backslash\Delta \times \mathbb{R}$.

2.8 Summarizing the dynamical boundary of ($T^1\mathbb{H}$, g_t)

- For a given hyperbolic surface X, the geodesic flow gives a dynamical system with phase space T^1X.
- We studied the flow defined on $T^1\mathbb{H}$, since $\mathbb{H}$ is the universal cover of any hyperbolic surface.
- A forward boundary (∂^+) and backward boundary (∂^-) were naturally constructed and both are topologically equivalent to S^1.
- The boundary ($\partial^- \times \partial^+$) distinguishes orbits of the system.
- The set $(\partial^+ \times \partial^-)\backslash\Delta \times \mathbb{R}$ distinguishes orbits and locations on orbits allowing the full $T^1\mathbb{H}$ to be recovered.

3 The torus

We give a brief introduction to hyperbolic automorphisms of the two-dimensional torus, $\mathbb{T}^2$. In particular, their link with linear maps on $\mathbb{R}^2$. Lastly, we construct a “dynamical boundary at infinity" for this system.

3.1 Hyperbolic toral automorphisms

A n-dimensional torus is defined to be $\mathbb{T}^n = S^1 \times ... \times S^1 = \mathbb{R}^n/\mathbb{Z}^n$, where S^1 is the unit circle ([HK]). For the purposes of our discussion, we consider the

two-dimensional torus: $\mathbb{T}^2 = S^1 \times S^1 = \mathbb{R}^2/\sim$. Here we define the equivalence relation $\sim$ as follows: for $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$, $(x_1, y_1) \sim (x_2, y_2)$ if and only if $(x_1 - x_2, y_1 - y_2) \in \mathbb{Z}^2$. This is induced from the natural action of $\mathbb{Z}^2$ on $\mathbb{R}^2$, making $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. The relation is illustrated in Figure 3.1.1.

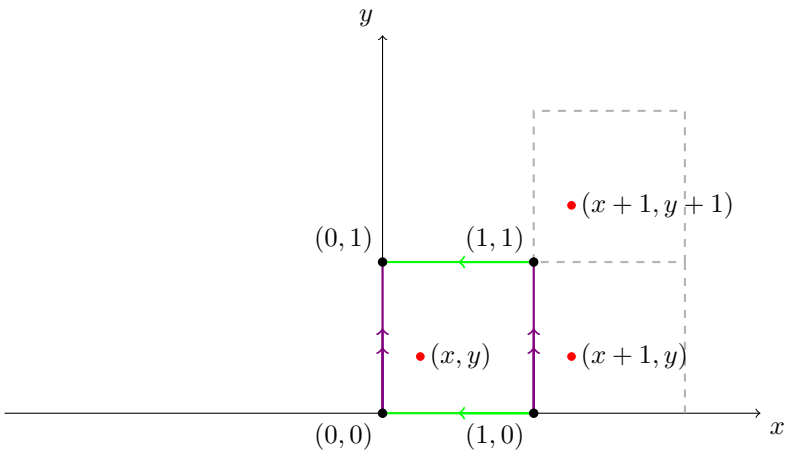


Figure 3.1.1: The green edges are identified and so are the purple edges. The red points represent the same point in $\mathbb{T}^2$.

Consider a map $f : \mathbb{R}^2 \to \mathbb{R}^2$. If f respects the equivalence relation given above, i.e., if $(x_1, y_1) \sim (x_2, y_2)$ implies $f(x_1, y_1) \sim f(x_2, y_2)$ then f induces a map on $\mathbb{T}^2$. This map is denoted $\tilde{f} : \mathbb{T}^2 \to \mathbb{T}^2$ and defined by $\tilde{f}(x, y) = [f(x, y)]$.

Proposition 3.1.1 *A linear map $A : \mathbb{R}^2 \to \mathbb{R}^2$, given by $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $a, b, c, d \in \mathbb{Z}$, induces a map on $\mathbb{T}^2$.*

Proof . Let $(x_1, y_1) \sim (x_2, y_2)$. Since $(x_1 - x_2, y_1 - y_2) \in \mathbb{Z}^2$,

$A(x_1, y_1) - A(x_2, y_2) = (a(x_1 - x_2) + b(y_1 - y_2), c(x_1 - x_2) + d(y_1 - y_2)) \in \mathbb{Z}^2$.

Therefore, $(x_1, y_1) \sim (x_2, y_2)$ implies $A(x_1, y_1) \sim A(x_2, y_2)$. So, A respects

the equivalence relation $\sim$ and hence induces a map on $\mathbb{T}^2$.

□

Definition 3.1.2 We say that a map $T : \mathbb{T}^2 \to \mathbb{T}^2$ is a *linear toral automorphism* if it is induced by a linear map A with integer entries and determinant equal to ± 1. In addition, T is *hyperbolic* if the eigenvalues of A are not of norm 1 (see [BS]).

It can be checked directly that linear toral automorphisms are invertible and continuous. Let A be a matrix representation of a linear hyperbolic toral automorphism. Note that $\prod_i \lambda_i = \text{Det}(A)$, where λ_i are the eigenvalues of A, $\text{Det}(A)= 1$, and $|\lambda_i| \neq 1$ imply that A has two distinct eigenvalues with $\lambda_2 = \frac{1}{\lambda_1}$. Also, note that $\text{Det}(A^n) = (\text{Det}(A))^n = \pm 1$ and the eigenvalues of A^n are $|\lambda_i^n| \neq 1$. So, the n^{th} iteration of a linear hyperbolic toral automorphism is itself a linear hyperbolic toral automorphism.

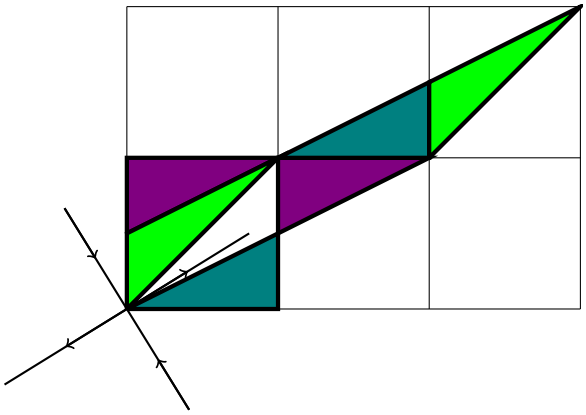


Figure 3.1.2: The image of $\mathbb{T}^2$ under a hyperbolic toral automorphism with the same identification as shown in Figure 3.1.1.

We abuse notation and refer to a linear automorphism, T, by the matrix that induces it. By fixing a linear hyperbolic toral automorphism, A, we can consider the discrete reversible dynamical system ($\mathbb{T}^2$, A). The $\mathbb{T}^2$ is the phase space and the law specifying the evolution of states is given by the iterations of A and A^{-1}. We continue by studying the dynamics in $\mathbb{R}^2$, the universal cover of $\mathbb{T}^2$.

Let E_{λ_i} denote the eigenspace of λ_i (i.e. the set of all eigenvectors corresponding to λ_i). It is clear that applying A^n to a vector in E_{λ_i} returns another vector in E_{λ_i} Therefore, E_{λ_i} is *A-invariant* for a fixed λ_i.

In this case, the two eigenspaces of A are E_λ and $E_{1/\lambda}$, where λ is the eigenvalue of A that is greater than one. The orbit of points contained in E_λ or $E_{1/\lambda}$ is easily understood. A expands by a factor of λ when applied to vectors in E_λ, and contracts by a factor of $1/\lambda$ when applied to vectors in $E_{1/\lambda}$ (see [BS]).

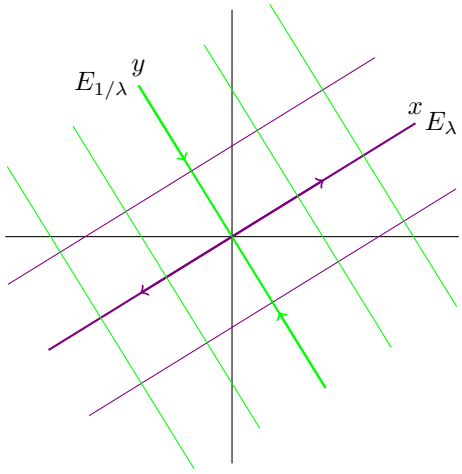


Figure 3.1.3: Eigendirections giving stable and unstable manifolds in $(\mathbb{T}^2, A)$.

The eigenspaces of A can be used to understand the behavior of the system. As explained in [BS], for $(x, y) \in \mathbb{R}^2$, the line passing through (x, y) and parallel to vectors in E_λ is $W^u(x, y)$. Similarly, the line passing through (x, y) and parallel to vectors in $E_{1/\lambda}$ is $W^s(x, y)$.

3.2 Dynamical boundary at infinity on the torus

We constructed the dynamical boundary of $\mathbb{T}^2$ in its universal cover, $\mathbb{R}^2$. The hyperbolic toral automorphism, A, can be expressed in terms of the basis corresponding to its eigenvectors by $A_d = \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$. With respect to these coordinates, the stable and unstable manifolds are given by:

$$W^s(x_o, y_o) = \{(x_o, y) : y \in \mathbb{R}\}$$

and

$$W^u(x_o, y_o) = \{(x, y_o) : x \in \mathbb{R}\}.$$

We begin by constructing a forward boundary. First, consider the following intervals partitioning $\mathbb{R}_{>0} \times \{0\}$:

$$\vdots$$
$$(1/\lambda^3, 1/\lambda^2] \times \{0\}$$
$$(1/\lambda^2, 1/\lambda] \times \{0\}$$
$$(1/\lambda, 1] \times \{0\}$$
$$(1, \lambda] \times \{0\}$$
$$(\lambda, \lambda^2] \times \{0\}$$
$$\vdots$$

Lemma 3.2.1 *There is a bijective correspondence between points in an interval of type* $(\lambda^{m-1}, \lambda^m] \times \{0\}$*, for* $m \in \mathbb{Z}$*, and the orbits of points in* $\mathbb{R}_{>0} \times \{0\}$*.*

Proof . Fix an $m \in \mathbb{Z}$. For any $(x, 0) \in (\lambda^{m-1}, \lambda^m]$, identify it with its own orbit, $\mathcal{O}(x, 0)$. Since $A_d^l(x, 0) = (\lambda^l x, 0) \in (\lambda^{l+m-1}, \lambda^{l+m}]$ for all $l \in \mathbb{Z}$, $(x, 0)$ is the only point in $(\lambda^{m-1}, \lambda^m] \times \{0\} \cap \mathcal{O}(x, 0)$. So, this identification is one-to-one.

Next, for an arbitrary $(x, 0) \in \mathbb{R}_{>0} \times \{0\}$, we have $(x, 0) \in (\lambda^{k-1}, \lambda^k]$ for some $k \in \mathbb{Z}$. Therefore, $A_d^{m-k}(x, 0) \in (\lambda^{m-1}, \lambda^m]$ and $\mathcal{O}(x, 0)$ is identified with $A_d^{m-k}(x, 0) \in (\lambda^{m-1}, \lambda^m]$. Hence, this identification is onto. □

Fix the interval $(1, \lambda] \times \{0\}$. In fact, Lemma 3.2.1 can be used to identify the future behaviors of all points in $\mathbb{R}_{>0} \times \mathbb{R}$ with $(1, \lambda] \times \{0\}$. For $(x, y) \in (\lambda^{k-1}, \lambda^k) \times \mathbb{R}$, make the identification: $(x, y) \sim_f A^{1-k}(x, 0) \in (1, \lambda] \times \{0\}$. This can also be understood as identifying (x, y) with $(x, 0)$ and then applying the identification in the proof of 3.2.1. This can be extended to $\mathbb{R}_{<0} \times \mathbb{R}$ giving the next Corollary.

Corollary 3.2.2 *There is a bijective correspondence between points in an interval of type* $((\lambda^{m-1}, \lambda^m] \cup [-\lambda^m, -\lambda^{m-1})) \times \{0\}$ *and the orbits of points in* $\mathbb{R}_{\neq 0} \times \{0\}$*.*

The interval $((1, \lambda] \cup [-\lambda, -1)) \times \{0\}$ is the forward dynamical boundary, (∂^+). The future behaviors of all points in $\mathbb{R}_{\neq 0} \times \mathbb{R}$ can be identified with (∂^+) by naturally extending the identification $\sim_f$ to include points in $\mathbb{R}_{<0} \times \{0\}$.

Similarly, the interval $\{0\} \times ((1, \lambda] \cup \times [-\lambda, -1))$ is the backward dynamical boundary, (∂^-). The past behaviors of any points in $\mathbb{R} \times \mathbb{R}_{\neq 0}$ can be identified with points in (∂^-) as follows: for $(x, y) \in \mathbb{R} \times ((\lambda^{k-1}, \lambda^k] \cup [-\lambda^k, -\lambda^{k-1}))$, $(x, y) \sim_b A^{1-k}(0, y) \in \{0\} \times ((1, \lambda] \cup [-\lambda, -1))$.

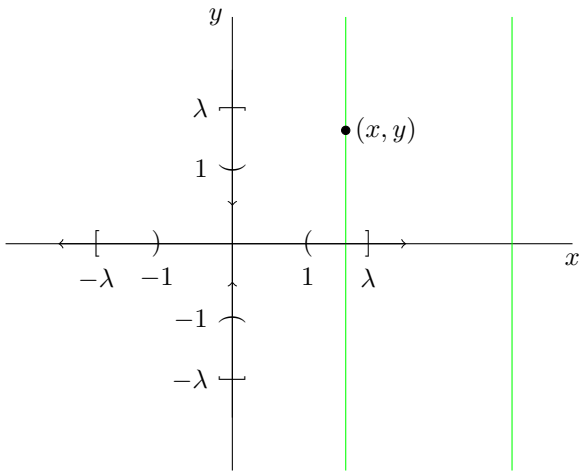


Figure 3.2.1: Forward and backward dynamical boundaries. The two green lines are stable manifolds that are identified by the forward boundary.

We define the dynamical boundary at infinity of this system by $(\partial^- \times \partial^+)$. This boundary identifies points in $\mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$ based on past and future behaviors as follows: for $(x_1, y_1), (x_2, y_2) \in \mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$, $(x_1, y_1) \sim (x_2, y_2)$ if and only if $(x_1, y_1) \sim_f (x_2, y_2)$ and $(x_1, y_1) \sim_b (x_2, y_2)$. It can be checked directly from the definition that $(x, y) \sim (\lambda^m x, \lambda^{-n} y)$ for any $m, n \in \mathbb{Z}$. All points in the same orbit are identified.

The dynamical boundary, $(\partial^- \times \partial^+)$ is a structure on points in $\mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$, but not on $\mathbb{R} \times \mathbb{R}$. The origin is not included in $(\partial^- \times \partial^+)$. Hence, all points with the same past as $(0, 0)$ (the $x-$axis) and the same future as $(0, 0)$ (the $y-$axis) are excluded from the identification. Therefore, $\{0\} \times \mathbb{R}$ and $\mathbb{R} \times \{0\}$ are not included in this dynamical boundary construction.

Note that in what follows, we abuse notation by also expressing the coordinates of the boundary as (x, y). The dynamical boundary $(\partial^- \times \partial^+)$ does not uniquely distinguish orbits (see Figure 3.2.2.).

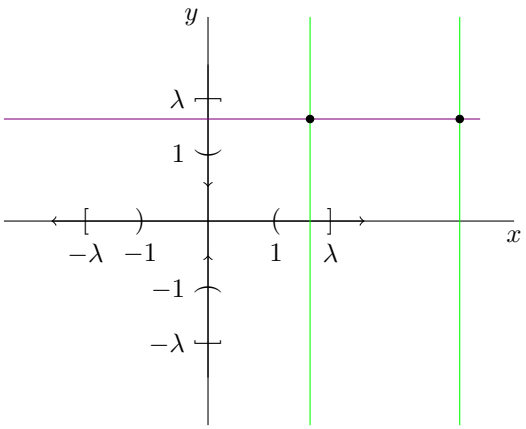


Figure 3.2.2: The two black points are identified with each other, but are not in the same orbit.

Lemma 3.2.3 *Any* $(x, y) \in (\partial^- \times \partial^+)$ *corresponds to a countable number of orbits characterized by* $(x, \lambda^{-m} y)$*, for* $m \in \mathbb{Z}$*.*

Proof . Take $(x, y) \in (\partial^- \times \partial^+)$,

$$(x, y) \sim (\lambda^n x, \lambda^{-l} y) \sim (x, \lambda^{-(l+n)} y),$$

since $A^{-n}(\lambda^n x, \lambda^{-l} y) = (x, \lambda^{-(l+n)} y)$. So, (x, y) corresponds to at most $\mathbb{Z}$ distinct orbits. Also,

$$(x, \lambda^{-m'} y) \in \mathcal{O}(x, \lambda^{-m''} y) \text{ if and only if}$$

$A_d^k(x, \lambda^{-m'} y) = (\lambda^k x, \lambda^{-(m'+k)} y) = (x, \lambda^{-m''} y)$ for some $k \in \mathbb{Z}$ if and only if

$k = 0$ and $m' = m''$. So, there are exactly $\mathbb{Z}$ distinct orbits identified by $m = l + n \in \mathbb{Z}$. □

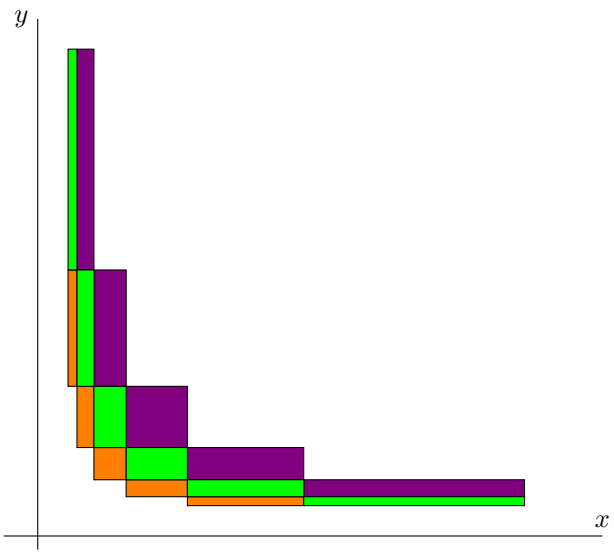


Figure 3.2.3: Each color distinguishes a family of orbits corresponding to all possible past and futures.

Therefore, we see that $(\partial^- \times \partial^+) \times \mathbb{Z}$ gives us a bijective correspondence between the distinct orbits of points in $\mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$ based and their pasts and futures.

Proposition 3.2.4 *There is a one-to-one correspondence with* $(\partial^- \times \partial^+) \times \mathbb{Z}^2$ *and the orbits and location on the orbits of points in* $\mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$ *given by* $(x, y, m, n) \leftrightarrow A_d^n(x, \lambda^{-m} y)$.

Proof . By Lemma 3.2.1, for any point in $\mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$, its orbit is uniquely identified with $(\partial^- \times \partial^+) \times \mathbb{Z}$ by $(x, \lambda^{-m} y)$. Any location on this orbit can be obtained by $A_d^n(x, \lambda^{-m})$, for some $n \in \mathbb{Z}$.

□

3.3 Summarizing the dynamical boundary of $(\mathbb{T}^2, A_d)$

- The dynamical system studied was iterations of a hyperbolic toral automorphism on $\mathbb{T}^2$.
- We studied the dynamics on $\mathbb{R}^2$, the universal cover of $\mathbb{T}^2$.
- A forward boundary (∂^+) and a backward boundary (∂^-) were constructed, distinguishing families of orbits of points in $\mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$.
- The set $(\partial^- \times \partial^+) \times \mathbb{Z}^2$ distinguishes orbits and locations on the orbits of points in $\mathbb{R}_{\neq 0} \times \mathbb{R}_{\neq 0}$.

4 Further questions

The boundary $(\partial^- \times \partial^+)$ is constructed in $\mathbb{R}^2$. A question for further studies is to understand this boundary under the identification of the torus: $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. In addition, it would be interesting to see whether this boundary can be used to retrieve information about the original dynamical system on $\mathbb{T}^2$. One might also ask whether the existence of such a boundary at infinity can be generalized to any "reasonable" dynamical system.

Acknowledgments

I would like to give a very special thanks to my advisor, Dr. Bryce Weaver, for all his time, dedication, and guidance throughout this entire project. I would also like to thank Professor Chris Connell for directing the math REU program at Indiana University, Bloomington and for helping make this research experience possible.

References

[BS] Brin, M. and Stuck, G., *Introduction to Dynamical Systems*, Cambridge University Press, New York, 2002.

[CM] Chernov, N.; Markarian, R. *Chaotic Billiards*, American Mathematical Society, Providence, R.I., 2006.

[DC] Do Carmo, M.P., *Riemannian Geometry*, Birkhäuser, Boston, 1992.

[HK] Katok, A. and Hasselblatt, B., *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, New York, 1995.

[Ka] Katok, S., *Fuchsian Groups, Geodesic Flows on Surfaces of Constant Negative Curvature, and Symbolic Coding of Geodesics*, Clay Mathematics Proceedings, 8 (2008)

[Li] Lima, Y., *Lectures on Ratner's Theory: Dynamics Beyond Uniform Hyberbolicity* lecture notes distributed at University of Maryland, College Park on August 9, 2013.

Ends of algebraic surfaces

Eamon Quinlan Gallego[*]

Abstract

The main goal of this REU project was to investigate the curvature of algebraic surfaces in $\mathbb{R}^3$. In particular we show that the curvature of an algebraic surface goes to zero as one goes to infinity. However, in the process of doing so interesting results regarding points at infinity were derived which are also included in this report.

1 Points at infinity of algebraic surfaces

Here we consider algebraic surfaces, which are defined as subsets Σ of $\mathbb{R}^3$ given as solutions to polynomials, i.e.

$$\Sigma = \{(x, y, z) \in \mathbb{R}^3 : f(x, y, z) = 0\}$$

for some $f \in \mathbb{R}[x, y, z]$. Moreover, some issues regarding the fact that $\mathbb{R}$ is not algebraically closed imposes that we must restrict to irreducible polynomials in this section.

Here we investigate points at infinity of algebraic surfaces. There are two different notions of points at infinity for an algebraic surface, which here we call *algebraic* and *topological*.

1.1 Algebraic points at infinity

We begin with a general discussion of the projective space $\mathbb{R}P^3$.

The space $\mathbb{R}P^3$ can be seen as the union of $\mathbb{R}^3$ and a $\mathbb{R}P^2$, the later being usually denoted as $\mathbb{R}P^2_\infty$ - $\mathbb{R}P^2$ at infinity. This notion can be made formal as follows:

By definition, $\mathbb{R}P^3 = (\mathbb{R}^4 \backslash 0)/\sim$, where $\sim$ represents the equivalence relation $\boldsymbol{x} \sim \boldsymbol{y} \iff \boldsymbol{x} = \lambda \boldsymbol{y}$ for some $\lambda \in \mathbb{R}$, $\lambda \neq 0$.

Then the elements of $\mathbb{R}P^3$ can be denoted by choosing a representative $(x, y, z, w) \in \mathbb{R}^4 \setminus 0$ and writing $[x : y : z : w]$ for its equivalence class.

Now, if $[x : y : z : w] \in \mathbb{R}P^3$ is such that $w \neq 0$ then

$$[x : y : z : w] = [\frac{x}{w} : \frac{y}{w} : \frac{z}{w} : 1]$$

This shows that any such equivalence class $[x : y : z : w]$ has a representative with $w = 1$, which is easily seen to be unique.

[*]University of Glasgow

We can therefore write $\mathbb{R}P^3 = \{[x : y : z : 1]\} \cup \{[x : y : z : 0]\}$. But $\{[x : y : z : 1]\}$ can be clearly identified with $\mathbb{R}^3$ by $[x : y : z : 1] \to (x, y, z)$ - the uniqueness of the representative with $w = 1$ is important here - and $\{[x : y : z : 0]\}$ can be identified with $\mathbb{R}P^2$ by $[x : y : z : 0] \to [x : y : z]$. In this context, the $\mathbb{R}P^2$ is denoted as $\mathbb{R}P^2_\infty$.

Although the space $\mathbb{R}P^2$ is can be defined in a completely algebraic way there are various ways of realizing this space topologically. As it will be seen shortly, in order to compare the algebraic and topological points at infinity we want to realize $\mathbb{R}P^2$ as a quotient space of the sphere. For that we will use the map

$$\begin{array}{rccc} \phi : & \mathbb{S}^2 & \longrightarrow & \mathbb{R}P^2 \\ & (x, y, z) & \mapsto & [x : y : z] \end{array}$$

of the unit sphere onto $\mathbb{R}P^2$, which is a double covering.

Now let $\Sigma = \{(x, y, z) \in \mathbb{R}^3 : f(x, y, z) = 0\}$ for some irreducible polynomial $f \in \mathbb{R}[x, y, z]$. We can then consider the homogenization of f. The process of homogenizing a polynomial consists in adding a new variable w to f so that the resultant polynomial $\tilde{f}$ is a homogeneous polynomial in four variables such that $\tilde{f}(x, y, z, 1) = f(x, y, z)$. This last condition implies that the two polynomials coincide in the $\mathbb{R}^3$ part of $\mathbb{R}P^3$.

The process of homogenization is best illustrated by example. Say:

$$f(x, y, z) = x^4 + 3y^2z^2 - 5z^2 + x - 4$$

then we add w^n to each monomial choosing n so that we get a homogeneous polynomial of degree 4, obtaining:

$$\tilde{f}(x, y, z, w) = x^4 + 3y^2z^2 - 5z^2w^2 + xw^3 - 4w^4$$

A formal definition is given by $\tilde{f}(x, y, z, w) = w^{\deg(f)} f(x/w, y/w, z/w)$, where cancellations in the numerator and denominator by w are allowed.

Because $\tilde{f}$ is homogeneous it satisfies $\tilde{f}(\boldsymbol{x}) = 0 \implies \tilde{f}(\lambda \boldsymbol{x}) = 0$ for all $\lambda \in \mathbb{R}$ and hence we can consider the set of solutions in $\mathbb{R}P^3$, given by

$$\tilde{\Sigma} = \{[x : y : z : w] \in \mathbb{R}P^3 : \tilde{f}(x, y, z, w) = 0\}$$

Definition 1.1 Let $\Sigma = \{(x, y, z) \in \mathbb{R}^3 : f(x, y, z) = 0\}$ for some irreducible polynomial $f \in \mathbb{R}[x, y, z]$. Let $\tilde{f}(x, y, z, w) = w^{\deg(f)} f(x/w, y/w, z/w)$ be the homogenization of f and let $\tilde{\Sigma} = \{[x : y : z : w] \in \mathbb{R}P^3 : \tilde{f}(x, y, z, w) = 0\}$. Then the *algebraic points at infinity* of Σ are defined as $G_f = \phi^{-1}\big(\tilde{\Sigma} \cap \mathbb{R}P^2_\infty\big) \subset S^2$.

Remark 1.2 It is not clear from the definition that we may freely speak of algebraic points at infinity of Σ, since the construction relies on the polynomial f defining Σ. In fact, if we take $f(x, y, z) = x^2 + y^2 - z^2 - 1$ and $g(x, y, z) = (x^2 + y^2 - z^2 - 1)(x^2 + 1)$ it can easily be checked that although f and g give the same surface they give different algebraic points at infinity. However, it is enough to restrict to irreducible polynomials to solve this issue.

1.2 Topological points at infinity

Now we go on to define the topological points at infinity. As opposed to the algebraic points at infinity, these can be defined for general surfaces - and in fact, for general subsets of $\mathbb{R}^3$ - and we will do so here, although we are ultimately interested in algebraic surfaces.

As it is well known, the map $r \mapsto r/(r+1)$ gives a homeomorphism of $[0,\infty)$ onto $[0,1)$. Using this, a homeomorphism of $\mathbb{R}^3$ onto B_1, where B_1 is the unit ball, can be given by

$$\begin{array}{rccc} \psi: & \mathbb{R}^3 & \longrightarrow & B_1 \\ & \boldsymbol{x} & \mapsto & \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|+1} \end{array}$$

Under this homeomorphism we see that the unit sphere $\mathbb{S}^2$ can be identified as a "sphere at infinity". We would like to consider the image of our surface Σ under this map, and the set of points of the closure of the image that lie on this "sphere at infinity".

Definition 1.1 Let $\Sigma \subset \mathbb{R}^3$. The set of *topological points at infinity* of Σ is given by $L_\Sigma = \overline{\psi(\Sigma)} \cap \mathbb{S}^2 \subset \mathbb{S}^2$.

Remark 1.2 The closure in the above definition is taken by considering $\psi(\Sigma)$ as a subset of $\mathbb{R}^3$. Also, when Σ refers to an algebraic surface $\{(x,y,z) \in \mathbb{R}^3 : f(x,y,z) = 0\}$ we will write $_f$ for L_Σ.

The following lemma gives an equivalent definition.

Lemma 1.3 *Let $\Sigma \subset \mathbb{R}^3$. Then $p \in L_\Sigma$ if and only if there exists a sequence $(x_n) \subset \Sigma$ such that $\|x_n\| \to \infty$ and $\hat{x}_n = \frac{x_n}{\|x_n\|} \to p$*

Proof If $p \in L_\Sigma = \overline{\psi(\Sigma)} \cap S^2$ then there exists a sequence $(y_n) \subset \psi(\Sigma)$ such that $y_n \to p$. Then the sequence $x_n = \psi^{-1}(y_n) = \frac{y_n}{1-\|y_n\|}$ clearly satisfies the required conditions.
Conversely, if there exists such a sequence (x_n) then by considering $y_n = \psi(x_n)$ we immediately see that $p \in \overline{\psi(\Sigma)} \cap S^2 = L_\Sigma$. □

1.3 Results

Now that we have defined the two kinds of points at infinity we are ready to give a few results.

Lemma 1.1 *Let $H \in \mathbb{R}[x,y,z]$ be a homogeneous polynomial in three variables. Then $L_H = G_H$.*

Proof Let $\Sigma = \{(x,y,z) \in \mathbb{R}^3 : H(x,y,z) = 0\}$. First, notice that because H is homogeneous $H(x,y,z,0) = H(x,y,z)$ and hence:

$$G_H = \phi^{-1}(\Sigma) = \Sigma \cap \mathbb{S}^2$$

The fact that H is homogeneous also implies that

$$H(\boldsymbol{x}) = 0 \implies H(\lambda \boldsymbol{x}) = 0$$

for all $\lambda \in \mathbb{R}$, and hence $\psi(\Sigma) = \Sigma \cap B_1$.

Next, we claim that $\overline{\Sigma \cap B_1} \cap \mathbb{S}^2 = \Sigma \cap \mathbb{S}^2$. To see this first let $(x, y, z) \in \overline{\Sigma \cap B_1} \cap S^2$. Then there exists a sequence $(x_n, y_n, z_n) \in \Sigma \cap B_1$ such that $(x_n, y_n, z_n) \to (x, y, z)$. Since, for all n we get $H(x_n, y_n, z_n) = 0$, and H is a polynomial - and therefore continuous - we obtain $H(x, y, z) = 0$. This shows $\overline{\Sigma \cap B_1} \cap S^2 \subseteq \Sigma \cap S^2$.

For the other inclusion, suppose $(x, y, z) \in \Sigma \cap S^2$. Let

$$(x_n, y_n, z_n) = \frac{n}{n+1}(x, y, z)$$

Then $(x_n, y_n, z_n) \in B_1$ and because H is homogeneous

$$H(x_n, y_n, z_n) = (\frac{n}{n+1})^{\deg H} H(x, y, z) = 0$$

for all n. Hence $(x_n, y_n, z_n) \in \Sigma \cap B_1$ and $(x_n, y_n, z_n) \to (x, y, z)$. This shows $\overline{\Sigma \cap B_1} \cap \mathbb{S}^2 \supseteq \Sigma \cap \mathbb{S}^2$. This proves the claim.

Now we get

$$L_H = \overline{\psi(\Sigma)} = \overline{\Sigma \cap B_1} \cap \mathbb{S}^2 = \Sigma \cap \mathbb{S}^2$$

and hence $G_H = \Sigma \cap \mathbb{S}^2 = L_H$ as required. □

Because of this last result, when we speak about a homogeneous polynomial H we may say points at infinity without a risk of confusion. From now on, we denote by I_H the set of points at infinity of H when H is homogeneous.

Lemma 1.2 *Let $f \in \mathbb{R}[x, y, z]$ be a polynomial. Then f can be uniquely decomposed as $f = H + l$ where H is homogeneous,* $\deg(f) = \deg(H)$ *and* $\deg(f) > \deg(l)$. *Then $G_f = I_H$.*

Proof With regards to the decomposition, just take H to be made up of the terms of highest order of f, and then choose $l = f - H$. This decomposition is immediately seen to be unique.

It's now easy to see that $\tilde{f}(x, y, z, 0) = H(x, y, z)$. It then follows that $G_f = \phi^{-1}(\{\boldsymbol{x} \in \mathbb{R}P^2 : H(\boldsymbol{x}) = 0\}) = G_H = I_H$. □

Lemma 1.3 *Let $f \in \mathbb{R}[x, y, z]$ be a polynomial and $f = H + l$ be the decomposition described in the above lemma. Then $L_f \subseteq I_H$.*

Proof Let $\Sigma = \{(x, y, z) \in \mathbb{R}^3 : f(x, y, z) = 0\}$ and $(x, y, z) \in \mathbb{S}^2$ be an topological point at infinity of Σ. Then there exists a sequence $(x_n, y_n, z_n) \subset \psi(\Sigma) \cap \mathbb{S}^2$ such that $(x_n, y_n, z_n) \to (x, y, z)$.

We obtain that for all n $\frac{1}{1-r_n}(x_n, y_n, z_n) \in \Sigma$ where $r_n = \sqrt{x^2+y^2+z^2}$ and hence

$$\begin{aligned}
&f\big(\frac{x_n}{1-r_n}, \frac{y_n}{1-r_n}, \frac{z_n}{1-r_n}\big) = 0 \\
\Rightarrow &\big(\frac{1}{1-r_n}\big)^{\deg(f)} H(x_n, y_n, z_n) + l\big(\frac{x_n}{1-r_n}, \frac{y_n}{1-r_n}, \frac{z_n}{1-r_n}\big) = 0 \\
\Rightarrow &H(x_n, y_n, z_n) + (1-r_n)^{\deg(f)} l\big(\frac{x_n}{1-r_n}, \frac{y_n}{1-r_n}, \frac{z_n}{1-r_n}\big) = 0
\end{aligned}$$

Since $r_n \to 1$ and $\deg(f) > \deg(l)$ we conclude $H(x_n, y_n, z_n) \to 0$. This shows, again by continuity, that $H(x, y, z) = 0$, i.e.
$(x, y, z) \in \{\boldsymbol{x} \in \mathbb{R}^3 : H(\boldsymbol{x}) = 0\} \cap \mathbb{S}^2 = L_H = I_H$. □

2 Ends of algebraic surfaces

Before we analyse the curvature at infinity of algebraic surfaces, which was the ultimate goal of this project, we first want to define the notion of an *end*.

Even though this concept is highly intuitive describing it in abstract terms takes some work.

In this section we rigorously define an end and then show that an algebraic surface has a finite number.

Let $\Sigma \subset \mathbb{R}^3$ be an unbounded path connected component of an algebraic surface. We denote:

$$B_i = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| < i\}$$
$$S_i = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| = i\}$$

Because Σ is unbounded and path connected we get that $\Sigma \cap S_i \neq \emptyset$ for all $i \in \mathbb{N}$. Moreover, for all $i \in \mathbb{N}$, $\Sigma \cap S_i$ is a curve on the sphere, possibly including isolated points - if Σ intersected the surface on a whole disk then $\Sigma = S_i$ since Σ is defined by a polynomial. We now use the following theorem from [5] (Theorem 4.1).

Theorem 2.1 *Let R be a real closed field. For any variety X over R the set $\pi_0(X(R))$ of path connected components is finite.*

From the proof given in [5] it is also seen that the number of path connected components is bounded in terms of the degree of X. Applied to our problem, this has the following corollary.

Corollary 2.2 *As a subset of S_i, the curve $\Sigma \cap S_i$ has a finite number of path connected components. Moreover, this number is bounded for a given Σ.*

Now, for every $i \in \mathbb{N}$, let the path connected components of $\Sigma \cap S_i$ be $\{\alpha_j^{(i)}\}_{j=1}^{k_i}$ and, as discussed above, the k_i are bounded above by some integer R.

Let A denote the set of all such intersections, i.e.

$$A = \Sigma \cap \big(\bigcup_{i \in \mathbb{N}} S_i\big) = \{\alpha_j^{(i)}\}_{i=1\ j=1}^{\infty\ k_i}$$

We now need to introduce some notation: for any surface Ω with a path connected subset S we denote by $p.c.c.(S,\Omega)$ the path connected component of Ω on which S lies.

We may now introduce a relation $\triangleleft$ on A given by

$$\alpha_{j_1}^{(i_1)} \triangleleft \alpha_{j_2}^{(i_2)} \iff i_2 > i_1 \text{ and } \alpha_{j_2}^{(i_2)} \in p.c.c.(\alpha_{j_1}^{(i_1)}, \Sigma \setminus B_{i_1})$$

i.e. $\alpha_{j_1}^{(i_1)} \triangleleft \alpha_{j_2}^{(i_2)}$ if and only if $\alpha_{j_2}^{(i_2)}$ belongs to a bigger sphere than $\alpha_{j_1}^{(i_1)}$ and there exists a path on Σ strictly outside B_{i_1} that joins $\alpha_{j_1}^{(i_1)}$ and $\alpha_{j_2}^{(i_2)}$.
Under this relation we now may consider the set of increasing chains:

$$C = \left\{ \alpha_{j_1}^{(1)} \triangleleft \alpha_{j_2}^{(2)} \triangleleft \alpha_{j_3}^{(3)} \triangleleft ... : j_i \in \{1, 2, \cdots, k_i\} \right\}$$

and the set of path connected components of Σ described by such chains:

$$E = \{ \{ p.c.c.(\alpha_{j_i}^{(i)}, \Sigma \setminus B_i) : i \in \mathbb{N} \} : (\alpha_{j_i}^{(i)})_{i \in \mathbb{N}} \in C \}$$

This set can be provided with the equivalence relation $\sim$ described as follows: let $e_1, e_2 \in E$ be given by

$$e_1 = \{ p.c.c.(\alpha_{j_i}^{(i)}, \Sigma \setminus B_i) : i \in \mathbb{N} \}$$

$$e_2 = \{ p.c.c.(\alpha_{l_i}^{(i)}, \Sigma \setminus B_i) : i \in \mathbb{N} \}$$

for two chains $(\alpha_{j_i}^{(i)})_{i \in \mathbb{N}}, (\alpha_{l_i}^{(i)})_{i \in \mathbb{N}} \in C$

Then:

$$e_1 \sim e_2 \iff \exists N \in \mathbb{N} \ \forall n \geq N \ p.c.c.(\alpha_{j_n}^{(n)}, \Sigma \setminus B_n) = p.c.c.(\alpha_{l_n}^{(n)}, \Sigma \setminus B_n)$$

Definition 2.3 With the notation described above, an equivalence class of E under $\sim$ is called an *end of* Σ.

Lemma 2.4 *An unbounded path connected component Σ of an algebraic surface has finitely many ends.*

Proof As stated in Corollary 2.2, $\Sigma \cap S_i$ has a number of path connected components $k_i \leq R$ for all $i \in \mathbb{N}$. Let K_i be the number of path connected components of $\Sigma \setminus B_i$. It follows from the above that $K_i \leq R$.

Note that K_i is a (weakly)-increasing bounded sequence of positive integers, and is therefore ultimately stationary, i.e. $\exists N \in \mathbb{N} \ \ \forall n \geq N \ \ K_n = K$ for some integer K.

Now suppose for a contradiction that there were $K+1$ or more ends. Let these ends be represented by the chains $(\alpha_{a_{l_i}}^{(i)})_{i \in \mathbb{N}} \in C$ with $l \in \{1, 2, \ldots, K+1\}$. Recall that, by the definition of the relation $\triangleleft$, for any chain $(\alpha_{j_i}^{(i)})_{i \in \mathbb{N}} \in C$ we get

$$\forall M \in \mathbb{N} \ \ \forall n \geq M \ \ \alpha_{j_n}^{(n)} \in p.c.c.(\alpha_{j_M}^{(M)}, \Sigma \setminus B_M)$$

In our case, $\Sigma \setminus B_N$ has only K path connected components and hence by the pigeon-hole principle we must have two sequences that eventually belong to the same path-connected component. Without loss of generality let these two sequences be $(\alpha^{(i)}_{a1_i})_{i\in\mathbb{N}}$ and $(\alpha^{(i)}_{a2_i})_{i\in\mathbb{N}}$, and let e_1 and e_2 be the representatives of the ends they induce.

From the above discussion we deduce that

$$\forall n \geq N \; p.c.c(\alpha^{(n)}_{a1_n}, \Sigma \setminus B_n) = p.c.c.(\alpha^{(n)}_{a2_n}, \Sigma \setminus B_n)$$

and hence $e_1 = e_2$, giving a contradiction. □

3 Curvature of algebraic surfaces of revolution

We now start investigating the problem of showing that the curvature of an unbounded algebraic surface goes to zero as one goes to infinity along any end. But before we tackle the general case we would first like to consider a simpler case: surfaces of revolution.

Lemma 3.1 *Let $F \in \mathbb{R}[x, y, z]$ be a polynomial in three variables that defines the algebraic surface*

$$\Sigma = \{(x, y, z) \in \mathbb{R}^3 : F(x, y, z) = 0\}$$

Then Σ has rotational symmetry about the X-axis if and only if

$$\Sigma = \{(x, y, z) \in \mathbb{R}^3 : f(x, \sqrt{y^2 + z^2}) = 0\}$$

for some polynomial $f : \mathbb{R}^2 \to \mathbb{R}$ in two variables.

Proof Let Σ be such that it has rotational symmetry about the X-axis. Then the curve of intersection C of Σ with the XY-plane is given by $F(x, y, 0) = 0$.

Let $\boldsymbol{x} = (x, r\cos\theta, r\sin\theta)$ be an arbitrary point in $\mathbb{R}^3$ expressed in cylindrical coordinates. Then the symmetry of Σ about the X-axis implies that

$$\boldsymbol{x} \in \Sigma \iff (x, r, 0) \in C$$

and by letting $f(x, y) = F(x, y, 0)$ we see that

$$\Sigma = \{(x, y, z) \in \mathbb{R}^3 : f(x, \sqrt{y^2 + z^2}) = 0\}$$

as required.

Conversely, if

$$\Sigma = \{(x, y, z) \in \mathbb{R}^3 : f(x, \sqrt{y^2 + z^2}) = 0\}$$

for some polynomial f then

$$(x, r\cos\theta, r\sin\theta) \in \Sigma \iff (x, r, 0) \in S$$

for all θ and hence Σ has rotational symmetry about the X-axis. $\square$

Let $\Sigma = \{\boldsymbol{x} \in \mathbb{R}^3 : f(\boldsymbol{x}) = 0\}$ be an unbounded algebraic surface of revolution which has the x-axis as its axis of symmetry and e be an end of Σ.

As seen above, the intersection of Σ with the x, y-plane is given by $f(x, y, 0) = 0$. This curve has finitely many connected components by Theorem 2.1.

Lemma 3.2 *Let $\boldsymbol{x}_n = (x_n, y_n) \subset \mathbb{R}^2$ be a sequence of points in the curve C given by $f(x, y, 0) = 0$ where $x_n \to \infty$. Assume further that $\boldsymbol{x}_n$ is in the same path connected component P of C. Then if $y_n \nrightarrow 0$ we get that Σ must have infinite area.*

Proof We show this by proving that revolving P gives a surface of infinite area.

Since $y_n \nrightarrow 0$ we get that for some $\epsilon_0 > 0$ there are infinitely points (x_n, y_n) in the sequence such that $|y_n| > \epsilon_0$.

We now have two cases:

(i) The lines $y = \pm\epsilon_0$ intersects P in an interval. Then, since both are defined by polynomials, we get that P must be one of the lines $y = \pm\epsilon_0$. By revolving this about the x-axis we clearly get infinite area.

(ii) The line $y = \epsilon_0$ intersects P in points, or not at all. Then by Bezout's Theorem the number of points of intersection is finite. Therefore, eventually P must be either inside the region $-\epsilon_0 < y < \epsilon_0$ or outside. But since there are infinitely many points outside we get that the former is impossible.

Now we parametrize P by arc-length as $(x(s), y(s))$. Then the area A_R obtained when revolving P about the x-axis between $s = 0$ and $s = R$ is given by

$$A_R = 2\pi \int_0^R y(s)ds \geq 2\pi\epsilon_0 R$$

which clearly diverges as $R \to \infty$. $\square$

Proposition 3.3 *Let $F \in \mathbb{R}[x, y, z]$ be a polynomial that defines an algebraic surface Σ. Suppose that Σ has rotational symmetry along some axis and that its generating curve is unbounded. Let $(s_n) \subset \Sigma$ be a sequence of points in Σ such that $\|s_n\|$ diverges. Then $K(s_n) \to 0$ as $n \to \infty$, where K is the Gaussian curvature.*

Remark 3.4 Note that in general algebraic surfaces are not regular surfaces and hence the curvature may not be well defined. However, this is fixed by considering $\Sigma \setminus \boldsymbol{Sing}(F)$, where $\boldsymbol{Sing}(F) = \{\boldsymbol{x} \in \mathbb{R}^3 : \nabla(F) = 0\}$. But because $\Sigma \cap \boldsymbol{Sing}(F)$ is at most of dimension 1 this is a rather technical detail. For now, we assume that $s_n \notin \boldsymbol{Sing}(F)$ for all n.

Proof Without loss of generality assume the axis of symmetry is the X-axis. By Proposition 1.1 we know that $S = \{(x, y, z) \in \mathbb{R}^3 : f(x, \sqrt{y^2 + z^2}) = 0\}$ where

$f(x,v) = F(x,v,0)$. Because of symmetry K only depends on x and $\sqrt{y^2+z^2}$. By letting $v = \sqrt{y^2+z^2}$ we obtain:

$$K(x,v) = \left.\frac{f_v(f_{vv}f_x^2 - 2f_x f_v f_{xv} + f_v^2 f_{xx})}{v(f_x^2+f_v^2)^2}\right|_{(x,v)}$$

It remains to show that $K(x,v) \to 0$ as $x \to \infty$.

Without loss of generality we may assume that $s_n = (x_n, v_n, 0)$ where $x_n \to \infty$ and that the whole of s_n is contained in the same path-connected component P of C, where C is, as before, the curve of intersection of Σ with the x,y-plane. We divide the proof into two cases:

(i) If $v_n \nrightarrow 0$ then S has infinite area by Lemma 3.2 and by [4] (Proposition 1) we get that $K \to 0$ as $x \to \infty$.

(ii)If $v_n \to 0$ then consider the map

$$\begin{array}{rccc} \Phi: & \mathbb{R}^2 \setminus Y & \longrightarrow & \mathbb{R}^2 \setminus Y \\ & (x,y) & \mapsto & (1/x, y) \end{array}$$

where Y denotes the y-axis. In particular, consider the image P' of $P \setminus \{(0,v) : v \in \mathbb{R}\}$. It's easy to see that the image C' of C is given by $f(\frac{1}{x}, v) = 0$.

C' can be made algebraic by multiplying the equation $f(\frac{1}{x}, v) = 0$ by $x^{\deg_x(f)}$ (in this process the origin is be added to the solution).

By looking at the image under ϕ of the sequence s_n we see that P' contains a branch through the origin. This branch may be locally expressed as a Puiseux series in x:

$$v(x) = b_1 x^{\frac{1}{n}} + b_2 x^{\frac{2}{n}} + \cdots$$

The inverse image under ϕ of the interval on which the above series converges is an interval of the form $(-\infty, M) \cup (R, \infty)$ for some M, R. We conclude that in this interval P may be expressed as a series of the form:

$$v(x) = b_1 (\frac{1}{x})^{\frac{1}{n}} + b_2 (\frac{1}{x})^{\frac{2}{n}} + \cdots$$

for some $n \in \mathbb{N}$.

For $f(x,v) = v - v(x)$ the curvature formula reduces to

$$K(x,v) = \left.\frac{-v_{xx}}{v(v_x^2+1)}\right|_{(x,v)}$$

To show that $K \to 0$ as $x \to \infty$ we track the smallest order terms in $1/x$.

Let $I = \min\{i \in \mathbb{N} : b_i \neq 0\}$. Then the smallest non-zero term in the numerator is $\frac{I}{n}(\frac{I}{n}+1) b_I (\frac{1}{x})^{(I/n+2)}$ where as in the denominator it is $b_I (\frac{1}{x})^{(I/n)}$. We conclude that $K \to 0$ as $x \to \infty$ as required. $\square$

4 Curvature at infinity

4.1 Curvature of analytic points at infinity

As it will be seen later, eventually a proof was found for the fact that algebraic surfaces have curvature going to zero as one goes off to infinity. The proof relies on showing that the surface may be expressed as union of graphs of a certain form and, potentially, this property may not be unique to algebraic surfaces.

Therefore we would like to introduce the kind of surfaces that this property holds - which we call analytic at infinity - in this section and show that indeed the curvature goes to zero at infinity. In the next section, we show that any algebraic surface is analytic at infinity.

We begin by recalling the equivalent definition of a topological point at infinity given in Lemma 1.5.

Definition 4.1 Let Σ be an unbounded surface. A point $p \in \mathbb{S}^2$ is called a *topological point* at infinity of Σ if there exists a sequence $(x_n) \subset \Sigma$ such that $\|x_n\| \to \infty$ and $\hat{x}_n = \frac{x_n}{\|x_n\|} \to p$. The sequence (x_n) is called a *characteristic sequence* of p.

In this section, we will refer to these points as points at infinity.

Suppose that Σ is an unbounded surface and $p \in \mathbb{S}^2$ is a point at infinity of Σ with characteristic sequence x_n. Let Π_p be the plane perpendicular to p and let u, v be an orthonormal basis for Π_p such that $\{p, u, v\}$ forms a right handed orthonormal basis for $\mathbb{R}^3$.

We may now consider the map

$$\begin{array}{rccc} \Phi_p : & \mathbb{R}^3 \setminus \Pi_p & \longrightarrow & \mathbb{R}^3 \setminus \Pi_p \\ & \lambda p + \mu u + \eta v & \mapsto & \frac{1}{\lambda} p + \frac{\mu}{\lambda} u + \frac{\eta}{\lambda} v \end{array}$$

Note that Φ_p is a homeomorphism and it is its own inverse, i.e. $\Phi_p \circ \Phi_p = Id$. Moreover, Φ_p is a PGL_4 transformation given by the matrix

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

in the $\{p, u, v, w\}$ basis.

Lemma 4.2 *With the notation described above,* $\Phi_p(x_n) \to (0, 0, 0)$

Remark 4.3 Potentially, some points of (x_n) could lie in Π_p, and then $\Phi(x_n)$ would not be defined. However, this number has to be finite and hence we can ignore this situation.

Proof First of all, we express (x_n) in the (u, v, p) coordinates as $x_n = \lambda_n p + \mu_n u + \eta_n v$. Now, from the definition, we have that $\|x_n\| \to \infty$ and $\frac{x_n}{\|x_n\|} \to p$. By denoting $r_n = \sqrt{\lambda_n^2 + \mu_n^2 + \eta_n^2}$ these imply that

$$r_n \to \infty$$

$$\frac{\lambda_n}{r_n} \to 1$$
$$\frac{\mu_n}{r_n} \to 0$$
$$\frac{\eta_n}{r_n} \to 0$$

and therefore

$$\begin{aligned}\Phi_p(x_n) &= \frac{1}{\lambda_n}(p + \mu_n u + \eta_n v) \\ &= \frac{r_n}{\lambda_n}\left(\frac{1}{r_n}p + \frac{\mu_n}{r_n}u + \frac{\eta_n}{r_n}v\right) \\ &\to 1(0p + 0u + 0v+) = (0,0,0)\end{aligned}$$

as required. □

Corollary 4.4 $(0,0,0) \in \overline{\Phi(\Sigma \setminus \Pi_p)}$

From now on, we assume that the vector $(\lambda, \mu, \eta) \in \mathbb{R}^3$ is given with respect to the basis $\{p, u, v\}$, i.e $(\lambda, \mu, \eta) = \lambda p + \mu u + \eta v$. Let's also denote

$$F_\epsilon = \{(\lambda, \mu, \eta) \in \mathbb{R}^3 : 0 < \lambda < \epsilon,\ -\epsilon < \mu < \epsilon,\ -\epsilon < \eta < \epsilon\}$$

Definition 4.5 Let Σ be an unbounded surface and let p be a point at infinity of Σ. Then p is called analytic if, using the notation described above, $\Phi_p(\Sigma \setminus \Pi_p)$ can be written, in a neighbourhood F_ϵ, as a union of finitely many graphs $\eta_k(\lambda, \mu)$ of the form

$$\eta^{(k)} = \sum_{i,j} a_{ij}^{(k)} \lambda^{i/n_k} \mu^{j/n_k} \tag{1}$$

where $a_{ij}^{(k)} \in \mathbb{R}$, $n_k \in \mathbb{N}$ and the exponents $(i/n_k, j/n_k)$ lie in a strictly convex cone of $\frac{1}{n_k}\mathbb{Z}^2_{>0}$, and the summation is taken in this cone.

For the definition of strictly convex cone, see [1].

Definition 4.6 Let Σ be an unbounded surface. If all points at infinity of Σ are analytic then Σ is said to be *analytic at infinity.*

Theorem 4.7 *Let Σ be an unbounded surface and p a point at infinity of Σ which is analytic. Then for every characteristic sequence (x_n) of p we get that $K(x_n) \to 0$ as $n \to \infty$, where K is the Gaussian curvature.*

Proof We may assume that $x_n \notin \Pi_p$ for all n. Then $(\Phi(x_n)) \subset \Phi(\Sigma \setminus \Pi_p)$. This implies that $(\Phi(x_n))$ lies in the finitely many graphs of the form 1. By applying the transformation Φ_p to the neighbourhood F_ϵ we conclude that Σ can be expressed as a union of finitely many graphs of the form

$$\eta^{(k)} = \sum_{i,j} a_{ij}^{(k)} \left(\frac{1}{\lambda}\right)^{(i/n_k)-1} \left(\frac{\mu}{\lambda}\right)^{j/n_k}$$

in a region $\{(\lambda, \mu, \eta) \in \mathbb{R}^3 : R < \lambda < \infty,\ -R < \mu < R,\ -R < \eta < R\}$. We now prove the claim by showing that the curvature goes to zero along each of these graphs as $\frac{1}{\lambda} \to 0$ and $\frac{\mu}{\lambda} \to 0$.

For a graph of this form, the curvature reduces to

$$K(\lambda, \mu) = \left. \frac{\eta_{\lambda\lambda}\eta_{\mu\mu} - \eta_{\lambda\mu}^2}{(1 + \eta_\lambda^2 + \eta_\mu^2)^2} \right|_{(\lambda,\mu)}$$

where we have dropped the index k for clarity. The exponents of $(1/\lambda)$ in the series expansion are all greater than $(2/n) - 1$ so after one differentiation by λ we get only positive exponents.

We therefore immediately recognize that the numerator has no constant term whereas the denominator has constant term 1. This implies that $K \to 0$ as required. □

Corollary 4.8 *Let Σ be analytic at infinity. Let $(x_n) \subset \Sigma$ be a sequence such that $\|x_n\| \to \infty$ and $\hat{x}_n = \frac{x_n}{\|x_n\|}$ converges. Then $K(x_n) \to 0$.*

4.2 Algebraic surfaces are analytic at infinity

As mentioned in the previous section, our main motivation when looking at surfaces analytic at infinity is studying algebraic surfaces. Here we show that they are analytic at infinity and hence have curvature going to zero at every point at infinity. We begin by using a well-known fact that allows us to speak of curvature for algebraic surfaces

Proposition 4.1 *Let $S = \{\boldsymbol{x} \in \mathbb{R}^3 : f(\boldsymbol{x}) = 0\}$ be an algebraic surface. Then $\Sigma = S \setminus \boldsymbol{Sing}(f)$ is a regular surface, where $\boldsymbol{Sing}(f) = \{\boldsymbol{x} \in \mathbb{R}^3 : \nabla f(\boldsymbol{x}) = 0\}$.*

For a proof, see [3] (Section 2-2 Proposition 2). Notice that the set $\boldsymbol{Sing}(f) \cap \Sigma$ is at most of dimension 1. From now on, we will call Σ an algebraic surface and assume that the singular set has been removed to make the surface regular.

We start with the following

Definition 4.2 Let $F \in \mathbb{C}[x, y, z]$. Then a formal power series $\phi \in \mathbb{C}[[x, y]]$ is called a solution of F if $F(x, y, \phi(x, y)) = 0$.

Lemma 4.3 *Let $F \in \mathbb{C}[x, y, z]$. Then there exists some N such that for every two power series solutions ϕ_1, ϕ_2 of F we get*

$$\phi_1 \equiv \phi_2 \mod \langle x, y \rangle^N \implies \phi_1 = \phi_2$$

Proof Express F as $F(x, y, z) = \sum_{i=0}^{N} z^i f_i(x, y)$ where $f_i \in \mathbb{C}[x, y]$. Then consider the map

$$\begin{array}{rccc} \Phi: & (\mathbb{C}[[x, y]])^{N+1} & \to & \mathbb{C}[[x, y]] \\ & (\phi_0, ..., \phi_N) & \mapsto & \sum_{i=0}^{N} \phi_i f_i(x, y) \end{array}$$

This is a $\mathbb{C}$-linear map and notice that $\phi \in \mathbb{C}[[x,y]]$ is a power series solution if and only if $(1, \phi, \phi^2, ..., \phi^N) \in \ker(\Phi)$.

Since $(\mathbb{C}[[x,y]])^{N+1}$ is a Noetherian $\mathbb{C}$-module, it's kernel is finitely generated over $\mathbb{C}$. Let $\big((\psi_0^{(0)}, ..., \psi_N^{(0)}), ..., (\psi_0^{(T)}, ..., \psi_N^{(T)})\big)$ be a basis for the kernel. We can choose this basis so that $(\psi_1^{(1)}, \psi_1^{(2)}, ..., \psi_1^{(t)})$ are linearly independent for some $t \leq T$ and $\psi_1^{(s)} = 0$ for all $s > t$.

From now on, denote $\psi_i^{(1)}$ as ψ_i. Now let $\phi_1 = \sum_{i=1}^t \alpha_i^{(1)} \psi_i$ and $\phi_2 = \sum_{i=1}^t \alpha_i^{(2)} \psi_i$. Then

$$\phi_1 \equiv \phi_2 \mod \langle x, y \rangle^n \iff \sum_{i=1}^t \big(\alpha_i^{(1)} - \alpha_i^{(2)}\big)(\psi_i)_n = 0$$

and hence it is enough to show that there exists some $N \in \mathbb{N}$ such that $\{(\psi_i)_n\}_i = 1^t$ are linearly independent.

To prove this claim let $b_n = \dim\langle (\psi_i)_n \rangle_i = 1^t$. Then b_n is a (weakly)-increasing sequence of positive integers which is bounded above by t. Hence it must converge to some value, say $s \leq t$. This would mean that $\exists N \in \mathbb{N}$ such that $\forall n \geq N$ we get $\dim\langle (\psi_i)_n \rangle_i = 1^t = s$. Now we want to show that $s = t$.

For a contradiction, suppose $s < t$. Then $\forall n \in \mathbb{N}$ $\exists \boldsymbol{\alpha}^{(n)} \in \mathbb{C}^t$, $\boldsymbol{\alpha}^{(n)} \neq 0$ such that $\sum_{i=1}^t \alpha_i^{(n)} (\psi_i)_n = 0$.

We now let $A_n = \{\alpha \in \mathbb{C}^t : \sum_{i=1}^t \alpha_i (\psi_i)_n = 0\}$. Then the A_n are non-trivial subspaces of $\mathbb{C}^t$ and they form a descending chain. We conclude the chain must terminate in a non-trivial subspace. This implies that for some non-zero $\boldsymbol{\alpha} \in \mathbb{C}^t$ we get $\forall n \in \mathbb{N}$ $\sum_{i=1}^t \alpha_i (\psi_i)_n = 0$, meaning $\sum_{i=1}^t \alpha_i \psi_i = 0$, giving the required contradiction. □

Corollary 4.4 *Let $F \in \mathbb{C}[x,y,z]$ and $\phi \in \mathbb{C}[[x,y]]$ be such that $F(x, y, \phi(x,y)) = 0$. Then ϕ is convergent in some neighbourhood of the origin.*

Proof By Artin's Approximation Theorem [2] there exist convergent power series solutions $\tilde{\phi_n}$ such that

$$\tilde{\phi_n} \equiv \phi \mod \langle x, y \rangle^n$$

However, because of the lemma we get that, for big enough n, $\tilde{\phi_n} = \phi$. □

We now take the following result from [1]

Theorem 4.5 *Let $F \in \mathbb{C}[x,y,z]$ be a polynomial such that $z \nmid F$ and $F(0,0,0) = 0$. Then there exist power series ϕ of the form $\phi = \sum_{i,j} a_{ij} x^{i/n} y^{j/n}$ where $a_{ij} \in \mathbb{C}$, $n \in \mathbb{N}$ and the exponents $(i/n, j/n)$ lie in a strictly convex cone of the lattice $\frac{1}{n}\mathbb{Z}^2$ such that $F(x, y, \phi(x,y)) = 0$.*

Lemma 4.6 *The series ϕ obtained above are convergent on a neighbourhood of the origin.*

Proof First we note that since $F(0,0,0) = 0$ there are no negative exponents. Let $\phi = \sum_{i,j} a_{ij} x^{i/n} y^{j/n}$. Then $\phi'(x,y) = \phi(x^n, y^n)$ is a formal power series that solves $F'(x,y,z) = F(x^n, y^n, z)$ and hence, because of Corollary 4.12, it converges on a neighbourhood of the origin. Let E be its domain of convergence and denote $D_\epsilon = \{z \in \mathbb{C} : |z| < \epsilon\}$.

Because the domain of convergence of a complex power series is complete circular [6] (Proposition 2.3.15) we can find $\epsilon_1, \epsilon_2 > 0$ such that $D_{\epsilon_1} \times D_{\epsilon_2} \subset E$.

Then, because for all $z \in \mathbb{C}$ and $k, m \in \mathbb{N}$, $|z|^{k/m} = |z^k|^{1/m}$ - we get that if we let $\tilde{\epsilon_i} = \epsilon_i^n$ for $i = 1, 2$ then ϕ converges on $D_{\tilde{\epsilon_1}} \times D_{\tilde{\epsilon_2}}$. □

Moreover, because of Remark 4.2 in [1] we obtain d series - counting multiplicity - where $d = \deg_z(F)$. These results also hold when restricting to real polynomials: in this case, the coefficients a_{ij} obtained must be real so that, when x and y are real, so is z. We are now ready to give a proof of the theorem.

Theorem 4.7 *Let Σ be an algebraic surface. Then Σ is analytic at infinity.*

Proof Let p be a point at infinity of Σ, let u, v be an orthonormal basis of Π_p - the plane perpendicular to p - so that $\{p, u, v\}$ forms a right-handed, orthonormal basis for $\mathbb{R}^3$.

We now use this basis to refer to vectors as $(\lambda, \mu, \eta) = \lambda p + \mu u + \eta v$. The surface Σ is then the zero-locus of some polynomial $f \in \mathbb{R}[\lambda, \mu, \eta]$, i.e. $\Sigma = \{(\lambda, \mu, \eta) : f(\lambda, \mu, \eta) = 0\} \setminus \boldsymbol{Sing}(f)$.

$\Phi_p(\Sigma)$ is then given by $f(\frac{1}{\lambda}, \frac{\mu}{\lambda}, \frac{\eta}{\lambda}) = 0$ which, for points $(\lambda, \mu, \eta) \notin \Pi_p$, is the same as $\lambda^{\deg(f)} f(\frac{1}{\lambda}, \frac{\mu}{\lambda}, \frac{\eta}{\lambda}) = 0$ which is now the zero locus of a real polynomial.

By [1] we can then find series of the form

$$\eta = \sum_{i,j} a_{ij} \lambda^{i/n} \mu^{j/n}$$

Since the polynomial is real, the a_{ij} are real and since the origin is included in the zero-locus we conclude that the exponents $(i/n, j/n)$ lie in the lattice $\frac{1}{n}\mathbb{Z}^2_{>0}$.

Finally, by Lemma 4.14 this series converges in a neighbourhood of the origin, and therefore in F_ϵ for some ϵ, as required. □

Corollary 4.8 *Let Σ be an algebraic surface, p a point at infinity of Σ and (x_n) its characteristic sequence. Then $K(x_n) \to 0$.*

5 Acknowledgements

I would like to thank the National Science Foundation for their founding, the University of Indiana at Bloomington for all their support throughout the whole of the project and especially to my advisor Professor Christopher Connell for his inestimable help and guidance. Working in Indiana University at Bloomington this summer has been a great experience.

References

[1] J. McDonald. Fiber polytopes and fractional power series. *Journal of Pure and Applied Algebra*, 104, 213-233 (1995).

[2] M. Artin. On the solutions of analytic equations. *Inventiones math.*, 5, 277-291 (1968).

[3] M. P. Do Carmo. Differential Geometry of Curves and Surfaces. *Prentice Hall* (1976).

[4] J.L.M. Barbosa and M. P. Do Carmo. On Regular Algebraic Surfaces with Constant Mean Curvature. arXiv:1403.7029.

[5] H. Delfs and M. Knebusch. Semialgebraic Topology Over a Real Closed Field I. *Math. Z.* 177, 107-129 (1981).

[6] S.G. Krantz. Function Theory of Several Complex Variables, Second Edition. *AMS Chelsea Publishing*, 1951.

Geometric Properties of Conformal Transformations on $\mathbb{R}^{p,q}$

Surya Raghavendran[*]

Abstract

In the present article, we show that conformal transformations on the generalized Minkowski space $\mathbb{R}^{p,q}$ map hyperboloids and hyperplanes into hyperboloids and hyperplanes. We show that this action is transitive when p or $q = 0$, and that this action has exactly 3 orbits otherwise. These properties are a generalization of well-known properties of Mobius (or fractional linear) transformations on $\mathbb{C}$ which map circles and lines to circles and lines.

1 Introduction

In the present article, we investigate geometric properties of conformal transformations on the generalized Minkowski space $\mathbb{R}^{p,q}$. Recall that $\mathbb{R}^{p,q}$ is the scalar product space consisting of $\mathbb{R}^{p+q}$ together with the indefinite quadratic form given by

$$Q(x) = (x^1)^2 + \cdots + (x^p)^2 - (x^{p+1})^2 - \cdots - (x^{p+q})^2.$$

Conformal transformations on $\mathbb{R}^{p,q}$ can be described by an action of the indefinite orthogonal group $O(p+1, q+1)$ on a suitable conformal compactification. In the case of $\mathbb{R}^{2,0} \cong \mathbb{C}$ and $\mathbb{R}^{4,0} \cong \mathbb{H}$, the action of $O(p+1, q+1)$ by conformal transformations is transitive on the moduli space of spheres and hyperplanes in the ambient space (c.f. [1, 5]). A similar result for the case of $\mathbb{R}^{1,1}$ is also known (c.f. [8]). In this paper, we investigate this action in higher dimensions.

References

[1] C.Bisi, G.Gentili, *Mobius transformations and the Poincare distance in the quaternionic setting.* Indiana Univ. Math. Journal, 58, 2009, also arXiv:0805.0357.

[2] S.Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces.* Academic Press, 1978

[*]University of Texas at Austin

[3] T.Y.Lam, *Introduction to Quadratic Forms over Fields*. American Mathematical Soc.

[4] M.Arayapoor, *The Penrose Transform in the Split Signature*. Differential Geometry and its Applications, 30, 2008, also arXiv:0812.3692

[5] E.Gwynne, M.Libine, *On a quaternionic analogue of the cross-ratio*. Advances in Applied Clifford Algebras, 22, 1041-1053, 2012, also arXiv:1112.0612.

[6] M.Schottenloher, *A mathematical introduction to conformal field theory*. Springer-Verlag, 759, Lecture Notes in Physics, 2nd, 2008.

[7] R.S.Ward, R.O.Wells, *Twistor Geometry and Field Theory*. Cambridge University Press, Cambridge Monographs on Mathematical Physics, 1991.

[8] I.M.Yaglom, *A simple non-Euclidean geometry and its physical basis*. Springer, 1979.

Reaction Time in General Recognition Theory

Aleina Wachtel *

Abstract

General Recognition Theory (GRT), developed by Townsend Laboratory for Mathematical Psychology, is used to examine multiple dimensions of stimuli in Signal Detection Theory (SDT). SDT mathematically models perceptual and decisional processes within cognitive psychology. We are interested in incorporating reaction time in addition to accuracy within GRT, and we hope to diminish the inability to distinguish failures of decisional orthogonality and perceptual separability. Through the use of separable and integral dimensional stimuli to manipulate perceptual separability, as well as experimentally induced failures of decisional orthogonality via positive and negative biases in response, we show that RT-GRT correctly models failures of the varieties of perceptual independence.

1 Mathematical Psychology

By modeling cognitive processes, the field of mathematical psychology seeks to make psychological concepts more rigorous. Using mathematics to model perceptual and decisional processes as separate mechanisms, we can begin to have a deeper understanding of what occurs within the brain between the moment a stimulus is shown to a participant to the moment the participant makes a decision about the stimulus.

1.1 Stimuli

For the sake of clarity, we define terminology associated with cognitive psychology experiments. *Dimensions* are the characteristics of stimuli that are varied purposefully within an experiment, e.g., color and shape. *Features*, on the other hand, are the characteristics of a specific stimulus, e.g., red and circle [1].

The dimensions of a stimulus can be classified into one of two categories: separable or integral. Although experimentally the dimensions of varying stimuli are always separate, meaning the values associated with each dimension must be numerically different or the stimuli would otherwise be considered identical, we use separable and integral to describe perceptual effects. Separable dimensions of stimuli do not interfere with each other perceptually, whereas integral dimensions demonstrate a failure of selective attention between the dimensions

*Harvey Mudd College, awachtel@g.hmc.edu

of the stimulus. For example, in previous psychological literature, it has been shown that color and shape are separable dimensions; saturation and brightness, however, are two integral dimensions [2]. Thus, a participant may believe that stimuli varying in brightness values are changing in saturation or vice versa.

The *salience* of a stimulus is the level of "perceptual difficulty" associated with that stimulus, meaning that a stimulus with high salience is easy for a participant to perceive, understand, and use in an identification or categorization task.

1.2 Signal Detection Theory

Signal Detection Theory (SDT) is a model used to separate perceptual effects from decisional effects. SDT specifically works within one dimensional stimuli, where most often participants identify the presence or absence of a particular signal alongside a constant, unidentifiable noise. We consider only accuracy, not reaction time, when using SDT to model perceptual and decisional processes [1].

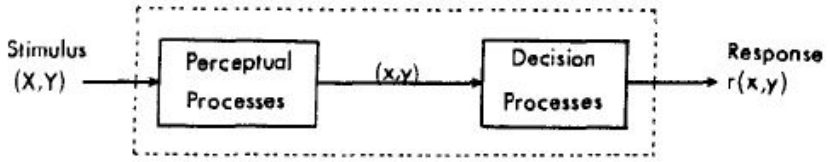


Figure 1.2.1: Seen above, the items within the dotted box represent what is occurring cognitively. We can empirically observe anything outside of this box, meaning the stimulus as well as the response are experimentally observable, where the response is merely a function of the stimulus shown. SDT furthers our understanding by mathematically modeling the intermediate stages of an identification task [1].

We assume that there exists a non-observable "perceptual space" that we can model using a simple Gaussian distribution. This is beneficial to help understand how, within our perceptual space, we can account for varying degrees of perceptual salience. For example, when shown a particular stimulus, at any given time the participant may identify the stimulus in a slightly different way. By representing what is perceptually occurring as a normal Gaussian distribution, we account for the variance in perception caused by cognitive noise. However, SDT is limited in its ability to model these effects when there exists more than one dimension to a stimulus.

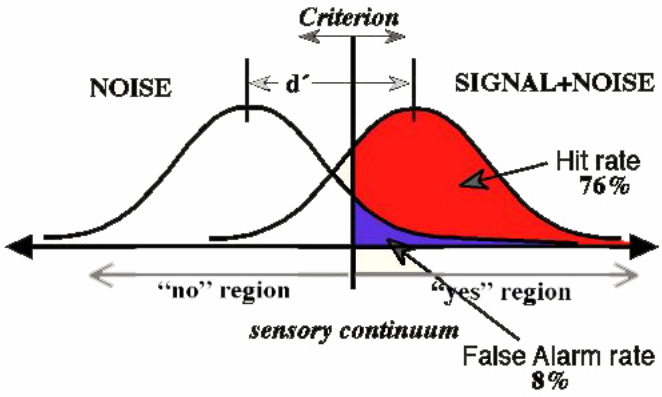


Figure 1.2.2: A visual representation of Signal Detection Theory in one dimension. In this task, participants are asked to identify the presence or absence of an auditory signal [4].

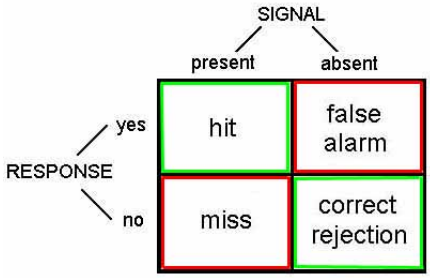


Figure 1.2.3: Matrix of possible stimulus-response combinations associated with the curves shown in Figure 2. The "Hit" and "Correct Rejection" boxes correspond to the large areas within the two curves where there is no overlap. The overlapping area in the "no region" is a miss, and the overlapping area in the "yes region" is a false alarm [3].

1.3 General Recognition Theory

In order to examine various stimuli with multiple dimensions, General Recognition Theory (GRT) was created [1]. Instead of being limited to solely classifying the presence or absence of a stimulus, GRT allows for the use of multidimensional stimuli within categorization and identification tasks, such as correctly locating a stimulus with several particular dimensions (e.g., red circle). In addition to permitting the use of multidimensional stimuli, GRT is instruction-independent, i.e., a wide variety of response instructions may be used within any given experiment. It is important to note that GRT only accounts for accuracy and does not consider reaction time [1].

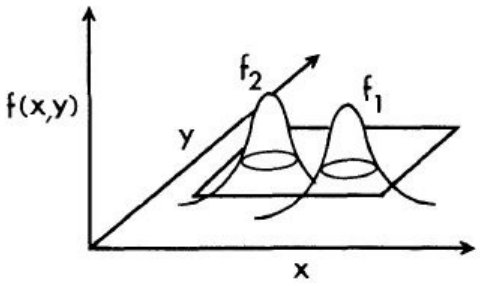


Figure 1.3.1: f_1 and f_2 are bivariate Gaussian distributions of perceptual effects corresponding to two stimuli, each composed of two dimensions, x and y. $f(x, y)$ is the probability value for each function at a specific point. Again, we assume there exists a Euclidean perceptual space in which these curves represent the perceptual processes occurring during an experiment [1].

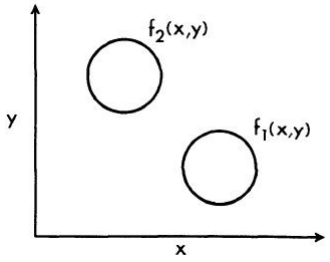


Figure 1.3.2: Contours of equal probability associated with Figure 4. These circles represent a perceptual space in which the participant recognizes certain stimuli to convey certain characteristics [1].

1.4 Varieties of Perceptual Independence

The components A and B of the two-dimensional stimulus AB are said to be independently perceived if the perception of each is in no way contingent on or interacts with the perception of the other [1].

$$P(AB) = P(A)P(B)$$

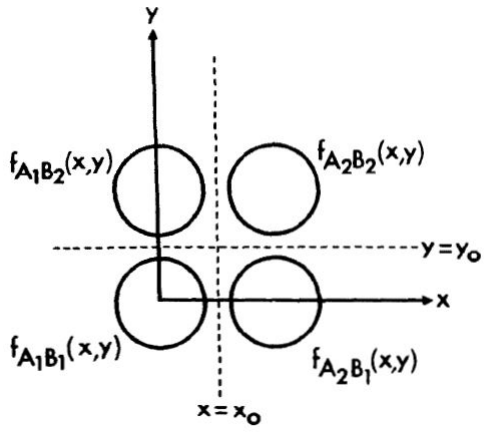


Figure 1.4.1: Represented here are the contours of equal probability and decision bounds from a complete identification experiment with stimuli A_1B_1, A_1B_2, A_2B_1, and A_2B_2, where A_i denotes a stimulus value on the x dimension and B_j denotes a stimulus value on the y dimension. Each circle represents the perception of each stimulus, and the dotted lines represent the decision bounds separating each possible response. In this figure, all varieties of perceptual independence hold [1].

What we seek to visually represent are the occurrences in which independence does not occur, i.e., failures of the varieties of perceptual independence: *perceptual separability*, *perceptual independence* (correlation), and *decisional orthogonality*. This summer, we focused solely on failure of perceptual separability and failures of decisional orthogonality. We define these terms below.

Definition 1.1 [Statistical Independence] Two arbitrary random variables W and Z are said to be *statistically independent* if the probability distribution of W does not depend on the value of Z (and vice versa) or, more formally, if and only if

$$f(W|Z) = f(W) \text{ and } f(Z|W) = (Z)$$

.

Definition 1.2 [Perceptual Independence] *Perceptual independence* of components A and B holds in stimulus A_iB_j if and only if the perceptual effects of A and B are statistically independent, that is, if and only if

$$f_{A_iB_j}(x,y) = g_{A_iB_j}(x)g_{A_iB_j}(y)$$

for all x and y and where $g_{A_iB_j}$ is the marginal distribution on dimension x when the stimulus is A_iB_i.

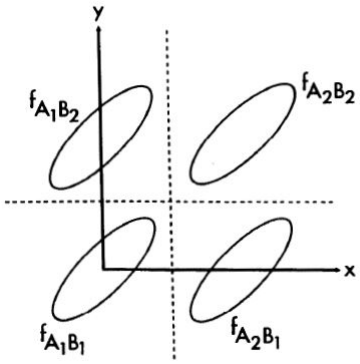


Figure 1.4.2: Contours of equal probability from a complete identification experiment in which perceptual separability and decisional orthogonality hold but perceptual independence fails. Depending on the slope associated with each ellipse, we can see a correlation between the dimensions of our stimuli [1].

Definition 1.3 [Perceptual Separability] Consider the complete identification experiment with stimuli A_1B_1, A_1B_2, A_2B_1, and A_2B_2. The components A and B are *perceptually separable* if the perceptual effect of one component does not depend on the level of the other, that is, if

$$g_{A_iB_1}(x) = g_{A_iB_2}(x) \qquad \text{for i} = 1, 2 \text{ and}$$
$$g_{A_1B_j}(y) = g_{A_2B_j}(y) \qquad \text{for j} = 1, 2$$

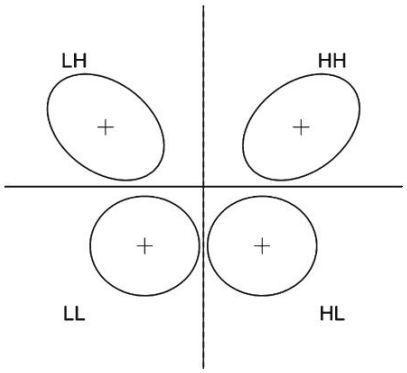


Figure 1.4.3: Contours of equal probability from a complete identification experiment in which decisional orthogonality holds but perceptual separability and independence fails. We see there exists a failure perceptual separability by examining the space between the stimuli in the LH and HH stimuli versus the LL and HL stimuli [5].

Definition 1.4 [Decisional Orthogonality] Consider the complete identification experiment with stimuli A_1B_1, A_1B_2, A_2B_1, and A_2B_2. The components A and B are *decisionally separable*, also referred to as *decisionally orthogonal*, if the decision about one component does not depend on the level of the other, that is, if the decision bounds in the general recognition theory are parallel to the coordinate axes [1]. In other words, we visually know when there exists a failure of decisional orthogonality when the dotted decisional bounds are no longer perpendicular.

1.5 Silbert & Thomas

In a paper published in 2013, Silbert and Thomas demonstrated failure of decisional orthogonality is not identifiable in the Gaussian GRT model with either of two response selection models. In this study, it was shown that "GRT relies on two basic assumptions to account for factorial identification data: (1) random perceptual effects [which we account for with a Gaussian distribution and circular perceptual contours] and (2) deterministic decision bounds that exhaustively partition perceptual space" [5]. Thus, we see the largest issues arise within our models when we assume there is initially no failure of decisional orthogonality.

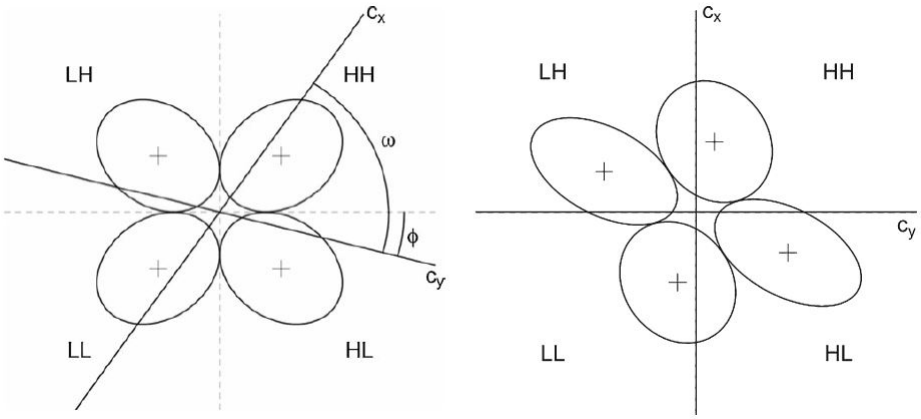


Figure 1.5.1: Seen on the left, we have perceptual separability with a linear failure of decisional orthogonality. On the right, however, after a rotated and sheared model configuration, we have a mathematically identical perceptual representation (up to area preserving linear transformations) in which decisional orthogonality holds, but there is a failure of perceptual separability. These two images are related by transformations in the special linear group of 2x2 matrices of determinant 1, a.k.a. SL(2,R) [5].

2 RT-GRT Project

2.1 Motivations

Considering these findings, we designed the Reaction Time in General Reconition Theory (RT-GRT) Project with two motivations in mind: (1) to try and incorporate reaction time into GRT, given that accuracy is normally the only aspect considered when analyzing response data, and (2) to address the concerns exhibited in the publication of Silbert and Thomas.

Within cognitive psychology, there are two aspects to data collected within an experiment: accuracy and reaction time. Accuracy is the percentage of correct responses in a categorization or identification task out of all given responses, and reaction time is the time (milliseconds) it takes for a participant to elicit a single response. It is important to consider both of these aspects, since it is only the consideration of both that allows for comparisons between data points of varying participants. For example, participant A with accuracy of 80 percent is not inherently "better" at a task than participant B with accuracy of 70 percent if the participant A had a much larger (slower) reaction time than participant B. Thus, we hope that by incorporating reaction time into GRT, it is possible that the shear model transform from a failure of decisional orthogonality to a failure of perceptual separability exhibited in Silbert and Thomas is no longer valid.

We hope to address these concerns over the course of two experiments. In these experiments, we induce failures of either one or both decisional orthogonality and perceptual separability. By generating these failures from the onset of each experiment, and thus knowing what the intended distributions should look like, we hope to demonstrate that there are necessary differences between the two representations of the failures of perceptual independence.

		Decisional Orthogonality?	
		YES	NO
Perceptual Separability?	YES	1a) no bias	1a) bias +/-
	NO	1b) no bias	1b) bias +/-

Figure 2.1.1: Graphic representation of the two intended experiments.

2.2 Experiment Design

This summer, Townsend Lab designed what we call Experiment (1a). In this full identification experiment, we maintain perceptual separability through the use

of separable dimensional stimuli and experimentally induce failures of decisional orthogonality through a series of biases. The task is to categorize four stimuli in accordance with the number pad on a keyboard, using the keys '1', '3', '7', and '9' (see Figure 14).

2.2.1 Creation of Stimuli

Previous psychological literature has demonstrated horizontal offset of a vertical line in a block stimulus and the saturation of a block stimulus are two separable dimensions [2]. Thus, there is no failure of selective attention between line placement and saturation, and we can vary each dimension without worry of a failure of perceptual separability.

Initially, we created four stimuli as seen in Figure 11. Each stimulus is 100x100 pixels in size, presented so as to subtend 2.5° visual angle. A visual angle denotes the size of a stimulus in relation to how far away a participant is seated from the stimulus (in this case, presented on a computer screen). However, after many trials of pilot testing, we found the categorization task to be too easy, i.e., participants' accuracy levels were too high to model efficiently with GRT.

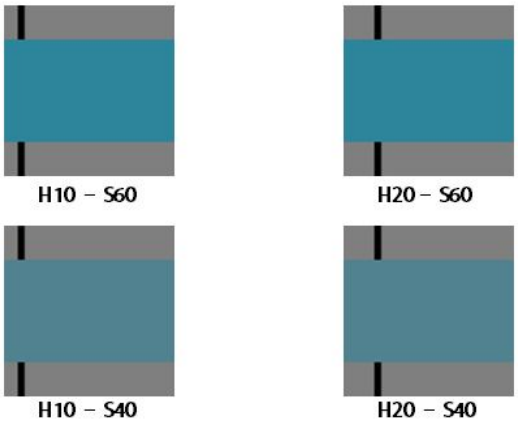


Figure 2.2.1: Original four stimuli created for Experiment (1a). The numerical value with "H" is the number of pixels of the displacement of the vertical line from the far left side of the stimulus. The numerical value with "S" is the saturation (chroma) value obtained from the Munsell color space.

Thus, through the Method of Constant Stimuli, we found the "perceptual thresholds" for each dimension and created the stimuli seen in Figure 12. To do so, we created stimuli with horizontal offset values of H10, H11, H12, ..., H20 with constant saturation and stimuli with saturation values S40, S42, S44, ..., S60 with constant horizontal offset values. We then implemented a focused

attention task, i.e., with one varying dimension, in which participants attempted to identify which stimulus was the lowest extreme (in this case H10 and S40) among the rest of the stimuli varying in one dimension. From the response data, we were able to determine where in the spectrum of line placement and saturation we could obtain the desirable 70 percent accuracy needed for GRT. In doing so, we found that using stimuli with a 4 pixel difference in line placement and 12 value difference in saturation in a full categorization task approximately generated the accuracy percentages we needed.

For the most intuitive response mapping, we arranged the stimuli such that from left to right the line placement shifts from left to right, and from bottom to top the block becomes more saturated. Thus, the lowest saturated stimulus with the line furthest to the left corresponds to the response '1', and the most saturated stimulus with the line furthest to the right corresponds to the response '9' (see Figure 14).

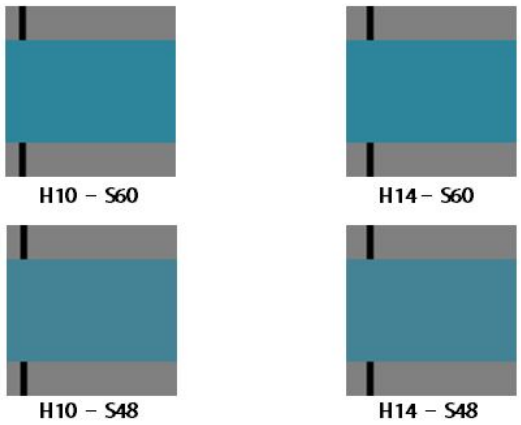


Figure 2.2.2: Final four stimuli created for Experiment (1a).

2.2.2 Data Collection Plan

Having established that our stimuli would sufficiently demonstrate perceptual separability, we embarked on the process of experimentally inducing failures of decisional orthogonality. We sought to do this through a series of biases. As shown in Figure 13, the experiment is designed to take five days. The first day, labeled as 'practice', is used to familiarize the participant with the categorization task. First, the participant completes focused attention tasks, where only one dimension of the stimulus varies. Then, the participant goes through six blocks of divided attention tasks, categorizing the stimuli according to both dimensions. All data from the first day is thrown out to account for practice effects, and only the data from the remaining four days of the experiment is analyzed.

Day 1 (practice)	Day 2 (no bias)	Day 3 (+/- bias)	Day 4 (-/+ bias)	Day 5 (no bias)
2 blocks - focused attention	8 blocks - divided attention	8 blocks - divided attention	8 blocks - divided attention	8 blocks - divided attention
6 blocks - divided attention	-----	-----	-----	-----

Figure 2.2.3: Day-by-day breakdown of the various experimental conditions.

The second day, the participant completes a normal GRT categorization task with no bias; the participant will also end the experiment with this condition. However, on the third and fourth days, the participant is randomly assigned either a positive bias on the third day and a negative bias on the fourth, or vice versa. It is within these two days we experimentally induce a failure of decisional orthogonality.

2.3 Decisional Orthogonality Biases

In order to manipulate decision bounds, we must alter the way in which a participant responds to a stimulus. It is important to note that we are not altering the stimuli themselves; rather, we are changing the number of points (and therefore dollars) awarded for particular stimulus-response combinations.

2.3.1 No Bias Condition

As stated previously, there are no biases on the focused attention tasks, as well as the divided attention tasks on day 1, 2, or 5. Figure 14 shows the instructions shown in the no bias condition.

On the screen, the participant is briefly shown one of the four possible stimuli. As fast and as accurately as possible, the participant must classify which of the four stimuli they believe they have just seen, and in response, presses one of the four appropriate keys. We incentivize this task by awarding positive points for each correct response and giving negative points for each incorrect response. The payoff matrix associated with the no bias condition can be seen in Figure 15A. Thus, along the diagonal, a participant receives 6 points for correctly categorizing the stimulus and loses 2 points for incorrectly doing so. GRT models this process, and theoretically we obtain Figure 15B. To ensure that the points create enough of an incentive to perform with high accuracy and fast reaction time, we randomly select 48 of the hundreds of trials the participant completes in one session. For each point earned in those trials, the participant is awarded 1 cent per point for a possible \$1 to \$3 bonus. This is in addition to the \$9 payment the participants receive each day upon successful completion of the experiment.

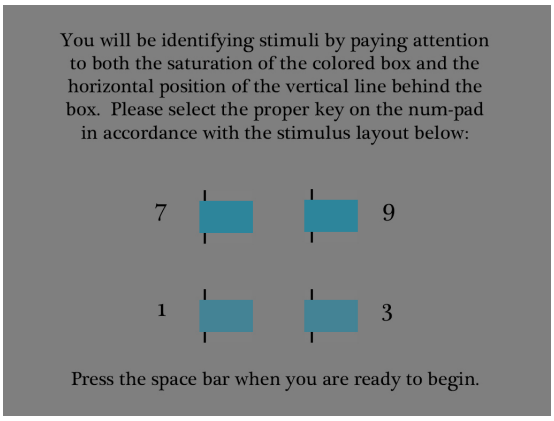
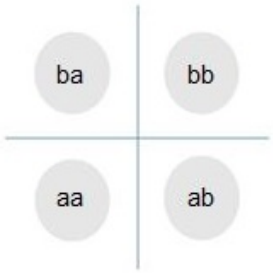


Figure 2.3.1: Instructions associated with the no bias condition.

	Responses				
Stimuli	**aa**	**ab**	**ba**	**bb**	**sum**
aa	6	-2	-2	-2	0
ab	-2	6	-2	-2	0
ba	-2	-2	6	-2	0
bb	-2	-2	-2	6	0
sum	0	0	0	0	

(A)



(B)

Figure 2.3.2: a) Payoff matrix associated with the no bias condition. b) Contours of equal probability associated with no bias condition.

2.3.2 Positive Bias Condition

The positive bias condition randomly occurs on either the third of fourth day of experimentation. This condition solely biases divided attention tasks. Figure 16 shows the instructions for the positive bias condition.

Like the no bias condition, the participant is briefly shown one of the four possible stimuli and the participant must classify each stimulus as quickly and accurately as possible. However, in order to experimentally induce a positive linear failure of decisional orthogonality, we must emphasize the responses associated with stimuli *aa* and *bb*. We do this by modifying the point-system incentive.

We still incentivize this task by awarding positive points for each correct response and giving negative points for each incorrect response, however, as seen in the payoff matrix associated with the positive bias condition in Figure 17A, the point values are no longer consistent.

Instead, when the emphasized responses are correct, the participant earns 7 points (see emphasized columns) rather than the 5 points a participant can earn for correctly categorizing the other two stimuli. Additionally, if a participant responds one of the two emphasized responses incorrectly, i.e., the participant presses '1' or '9' when the stimulus shown corresponds to the response of '7' or '3', the participant only loses 1 point. The larger penalty of 3 points is given when the participant responds with an unemphasized response incorrectly. This point system promotes a strategy in which the participant responds with the '1' or '9' keystroke more often since responding in this way is the most advantageous way to earn the most points.

Like the no bias condition, we randomly select 48 of the trials the participant completes and award 1 cent per point. Since the participant does not know which 48 trials will be selected, this ensures that the participant continues to perform as well as possible throughout the entire session.

If we can successfully induce this type of failure of decisional orthogonality, GRT will model the contours of equal probability for this condition as seen in Figure 17B.

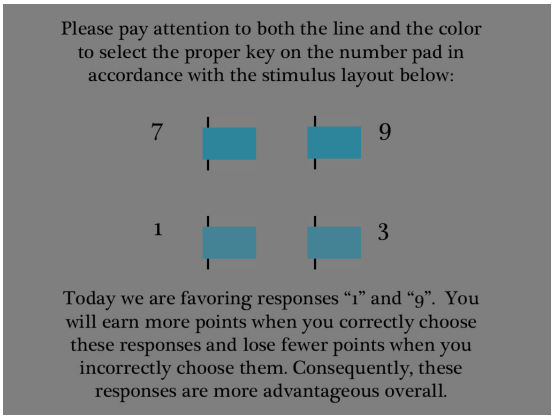
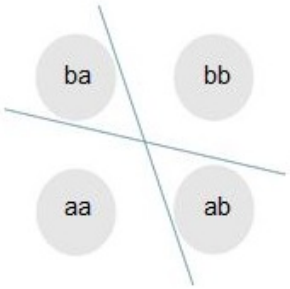


Figure 2.3.3: Instructions associated with the positive bias condition.

	Responses				
Stimuli	aa	ab	ba	bb	sum
aa	7	-3	-3	-1	0
ab	-1	5	-3	-1	0
ba	-1	-3	5	-1	0
bb	-1	-3	-3	7	0
sum	4	-4	-4	4	

(A)



(B)

Figure 2.3.4: a) Payoff matrix associated with the positive bias condition. Note the columns associated with stimuli *aa* and *bb* are favored in response with a value of positive (+) 4. b) Contours of equal probability associated with positive bias condition. The regions associated with stimuli *aa* and *bb* are larger than the regions associated with the other two stimuli.

2.3.3 Negative Bias Condition

The negative bias condition is completely analogous to the positive bias condition. The participant completes the negative bias condition experiment on whichever day was not allotted for the positive bias condition. Figure 18 shows the instructions for this condition.

We modify the point system such that the other two responses are emphasized in exactly the same way. The payoff matrix associated with negative bias condition can be seen in Figure 19A. With the same monetary bonuses for 48 of the trials, we hope that GRT models the contours of equal probability for this condition as seen in Figure 19B.

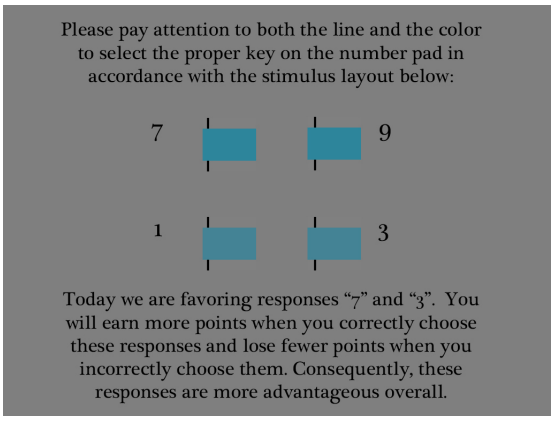
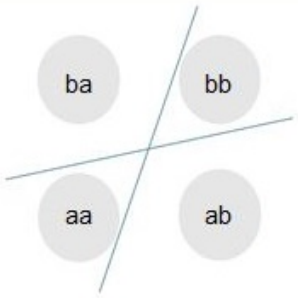


Figure 2.3.5: Instructions associated with the negative bias condition.

	Responses				
Stimuli	aa	ab	ba	bb	sum
aa	5	-1	-1	5	0
ab	-3	7	7	-3	0
ba	-3	-1	-1	-3	0
bb	-3	-1	-1	-3	0
sum	-4	4	4	-4	

(A)



(B)

Figure 2.3.6: a) Payoff matrix associated with the negative bias condition. Note the columns associated with stimuli *ab* and *ba* are favored in response with a value of positive (+) 4. b) Contours of equal probability associated with negative bias condition. The regions associated with stimuli *ab* and *ba* are larger than the regions associated with the other two stimuli.

3 Discussion

By collecting data from all three conditions outlined above, we hope to show that RT-GRT, by including reaction time, successfully distinguishes between failures of perceptual separability and failures of decisional orthogonality. Currently we have completed the experimentation design for Experiment (1a), dealing with separable dimensions, but upon completion of the collection of our data, we

wish to begin designing Experiment (1b) with the use of integral dimensional stimuli. As shown in Figure 20, we will once again use the method of constant stimuli to create appropriate stimuli of varying brightness and saturation for a nearly identical categorization task.

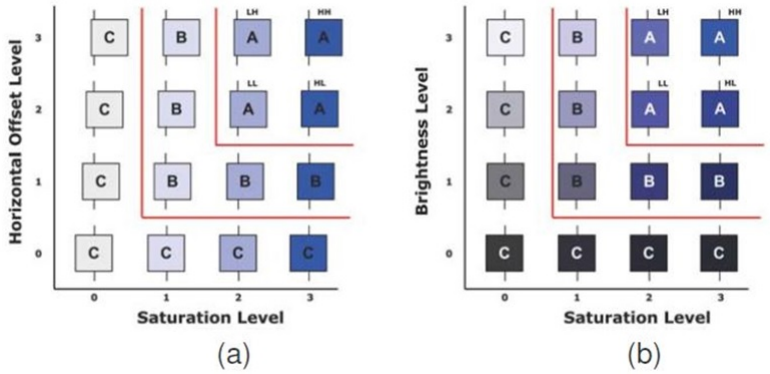


Figure 3.0.7: a) Experiment (1a) with separable dimensions of horizontal offset of vertical line and saturation. b) Experiment (1b) with integral dimensions of brightness and saturation

Thus far we have collected data points from three participants over the course of two weeks' worth of pilot testing. If RT-GRT models the failures of decisional orthogonality as expected, we plan to pursue with finalized participant experimentation.

Acknowledgments

I would like to thank everyone in Townsend Laboratory for Mathematical Psychology: Devin Burns, without whom this project could not have even been started; James Townsend, my advisor who so graciously invited me to work here at Indiana University over the summer; Haiyuan Yang, Arash Khodadadi, and Robert Hawkins, three phenomenal resources in the lab; and Patricia Knapp and Meelia Palakal, whose help to run participants for pilot testing was extremely appreciated! I would also like to thank the NSF for funding this research program, as well as the Department of Mathematics and the Department of Psychological and Brain Sciences at Indiana University, Bloomington. Additional thanks to Chris Connell for running the IU Math REU program, as well as Annie Carter for assistance on this report.

References

[1] F. Gregory Ashby & James T. Townsend: Varieties of Perceptual Independence. American Psychological Association. *Psychological Review*, 1986, Vol. 93, No. 2, 154-179.

[2] Mario Fific, Robert M. Nosofsky, & James T. Townsend: Information-Processing Architectures in Multidimensional Classification: A Validation Test of the Systems Factorial Technology. *Journal of Human Experimental Psychology: Human Perception and Performance*, 2008, 34, 356-375.

[3] David Heeger: Signal Detection Theory. Department of Psychology, Stanford University, 1998. `http://www-psych.stanford.edu/~lera/psych115s/notes/signal/`

[4] The Theory of Signal Detection. RIT Center for Imaging Science. `http://www.cis.rit.edu/people/faculty/montag/vandplite/pages/chap_5/ch5p1.html`

[5] Noah H. Silbert & Robin D. Thomas: Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Society*, 2012. *Psychon Bull Review* (2013) 20:1-20.

The CW-Complex of Translation Surfaces of Genus 2 with One Singularity

Elizabeth Winkelman *

Abstract

A well-rounded translation surface X of genus 2 with one singularity of 6π can be described as a set of closed planar eight-bar linkages. A linkage L in this set can be regarded as a collection of eight vectors $\{\vec{v_1}, ..., \vec{v_8}\}$ in the plane whose sum equals zero. This paper will focus on the subset of unit-length linkages with an interior distance of one between distinct vertices. Each such linkage arises as the minimal vector decomposition of X. The configuration space of linkages satisfying these conditions is a CW complex whose geometrical and topological properties will be presented.

1 Introduction

A translation surface X is a topological surface whose transition functions are translations. Associated to a translation surface is a natural notion of arclength. From this notion of length we can develop the concept of a minimal vector decomposition for a translation surface. Given a set α of arcs on a translation surface X, one defines $\ell(\alpha) = \inf\{\ell(\gamma)|\gamma \in \ell\}$. Each homology class $\alpha \in H_1(X)$ may be regarded as a collection of loops in the surface that pass through the singularity x_0. A nonzero homology class α is said to be *minimal* if for all other nonzero homology classes β we have $\ell(\beta) \geq \ell(\alpha)$. If α is minimal then it contains a unique arc that realizes the infimum. X is said to be *well-rounded* if the set of $\alpha \in H_1(X)$ which are minimal spans $H_1(X)$.

If X is a well-rounded surface of genus 2, then there will be at least four minimal homology classes and hence at least four shortest arcs representing these homology classes. Note that each of these arcs has the same length. These arcs will create closed loops in X, so the surface can be cut apart along these arcs to yield polygonal regions in $\mathbb{R}^2$ with equal length sides [2].

If there are exactly four minimal homology classes then there is only one polygon K. K can be rescaled so that it has unit-length sides. The edges of K correspond to the arcs along which we cut the surface; two edges for each arc. To rebuild the surface X from K we paste together the two sides that came from cutting the same arc.

A linkage L is the minimal vector decomposition of X if the vectors which form L define the edges of K. The vectors of L form the set $\{\vec{v_i}\} \in \mathbb{R}^2$ for

*University of Rochester

$i \in (1, 8)$ connected tip to tail counter-clockwise. Since K is one piece, the linkage is closed and so $\sum \vec{v}_i = 0$. The edge identification of K used to form X via pasting is represented in L by $\vec{v}_i = -\vec{v}_{i+4(mod8)}$ where the edges of K described by $\vec{v}_i$ and $\vec{v}_{i+4(mod8)}$ are pasted together.

This paper will focus on the set of closed planar unit-length eight-bar linkages which arise as the minimal vector decomposition of a well-rounded translation surface of genus 2 with a singularity of 6π.

Due to the minimality of the classes in $H_1(X)$, the interior distance between distinct vertices is at least one (the minimal length). The set of all such linkages has a configuration space which is a subset of $\mathbb{R}^{16}$ and as a result our subset has a natural subspace topology which we will explore.

To better understand these linkages and their configuration space, we next explore the implications of the above conditions on the structure of a linkage.

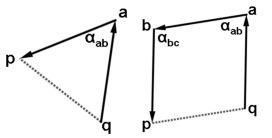
1.1 Interior Distance

The requirement that the interior distances between distinct vertices is no less than one determines the conditions on the angles between vectors. There are two simple cases and one slightly complicated case of how to form linkages in the configuration space.

Consider a subset of the vectors describing a linkage. Require that the vectors be connected, so that they form a chain. Let α_{ij} denote the angle between $\vec{v}_i$ and $\vec{v}_j$ where $\vec{v}_i$ is encountered first. (When determining whether $\vec{v}_i$ comes before or after $\vec{v}_j$, linkages are labeled in a counter-clockwise manner.) If $j = i + 4(mod8)$ then $\alpha_{ij} = 0$.

Define q and p to be vertices at the start and end of the vector chain with the line segment connecting them denoted qp. Then the requirement that linkages have an interior distance no less than one means that for all sets of vector chains, $|qp| \geq 1$.

Proposition 1.1 *If $|qp| \geq 1$, then*
i) for a vector chain of two vectors $\{\vec{v}_a, \vec{v}_b\}$, $\alpha_{qp} \geq \frac{\pi}{3}$,
ii) for a vector chain of three vectors $\{\vec{v}_a, \vec{v}_b, \vec{v}_c\}$, $\alpha_{ab} + \alpha_{bc} \geq \pi$.



Proof

i) Let $\vec{v}_a$ and $\vec{v}_b$ meet at the vertex a. Then the vertices q, a, p form the triangle qap.

If $|qp| = 1$, then the triangle qap is an equilateral triangle with unit length sides. Thus all interior angles are $\frac{\pi}{3}$, so $\alpha_{ab} = \frac{\pi}{3}$.

If $|qp| > 1$, then the triangle qap is an isosceles triangle. Since $|\vec{v}_a| = |\vec{v}_b| = 1$, qp is the longest side of triangle qap and so the angle opposite will be the largest interior angle of the triangle. This is α_{ab}.

Suppose $\alpha_{ab} < \frac{\pi}{3}$. Then the sum of the angles at vertex q and vertex p will be greater than $\frac{2\pi}{3}$ since the sum of all the angles must be π. This requires at

least one of those two angles to be greater than $\frac{\pi}{3}$; contradicting that α_{ab} is the largest interior angle. Thus it must be that $\alpha_{ab} > \frac{\pi}{3}$.

Therefore $\alpha_{ab} \geq \frac{\pi}{3}$ if the vector chain has just two vectors.

ii) Let $\vec{v}_a$ connect to $\vec{v}_b$ which connects to$\vec{v}_c$ at vertices a and b. Then the vertices q, a, b, p form a quadrilateral.

If $|qp| = 1$, then quadrilateral $qabp$ is a rhombus with unit-length sides. Opposite angles will be equivalent, yielding the relation: $2\alpha_{ab} + 2\alpha_{bc} = 2\pi$, so $\alpha_{ab} + \alpha_{bc} = \pi$.

If $|qp| > 1$, then for all possible values $\gamma = \alpha_{ab} + \alpha_{bc}$ there exists trapezoid $qabp$ (which will never be a rhombus) such that $\alpha_{ab} = \alpha_{bc} = \frac{\pi}{2} + \delta$. Since lines through vertices a and b, perpendicular to line qp will intersect qp at a' and b' respectively, the triangle qaa' and the triangle pbb' will be right triangles and so $\delta > 0$. Thus $\alpha_{ab} + \alpha_{bc} > \pi$ if $|qp| > 1$.

Therefore $\alpha_{ab} + \alpha_{bc} \geq \pi$ if the vector chain has just three vectors.

□

Since each linkage in the configuration space will have four pairs of parallel vectors, Proposition 1.1 provides some insight into the structure of a linkage L in the space. If L is in case (i), then $\vec{v}_a$ and $\vec{v}_b$ will not be one of the vector pairs. If L is in case (ii) with $|qp| = 1$, then $\vec{v}_a$ and $\vec{v}_c$ will be parallel and could be one of the vector pairs describing L, but not necessarily, and $\vec{v}_b$, as in case (i) will not be paired with either $\vec{v}_a$ or $\vec{v}_c$.

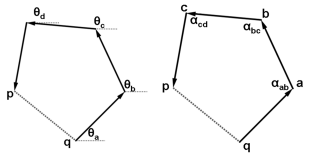
The third case is when there are four vectors $(\vec{v}_a, \vec{v}_b, \vec{v}_c, \vec{v}_d)$ in the vector chain. To solve this case we will use a different notation for describing the relationship between the vectors in the chain. A unit vector $\vec{v}_\alpha$ is described by $e^{i\theta_\alpha}$ in the complex plane. The α_{ij} used in Proposition 1.1 can be recovered from the theta notion by the relation:

$$\alpha_{ij} = \pi - |\theta_j - \theta_i| .$$

Since vectors are taken in a counter-clockwise manner, this can be written as $\alpha_{ij} = \pi - (\theta_j - \theta_i)$.

Proposition 1.2 *If $|qp| \geq 1$, then*
iii) for a vector chain of four vectors $\{\vec{v}_a, \vec{v}_b, \vec{v}_c, \vec{v}_d\}$,

$$cos(\alpha_{ab}) + cos(\alpha_{bc}) + cos(\alpha_{dc}) - cos(\alpha_{ab} + \alpha_{bc}) - cos(\alpha_{bc} + \alpha_{cd}) + cos(\alpha_{ab} + \alpha_{bc} + \alpha_{cd}) \leq \frac{3}{2}.$$



Proof

Let $\vec{v}_{qp}$ describe the vector from point q to point p. Then $|qp| = |\vec{v}_{qp}| \geq 1$. The sum of the four vectors is $e^{i\theta_a} + e^{i\theta_b} + e^{i\theta_c} + e^{i\theta_d} = \vec{v}_{qp}$, so

$$\left|e^{i\theta_a} + e^{i\theta_b} + e^{i\theta_c} + e^{i\theta_d}\right|^2 = |\vec{v}_{qp}|^2 \geq 1.$$

$\left|e^{i\theta_a} + e^{i\theta_b} + e^{i\theta_c} + e^{i\theta_d}\right|^2 = (e^{i\theta_a}+e^{i\theta_b}+e^{i\theta_c}+e^{i\theta_d})(e^{-i\theta_a}+e^{-i\theta_b}+e^{-i\theta_c}+e^{-i\theta_d})$. The inequality simplifies as follows:

$cos(\theta_b - \theta_a) + cos(\theta_c - \theta_a) + cos(\theta_c - \theta_b) + cos(\theta_d - \theta_a) + cos(\theta_d - \theta_b) + cos(\theta_d - \theta_c) \leq \frac{-3}{2}$

$cos(\pi - \alpha_{ab}) + cos(2\pi - \alpha_{bc} - \alpha_{ab}) + cos(\pi - \alpha_{bc}) + cos(3\pi - \alpha_{cd} - \alpha_{bc} - \alpha_{ab}) + cos(2\pi - \alpha_{cd} - \alpha_{bc}) + cos(\pi - \alpha_{dc}) \leq \frac{-3}{2}$

$-cos(\alpha_{ab}) + cos(\alpha_{bc} + \alpha_{ab}) - cos(\alpha_{bc}) - cos(\alpha_{cd} + \alpha_{bc} + \alpha_{ab}) + cos(\alpha_{cd} + \alpha_{bc}) - cos(\alpha_{dc}) \leq \frac{-3}{2}$

$cos(\alpha_{ab}) + cos(\alpha_{bc}) + cos(\alpha_{dc}) - cos(\alpha_{ab} + \alpha_{bc}) - cos(\alpha_{bc} + \alpha_{cd}) + cos(\alpha_{ab} + \alpha_{bc} + \alpha_{cd}) \leq \frac{3}{2}$

□

These three cases are all of the cases since if there were a chain of five vectors, then since a linkage is closed, on one side of the line qp would be five vectors $(\vec{v}_a, \vec{v}_b, \vec{v}_c, \vec{v}_d, \vec{v}_e)$ and so three connected vectors would be on the other side and fall within case (ii). Similarly, case (i) implies that a chain of six vectors lies on the other side of the line qp.

1.2 Genus 2 Surface

If a linkage L is in the configuration space, it is the minimal vector decomposition of a well-rounded translation surface of genus 2 with a singularity of 6π. L is closed and formed by eight vectors, so L is an octagon with an interior angle sum of 6π. In order for L to correspond to a genus 2 surface with a singularity of 6π, each vertex of L is identified with every other vertex of L.

As previously mentioned, L consists of four vectors and their additive inverses. Each vector and its inverse describes two edges in the polygonal region K which are pasted together to form the translation surface X. Since $\vec{v}_i = -\vec{v}_{i+4(mod8)}$, each linkage in the configuration space can be described by the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ of four vectors and a matching.

L be represented by the word formed by the indicies of the four vectors which describe it, $\{1, 2, 3, 4\}$, with every element appearing exactly twice. For example, if L is hypereliptic (that is has 180^o rotational symmetry), then the word 12341234 describes L, up to a relabeling of the vector pairs.

The way the vertices of a linkage L are identified is by looking at how the corresponding translation surface folds up. Each vector of L has an orientation and so the vertex that would be at the tip of $\vec{v_1}$ is the same as the vertex at the tail of $\vec{v_5}$ since by the matchings, the two vectors represent the same edge on the polygonal region described by L. This same vertex is at the tail of the vector following $\vec{v}_1$, say $\vec{v}_2$, and can be identified to another vertex on L which is at the tip of $\vec{v}_6$. This is done for all the vertices of L. All of the vertices must be identified together if L is in the configuration space. Figure 1.2.1 shows some examples of eight-bar linkages with their vertices identified.

It turns out that if a linkage L is in the configuration space then, when two adjacent vectors of different labels appear in a certain order, if they happen to be next to each other at a different point in L, then they cannot appear in the reverse order. For example, L cannot be represented by the word 12342134 because this means that in the structure of L, $\vec{v}_1$ is followed by $\vec{v}_2$ but later on $\vec{v}_2$ is followed by $\vec{v}_1$. Checking the vertex identifications, the linkage with word

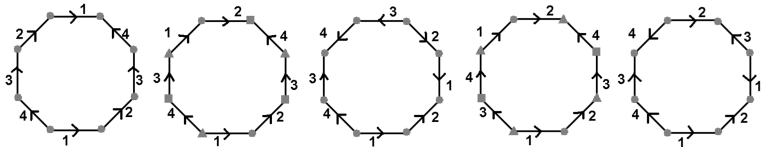


Figure 1.2.1: These octagons represent a linkage where the edge numbers correspond to vectors of that same label. The property that vectors of the same label are parallel is being ignored so as to focus on how to determine if a linkage is the decomposition of translation surface of genus 2 with a singularity of 6π. The arrow on an edge indicates the edge's orientation with respect to pasting, so some edges may be oriented in a different direction than the vector which describes it.

12342134 corresponds to a translation surface with three singularities of $\frac{9\pi}{4}$, $\frac{9\pi}{4}$ and $\frac{3\pi}{2}$ (as shown in Figure 1.2.1).

1.3 Allowable Linkages

Combining together all of the requirements for a linkage to be in the configuration space, there are only three types of linkages, up to a relabeling, which are allowed. These linkages have word representations 12341234, 12341423, and 12132434. Note, as seen in Figure 1.2.1, a linkage with word 12123434 satisfies the genus 2 requirement but not all interior distances between distinct vertices will be greater than or equal to one since 1212 corresponds to a closed four-bar linkage.

As a notational simplification, each of the allowed three linkage types can be written as being of type $(3,3,3,3), (3,2,2,1)$ and $(1,2,2,1)$. A linkage being of type (a,b,c,d) means that there are a edges in the linkage between the first edge and its second appearance and similarly for the other edges. When determining the separation between edges of the same label, the minimum separation is the one chosen. For example, a linkage of type 12341423 is of type $(3,2,2,1)$ but there are five edges between the third edge (3) and its second appearance, however starting at its second appearance yields a separation of two.

2 Constructing the Configuration Space

Let linkage L be in the configuration space. L is constructed from the vectors: $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$. Each linkage type can be described by a set of the α_{ij}, in a similar way as how only four of the eight vectors were needed to describe a linkage.

Proposition 2.1 *For each linkage type, only three angles are needed to know the entire structure of a linkage in the configuration space.*

Proof

A linkage in the configuration space is closed with eight unit-length sides, so it follows that only the angles between the vectors are needed to construct a linkage if its type is known, since that will provide information on the order of the vectors. Consider each linkage type separately:

$(3,3,3,3)$: If L is a linkage of type $(3,3,3,3)$, then L can be written as the word 12341234. The set of interior angles of L is $\{\alpha_{12}, \alpha_{23}, \alpha_{34}, \alpha_{41}\}$. Since the sum of the interior angles in L is 6π, it follows that $\alpha_{12} + \alpha_{23} + \alpha_{34} + \alpha_{41} = 3\pi$. Furthermore, $\alpha_{41} = 3\pi - (\alpha_{12} + \alpha_{23} + \alpha_{34})$, so only $\alpha_{12}, \alpha_{23}, \alpha_{34}$ are needed to describe L.

$(3,2,2,1)$: If L is a linkage of type $(3,2,2,1)$, then L can be written as the word 12341423. The set of interior angles of L is $\{\alpha_{12}, \alpha_{23}, \alpha_{34}, \alpha_{41}, \alpha_{14}, \alpha_{42}, \alpha_{31}\}$. The 414 part of L is in case (ii) of Proposition 1.1 and so $\alpha_{41} + \alpha_{14} = \pi$. It follows that only one of the two angles is needed to describe the 414 portion of L. Furthermore, if the vertex at α_{34} is labeled a and the vertex at α_{42} is labeled b, then let the vector from a to b be $\vec{v}_5$. Since the 414 part of L is in case (ii) of Proposition 1.1, $\vec{v}_5$ is parallel to $\vec{v}_1$. The word 4145 corresponds to a rhombus and the word 123523 corresponds to a hyperelliptic hexagon with equal-length sides. Therefore $\alpha_{12} = \alpha_{52}$ and $\alpha_{31} = \alpha_{35}$, so α_{12}, α_{23}, and α_{35} are the only angles needed to describe 123523. Additionally, $\alpha_{14} = \alpha_{54}$ and knowing α_{14} means knowing the entirety of 4145. Then $\alpha_{12}, \alpha_{23}, \alpha_{35}$ and α_{14} will describe L. Since $\alpha_{34} = \alpha_{35} + \alpha_{54} = \alpha_{35} + \alpha_{14}$, only $\alpha_{12}, \alpha_{23}, \alpha_{34}$ are needed to describe L.

$(1,2,2,1)$: If L is a linkage of type $(1,2,2,1)$, then L can be written as the word 12132434. The interior angles of L are all of unique label and form the set $\{\alpha_{12}, \alpha_{21}, \alpha_{13}, \alpha_{32}, \alpha_{24}, \alpha_{43}, \alpha_{34}, \alpha_{41}\}$. However, the structure of L is such that if $\vec{v_5}$ goes from the vertex at α_{13} to the one at α_{41} and $\vec{v_6}$ is defined similarly, only between α_{24} and α_{41}, then the words 1215, 3265, and 4346 all correspond to a rhombus and are thus in case (ii) of Proposition 1.1. As seen with linkages of type $(3,2,2,1)$, only one interior angle is needed to describe a rhombus. Thus, only $\alpha_{12}, \alpha_{32}, \alpha_{34}$ are needed to describe L.

□

These sets of three angles can be used as coordinates in $\mathbb{R}^3$. Figure 2.0.1 shows which α_{ij} is used for each axis. Since each linkage type has a different construction, they cannot be plotted together using this method since the three coordinates provide no information about the structure of a linkage.

The boundary of the configuration space will be when one of the inequalities in Propositions 1.1 and 1.2 is an equality. This occurs when there exists a distance of length one between at least two nonadjacent vertices within a linkage.

2.1 The Boundary

Each L in the configuration space must satisfy certain requirements (previously mentioned). From these requirements a certain number of degrees of freedom can be assigned to L. If there exists an interior distance of length one between at least two nonadjacent vertices of L, then those vertices become locked at that position such that if any interior angle of L were to be changed to form linkage

	$(3,3,3,3)$	$(3,2,2,1)$	$(1,2,2,1)$
x	α_{12}	α_{12}	α_{12}
y	α_{23}	α_{23}	α_{32}
z	α_{34}	α_{34}	α_{34}

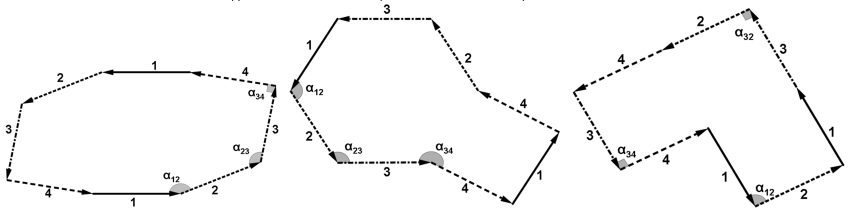


Figure 2.0.1: The table shows which α_{ij} was chosen to be used in the coordinate system for describing linkages of each type and which coordinate in $\mathbb{R}^3$ that α_{ij} corresponds to. An example of each linkage type is given with the three angles needed to describe it highlighted.

L', then L' will have a distance of one between the same vertices for which this is the case on L. The degrees of freedom of L are the number of angles which can be changed without breaking any locked distances that L may have. Since changing one angle will require at least one other angle to change, a linkage in the configuration space can have a maximum of three degrees of freedom.

The interior is the set of angles that satisfy the strict inequalities. It is the set of linkages which are described by the strict inequalities in Propositions 1.1 and 1.2. This set will be an open set bounded by the set of linkages in the configuration space with less than three degrees of freedom. An easy visualization is to picture where a point p is able to move. If p is on the interior of the configuration space then from p there is a point p' in the direction of a linear combination of the vectors $\vec{e}_i, \vec{e}_j, \vec{e}_k$ which is still on the interior, where $\vec{e}_i, \vec{e}_j, \vec{e}_k$ will always denote linearly independent unit vectors.

It then follows that a linkage with two degrees of freedom will be correspond to point p on a face of the configuration space. From p, only traveling in the direction of some linear combination of $\vec{e}_i, \vec{e}_j$ will go from p to p' still on the face. The face does not have to be planar for this fact to be true. For instance: If a person is walking on a hill, they can only choose a linear combination of going forward or turning to the right to change their location and maintain being on the hill regardless of whether they are traveling upward or downward with respect to some reference position.

Using the same logic, it can be seen that a linkage on an edge of the configuration space will have one degree of freedom and if on a vertex it will have no degrees of freedom.

The next three subsections describe in detail the structure of the configuration space for each linkage type. This will be three separate configuration spaces for which their union is the configuration space which has been referred

to up until now.

2.2 Linkage Type: $(3, 3, 3, 3)$

Let the linkages of type $(3, 3, 3, 3)$ be linkages with a word representation of 12341234, so the configuration space P_0 is plotted via a $(\alpha_{12}, \alpha_{23}, \alpha_{34})$-coordinate system. Table 1 shows what the faces, edges, and vertices of P_0 will be. It can be seen that there will be 12 faces, 30 edges, and 20 vertices.

The relations between the faces, edges, and vertices mentioned in Table 1 can be used to determine that the faces F_{1_0} through F_{4_0} will be hexagonal, F_{5_0} through F_{8_0} will be pentagonal, and the remaining faces F_{9_0} through F_{12_0} will be quadrilateral.

Since the vertices have no degrees of freedom, all of their interior angles are known with certainty and so their α_{ij} can be determined. Doing this for all of the vertices will yield a plot like that in Figure 2.2.1.

Vertex	Coordinates	Vertex	Coordinates
V_{1_0}	$(\frac{\pi}{3}, \frac{4\pi}{3}, \frac{\pi}{3})$	V_{11_0}	$(\frac{\pi}{3}, \pi, \frac{\pi}{3})$
V_{2_0}	$(\frac{4\pi}{3}, \frac{\pi}{3}, \pi)$	V_{12_0}	$(\pi, \pi, \frac{2\pi}{3})$
V_{3_0}	$(\frac{\pi}{3}, \pi, \frac{\pi}{3})$	V_{13_0}	$(\frac{2\pi}{3}, \frac{2\pi}{3}, \frac{4\pi}{3})$
V_{4_0}	$(\pi, \frac{\pi}{3}, \frac{4\pi}{3})$	V_{14_0}	$(\frac{\pi}{3}, \frac{2\pi}{3}, \frac{2\pi}{3})$
V_{5_0}	$(\pi, \pi, \frac{\pi}{3})$	V_{15_0}	$(\frac{4\pi}{3}, \frac{\pi}{3}, \frac{2\pi}{3})$
V_{6_0}	$(\pi, \frac{\pi}{3}, \frac{2\pi}{3})$	V_{16_0}	$(\frac{2\pi}{3}, \frac{4\pi}{3}, \frac{\pi}{3})$
V_{7_0}	$(\frac{\pi}{3}, \frac{2\pi}{3}, \pi)$	V_{17_0}	$(\frac{4\pi}{3}, \frac{2\pi}{3}, \frac{2\pi}{3})$
V_{8_0}	$(\frac{2\pi}{3}, \pi, \pi)$	V_{18_0}	$(\frac{\pi}{3}, \frac{4\pi}{3}, \frac{2\pi}{3})$
V_{9_0}	$(\pi, \frac{2\pi}{3}, \frac{\pi}{3})$	V_{19_0}	$(\frac{2\pi}{3}, \frac{\pi}{3}, \frac{4\pi}{3})$
V_{10_0}	$(\frac{2\pi}{3}, \frac{\pi}{3}, \pi)$	V_{20_0}	$(\frac{2\pi}{3}, \frac{2\pi}{3}, \frac{\pi}{3})$

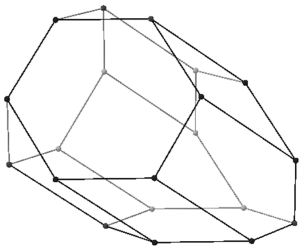


Figure 2.2.1: Coordinates for the vertices of P_0 and a plot of the vertices and edges.

To understand the properties of P_0, further investigation of the faces is needed. Knowing the curvature of each face will provide a better explanation as to the appearance of the configuration space in $\mathbb{R}^3$. The equations for each face are as follows:

$F_{1_0} : \alpha_{12} + \alpha_{23} + \alpha_{34} = \frac{8\pi}{3}$
$F_{2_0} : \alpha_{12} = \frac{\pi}{3}$
$F_{3_0} : \alpha_{23} = \frac{\pi}{3}$
$F_{4_0} : \alpha_{34} = \frac{\pi}{3}$

$F_{9_0} : \alpha_{12} + \alpha_{23} = 2\pi$
$F_{10_0} : \alpha_{23} + \alpha_{34} = 2\pi$
$F_{11_0} : \alpha_{12} + \alpha_{23} = \pi$
$F_{12_0} : \alpha_{23} + \alpha_{34} = \pi$

$F_{5_0} : -cos(\alpha_{12}) + cos(\alpha_{23}) - cos(\alpha_{34}) - cos(\alpha_{12} + \alpha_{23}) - cos(\alpha_{23} + \alpha_{34}) - cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) = \frac{-3}{2}$

F_{6_0} : $cos(\alpha_{12}) + cos(\alpha_{23}) - cos(\alpha_{34}) - cos(\alpha_{12} + \alpha_{23}) + cos(\alpha_{23} + \alpha_{34}) + cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) = \frac{3}{2}$

F_{7_0} : $-cos(\alpha_{12}) - cos(\alpha_{23}) - cos(\alpha_{34}) + cos(\alpha_{12} + \alpha_{23}) + cos(\alpha_{23} + \alpha_{34}) + cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) = \frac{-3}{2}$

F_{8_0} : $-cos(\alpha_{12}) + cos(\alpha_{23}) + cos(\alpha_{34}) + cos(\alpha_{12} + \alpha_{23}) - cos(\alpha_{23} + \alpha_{34}) + cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) = \frac{3}{2}$

This indicates that all of the faces are planar, except for faces $F_{5_0}, F_{6_0}, F_{7_0}, F_{8_0}$. ($F_{5_0}$ is graphed in Figure 2.2.2). Due to the curvature of these faces with respect to the rest of P_0, P_0 is not convex. Knowing the exterior of P_0, does not guarantee that the interior is well-behaved.

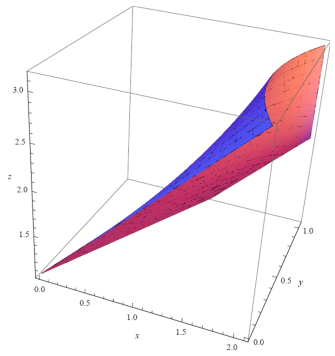


Figure 2.2.2: This is a plot of the equation describing the linkages on F_{5_0} with bounds determined by the vertices of the face. Faces F_{6_0}, F_{7_0}, F_{8_0} will have similar plots.

Proposition 2.1 *All the points $p \in \mathbb{R}^3$ satisfying the succeeding inequalities will correspond to a linkage L of type $(3,3,3,3)$ which satisfies the requirements outlined in Section 1. These inequalities describe the subset of $\mathbb{R}^3$ which is bound by the faces of P_0.*

(1) $\alpha_{12} + \alpha_{23} + \alpha_{34} \leq \frac{8\pi}{3}$

(2) $\alpha_{ij} \geq \frac{\pi}{3}$ *for all* i, j

(3) $\pi \leq \alpha_{12} + \alpha_{23} \leq 2\pi$

(4) $\pi \leq \alpha_{23} + \alpha_{34} \leq 2\pi$

(5) $-cos(\alpha_{12}) + cos(\alpha_{23}) - cos(\alpha_{34}) - cos(\alpha_{12} + \alpha_{23}) - cos(\alpha_{23} + \alpha_{34}) - cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) \geq \frac{-3}{2}$

(6) $cos(\alpha_{12}) + cos(\alpha_{23}) - cos(\alpha_{34}) - cos(\alpha_{12} + \alpha_{23}) + cos(\alpha_{23} + \alpha_{34}) + cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) \leq \frac{3}{2}$

(7) $-cos(\alpha_{12}) - cos(\alpha_{23}) - cos(\alpha_{34}) + cos(\alpha_{12} + \alpha_{23}) + cos(\alpha_{23} + \alpha_{34}) + cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) \geq \frac{-3}{2}$

(8) $-cos(\alpha_{12}) + cos(\alpha_{23}) + cos(\alpha_{34}) + cos(\alpha_{12} + \alpha_{23}) - cos(\alpha_{23} + \alpha_{34}) + cos(\alpha_{12} + \alpha_{23} + \alpha_{34}) \leq \frac{3}{2}$

Proof

A linkage L of type $(3,3,3,3)$ will be closed and have an interior angle sum of 6π with $\alpha_{12}+\alpha_{23}+\alpha_{34}+\alpha_{41}=3\pi$. By Proposition 1.1 we know that $\alpha_{ij} \geq \frac{\pi}{3}$ for all i,j, so (2) is satisfied. This also means that

$$\alpha_{12}+\alpha_{23}+\alpha_{34}=3\pi-\alpha_{41} \leq 3\pi-\frac{\pi}{3}=\frac{8\pi}{3},$$

so (1) is satisfied.

Additionally, Proposition 1.1 states that $\alpha_{ij}+\alpha_{jk} \geq \pi$ for all i,j,k. So

$$\alpha_{ij}+\alpha_{jk}=3\pi-\alpha_{kl}-\alpha_{li} \leq 3\pi-\pi=2\pi,$$

so (3) and (4) are satisfied.

Proposition 1.2 states that $cos(\alpha_{ij})+cos(\alpha_{jk})+cos(\alpha_{kl})-cos(\alpha_{ij}+\alpha_{jk})-cos(\alpha_{jk}+\alpha_{kl})+cos(\alpha_{ij}+\alpha_{jk}+\alpha_{kl}) \leq \frac{3}{2}$. By following the process used to prove Proposition 1.2, it can be determined that (5), (6), (7), (8) are satisfied. The difference in signs arises from the vector orientations in L. That is (5), (6), (7), (8) will be of the form given in Proposition 1.2 if close attention was paid to whether α_{ij} was between $\vec{v}_i$ and $\vec{v}_j$ or if it was between $\vec{v}_{i+4(mod8)}$ and $\vec{v}_{j+4(mod8)}$ when determining the equation for the face each inequality comes from.

□

2.2.1 Fundamental Domain

Despite being the plot of all linkages of type $(3,3,3,3)$ which satisfy all of the requirements outlined in Section 1, P_0 is not the proper configuration space for such linkages. Due to the symmetry of linkages of type $(3,3,3,3)$, P_0 contains sets of points which correspond to the same linkage. This is because if a linkage $L \in P_0$ is represented by the word 12341234 and plots as the point$(\alpha_{12},\alpha_{23},\alpha_{34})$, then there exists linkage $L' \in P_0$ which plots as the point $(\alpha_{23},\alpha_{34},\alpha_{41})$ and thus is represented by the word 23412341, a cyclic permutation of the labeling of L. Every linkage $L \in P_0$ will plot as the set of points $\{(\alpha_{12},\alpha_{23},\alpha_{34}),(\alpha_{23},\alpha_{34},\alpha_{41}),(\alpha_{34},\alpha_{41},\alpha_{12}),(\alpha_{41},\alpha_{12},\alpha_{23})\}$, with not all the points necessarily unique. An affine transformation can be applied to an element of this set to get another element in the set. Define this transformation as:

$$T(\vec{\alpha})=A\vec{\alpha}+b=\vec{\alpha'}.$$

Where

$$T\begin{pmatrix}\alpha_{ij}\\ \alpha_{jk}\\ \alpha_{kl}\end{pmatrix}=\begin{pmatrix}0&1&0\\0&0&1\\-1&-1&-1\end{pmatrix}\begin{pmatrix}\alpha_{ij}\\ \alpha_{jk}\\ \alpha_{kl}\end{pmatrix}+\begin{pmatrix}0\\0\\3\pi\end{pmatrix}=\begin{pmatrix}\alpha_{jk}\\ \alpha_{kl}\\ \alpha_{li}\end{pmatrix}.$$

Then $T^4(\vec{\alpha})=id$. The set of fixed points of T is $\{(\frac{3\pi}{4},\frac{3\pi}{4},\frac{3\pi}{4})\}$ which is the set containing the point corresponding to a linkage which makes a regular octagon. This point is the midpoint of P_0. The transformation S can be applied to the

coordinates of P_0 so as to make the point for the regular octagon be at the origin.

$$S\begin{pmatrix} \alpha_{ij} \\ \alpha_{jk} \\ \alpha_{kl} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \alpha_{ij} \\ \alpha_{jk} \\ \alpha_{kl} \end{pmatrix} - \begin{pmatrix} \frac{3\pi}{4} \\ \frac{3\pi}{4} \\ \frac{3\pi}{4} \end{pmatrix} = \begin{pmatrix} \beta_{ij} \\ \beta_{jk} \\ \beta_{kl} \end{pmatrix}.$$

The transformation S can be used to change from the α-coordinate system of P_0 to a β-coordinate system for the linkages of type $(3,3,3,3)$ and create the space P_{0_β}. The linear transformation $R = S \circ T \circ S^{-1}$ can be applied to the β-coordinates of P_0 to generate the set of points $\{(\beta_{12}, \beta_{23}, \beta_{34}), (\beta_{23}, \beta_{34}, \beta_{41}), (\beta_{34}, \beta_{41}, \beta_{12}), (\beta_{41}, \beta_{12}, \beta_{23})\}$ for each linkage. As with T, $R^4 = id$.

Proposition 2.2 *P_0 can be split into four quadrants where each quadrant will contain at least one of the coordinate representatives from the set of points representing a linkage $L \in P_0$. One of these quadrants will form the fundamental domain, P_1, of P_0.*

Proof Begin by dividing the boundary of P_0 into four pieces. Looking at the faces of P_0, it can be seen that:

$$T(F_{1_0}) = F_{2_0}, T^2(F_{1_0}) = F_{3_0}, T^3(F_{1_0}) = F_{4_0};$$

$$T(F_{5_0}) = F_{6_0}, T^2(F_{5_0}) = F_{7_0}, T^3(F_{5_0}) = F_{8_0};$$

$$T(F_{9_0}) = F_{10_0}, T^2(F_{9_0}) = F_{11_0}, T^3(F_{9_0}) = F_{12_0}.$$

The boundary of P_0 can then be divided into the four sets given by $\{F_{i_0}, F_{j_0}, F_{k_0}\}$, for $i \in (1,4), j \in (5,8), k \in (9,12)$, based on which faces are connected to one another.

Consider the set $\{F_{1_0}, F_{5_0}, F_{9_0}\}$. Connect the vertices V_{1_0}, V_{2_0}, V_{4_0}, V_{5_0}, V_{8_0}, $V_{13_0}, V_{16_0}, V_{17_0} V_{18_0}$ of this set to the midpoint of P_0. This will define the boundary of the space $P_1 \in P_0$.

For coordinate simplicity, apply R to the linkages in P_1 to get the space P_{1_β}. Then the edges connecting the vertices (V_{1_β}, V_{2_β}, V_{4_β}, $V_{5_\beta}, V_{8_\beta}, V_{13_\beta}$, V_{16_β}, V_{17_β}, V_{18_β}) to the origin can be described by the vertex. That is, every linkage on the edge between the origin and a vertex $V \in P_{1_\beta}$ is of the form tV where $t \in (0,1)$. Since R is linear, $R(tV) = tR(V)$. Thus tV is represented by a set of four points in P_{0_β} and those four points will appear on the edges connecting the origin to the image of V under R^n.

Every point in P_{1_β} is equivalent to $t\rho$ where ρ is a point on the boundary of P_{1_β} and so ρ is on one of the faces $F_{1_\beta}, F_{5_\beta}, F_{9_\beta}$. Since the faces are sent to other faces not in P_{1_β} under the transformations R, R^2, R^3; for all $t\rho \in P_{1_\beta}$ there exists a point in $R(P_{1_\beta}), R^2(P_{1_\beta}), R^3(P_{1_\beta})$.

Applying R^{-1} to P_{1_β} shows that for all $\rho \in P_1$ there exists a point in $T(P_1), T^2(P_1), T^3(P_1)$ and so P_1 is the fundamental domain of P_0 and the desired configuration space for linkages of type $(3,3,3,3)$.

□

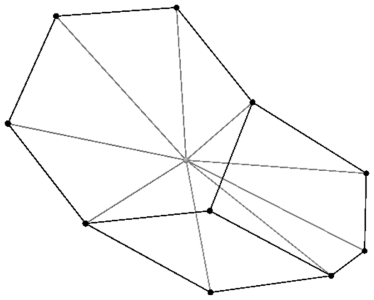


Figure 2.2.3: P_1.

2.3 Linkage Type: $(3, 2, 2, 1)$

Let linkages of type $(3, 2, 2, 1)$ with a word representation of 12341423, with the configuration space P_2, be plotted using a $(\alpha_{12}, \alpha_{23}, \alpha_{34})$-coordinate system. Table 2 shows what the faces, edges, and vertices of P_2 will be. It can be seen that there will be 8 faces, 18 edges, and 12 vertices.

The relations between the faces, edges, and vertices mentioned in Table 2 can be used to determine that the faces F_{1_2} and F_{2_2} will be hexagonal, and the remaining faces F_{3_2} through F_{8_2} will be quadrilateral.

Linkages of type $(3, 2, 2, 1)$ do not have the same amount of symmetry as those of type $(3, 3, 3, 3)$ do, so every point in P_2 corresponds to a unique linkage and there is no fundamental domain for the space which needs to be found.

Vertex	Coordinates	Vertex	Coordinates
V_{1_2}	$(\frac{\pi}{3}, \pi, \pi)$	V_{7_2}	$(\pi, \frac{\pi}{3}, \frac{4\pi}{3})$
V_{2_2}	$(\frac{\pi}{3}, \pi, \frac{4\pi}{3})$	V_{8_2}	$(\frac{2\pi}{3}, \frac{\pi}{3}, \frac{5\pi}{3})$
V_{3_2}	$(\frac{2\pi}{3}, \pi, \frac{2\pi}{3})$	V_{9_2}	$(\frac{\pi}{3}, \frac{2\pi}{3}, \frac{4\pi}{3})$
V_{4_2}	$(\frac{2\pi}{3}, \pi, \pi)$	V_{10_2}	$(\pi, \frac{2\pi}{3}, \frac{2\pi}{3})$
V_{5_2}	$(\frac{2\pi}{3}, \frac{\pi}{3}, \frac{4\pi}{3})$	V_{11_2}	$(\frac{\pi}{3}, \frac{2\pi}{3}, \frac{5\pi}{3})$
V_{6_2}	$(\pi, \frac{\pi}{3}, \pi)$	V_{12_2}	$(\pi, \frac{2\pi}{3}, \pi)$

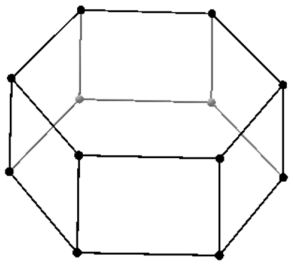


Figure 2.3.1: Coordinates for the vertices of P_2 and a plot of the vertices and edges.

As with P_0, the equations for the faces of P_2 can be determined to be:

$F_{1_2} : \alpha_{12} + \alpha_{23} + \alpha_{34} = \frac{7\pi}{3}$
$F_{2_2} : \alpha_{12} + \alpha_{23} + \alpha_{34} = \frac{8\pi}{3}$
$F_{3_2} : \alpha_{23} = \pi$
$F_{4_2} : \alpha_{12} + \alpha_{23} = \pi$
$F_{5_2} : \alpha_{12} = \pi$
$F_{6_2} : \alpha_{23} = \frac{\pi}{3}$
$F_{7_2} : \alpha_{12} = \frac{\pi}{3}$
$F_{8_2} : \alpha_{12} + \alpha_{23} = \frac{5\pi}{3}$

This indicates that all of the faces are planar and P_2 will be convex if the interior is well-behaved.

Proposition 2.1 *All the points $p \in \mathbb{R}^3$ satisfying the succeeding inequalities will correspond to a linkage L of type $(3, 2, 2, 1)$ which satisfies the requirements outlined in Section 1. These inequalities describe the subset of $\mathbb{R}^3$ which is bound by the faces of P_2.*

(1) $\frac{7\pi}{3} \leq \alpha_{12} + \alpha_{23} + \alpha_{34} \leq \frac{8\pi}{3}$
(2) $\pi \leq \alpha_{12} + \alpha_{23} \leq \frac{5\pi}{3}$
(3) $\frac{\pi}{3} \leq \alpha_{12} \leq \pi$
(4) $\frac{\pi}{3} \leq \alpha_{23} \leq \pi$

Proof

A linkage L of type $(3, 2, 2, 1)$ will be closed and have an interior sum of 6π with $\alpha_{12} + \alpha_{23} + \alpha_{34} + \alpha 41 = 3\pi$. Due to its structure, an additional vector, $\vec{v}_5$, can be added to the linkage which will be parallel to $\vec{v}_1$ and create the six-bar linkage represented 123523 and the four-bar linkage represented by 4145. Then by symmetry, $\alpha_{12} + \alpha_{23} + \alpha_{35} = 2\pi$ and $\alpha_{54} + \alpha 41 = \pi$, where $\alpha_{34} = \alpha_{35} + \alpha_{54}$.

By Proposition 1.1, $\alpha_{ij} \geq \frac{\pi}{3}$ for all i, j. Then $\alpha_{41} = \pi - \alpha_{54} \leq \pi - \frac{\pi}{3} = \frac{2\pi}{3}$. Subtracting α_{41} from $\alpha_{12}+\alpha_{23}+\alpha_{34}+\alpha 41 = 3\pi$ yields $\frac{7\pi}{3} \leq \alpha_{12}+\alpha_{23}+\alpha_{34} \leq \frac{8\pi}{3}$, so (1) is satisfied. Similarly, $\alpha_{12} + \alpha_{23} = 2\pi - \alpha_{35} \leq 2\pi - \frac{\pi}{3} = \frac{5\pi}{3}$.

Proposition 1.1 also gives that $\alpha_{ij} + \alpha_{jk} \geq \pi$ for all i, j, k. Combining this with $\alpha_{12} + \alpha_{23} \leq \frac{5\pi}{3}$ yields (2). We know that $\alpha_{12} + \alpha_{35} \geq \pi$, so $\alpha_{23} = 2\pi - (\alpha_{12} + \alpha_{35}) \leq \pi$. This satisfies (4) since $\alpha_{ij} \geq \frac{\pi}{3}$ for all i, j. Switching α_{12} and α_{23} gives (3). □

2.4 Linkage Type: $(1, 2, 2, 1)$

Let linkages of type $(1, 2, 2, 1)$ with a word representation of 12132434, with the configuration space P_3, be plotted using a $(\alpha_{12}, \alpha_{32}, \alpha_{34})$-coordinate system. Table 3 shows what the faces, edges, and vertices of P_3 will be. It can be seen that there will be 6 faces, 12 edges, and 8 vertices.

The relations between the faces, edges, and vertices mentioned in Table 3 can be used to determine that all of the faces will be quadrilateral.

Linkages of type $(1, 2, 2, 1)$ do not have the same amount of symmetry as those of type $(3, 3, 3, 3)$ do, so every point in P_3 corresponds to a unique linkage and there is no fundamental domain for the space which needs to be found.

As with the previous spaces, the equations for the faces of P_2 can be determined to be:

Vertex	Coordinates	Vertex	Coordinates
V_{1_3}	$(\frac{2\pi}{3}, \frac{\pi}{3}, \frac{2\pi}{3})$	V_{5_3}	$(\frac{2\pi}{3}, \frac{2\pi}{3}, \frac{\pi}{3})$
V_{2_3}	$(\frac{2\pi}{3}, \frac{\pi}{3}, \frac{\pi}{3})$	V_{6_3}	$(\frac{2\pi}{3}, \frac{2\pi}{3}, \frac{2\pi}{3})$
V_{3_3}	$(\frac{\pi}{3}, \frac{\pi}{3}, \frac{\pi}{3})$	V_{7_3}	$(\frac{\pi}{3}, \frac{2\pi}{3}, \frac{\pi}{3})$
V_{4_3}	$(\frac{\pi}{3}, \frac{\pi}{3}, \frac{2\pi}{3})$	V_{8_3}	$(\frac{\pi}{3}, \frac{2\pi}{3}, \frac{2\pi}{3})$

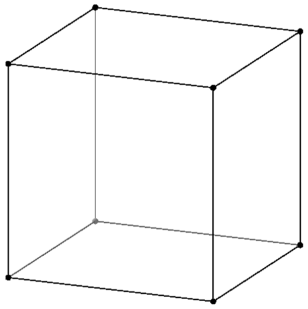


Figure 2.4.1: Coordinates for the vertices of P_3 and a plot of the vertices and edges.

$F_{1_3} : \alpha_{32} = \frac{\pi}{3}$ $\quad$ $F_{4_3} : \alpha_{34} = \frac{2\pi}{3}$
$F_{2_3} : \alpha_{12} = \frac{2\pi}{3}$ $\quad$ $F_{5_3} : \alpha_{34} = \frac{\pi}{3}$
$F_{3_3} : \alpha_{12} = \frac{\pi}{3}$ $\quad$ $F_{6_3} : \alpha_{32} = \frac{2\pi}{3}$

This indicates that all of the faces are planar and P_2 will be convex if the interior is well-behaved.

Proposition 2.1 *All the points $p \in \mathbb{R}^3$ satisfying the succeeding inequalities will correspond to a linkage L of type $(1, 2, 2, 1)$ which satisfies the requirements outlined in Section 1. These inequalities describe the subset of $\mathbb{R}^3$ which is bound by the faces of P_3.*

(1) $\frac{\pi}{3} \leq \alpha_{12} \leq \frac{2\pi}{3}$ $\qquad$ (2) $\frac{\pi}{3} \leq \alpha_{32} \leq \frac{2\pi}{3}$ $\qquad$ (3) $\frac{\pi}{3} \leq \alpha_{34} \leq \frac{2\pi}{3}$

Proof

A linkage L of type $(1, 2, 21)$ will be closed with a symmetric structure that causes $\alpha_{12} + \alpha_{21} = \pi$ and $\alpha_{43} + \alpha_{34} = \pi$. By Proposition 1.1, $\alpha_{ij} \geq \frac{\pi}{3}$ for all i, j. Combining this fact with the above properties yields $\frac{\pi}{3} \leq \alpha_{12} \leq \frac{2\pi}{3}$ (1) and $\frac{\pi}{3} \leq \alpha_{34} \leq \frac{2\pi}{3}$ (3).

The structure of L allows for an additional vector $\vec{v}_5$ to be added which will be parallel to $\vec{v}_2$. A vector $\vec{v}_6$ parallel to $\vec{v}3$ can also be added to L. This will form three quadrilaterals represented by 1215, 3265, and 4346. This means that $\alpha_{32} + \alpha_{26} = \pi$, so (2) is satisfied by subtracting $\alpha_{26} \geq \frac{\pi}{3}$.

□

3 Equivalence Relation on the Linkages

Each linkage which is in either P_1, P_2, or P_3 corresponds to the minimal vector decomposition of a well-rounded translation surface of genus 2. Let K be the polygonal region that has edges defined by the linkage $L \in \bigcup_{n=1}^{3} P_n$. Then the well-rounded translation surface corresponding to L is obtained by pasting the edges of K together. The word $(12341234, 12341423, 12132434)$ used to describe L can be replaced by a labeling scheme $(klmnk^{-1}l^{-1}m^{-1}n^{-1}$, $klmnk^{-1}n^{-1}l^{-1}m^{-1}$, $klk^{-1}ml^{-1}nm^{-1}n^{-1})$ for K. The exponent of each edge indicates the orientation of the edge. When pasting, edges k and k^{-1} may be pasted together, but not edges k and k. Figure 3.0.2 shows the pasting process for a polygonal region with labeling scheme $klk^{-1}ml^{-1}nm^{-1}n^{-1}$ corresponding to a linkage of type $(1, 2, 2, 1)$.

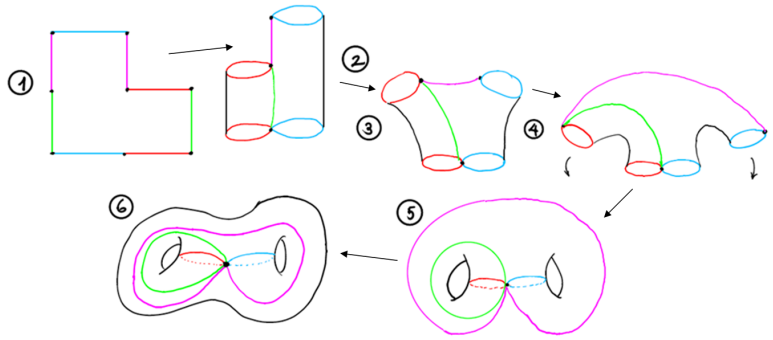


Figure 3.0.2: This diagram shows how pasting a polygonal region corresponding to a linkage of type $(1, 2, 2, 1)$ will result in a genus 2 surface. (Image by Christopher Judge.)

Before all of the edges of K are pasted together, any interior line connecting two vertices can be cut along to create two polygonal regions K_1 and K_2. These two regions can then be pasted together along any pair of edges with the same label, as in Figure 3.0.3, to make the polygonal region K'. Such a cutting and pasting will change the geometry of K but will not change the translation surface obtained by pasting together all of the edges of K [7]. That is, the topological space obtained from pasting together all of the edges of K will be the same as that obtained from K' and so the two regions are said to be equivalent.

Proposition 3.1 *Given a linkage $L \in \bigcup_{n=1}^{3} P_n$, there exists a polygonal region K with edges defined by L. K can be cut along any interior line segment joining two vertices ℓ and pasted together to form the polygonal region K' with a different labeling scheme than that of K. The edges of K' can be described by a linkage L'. Then $L' \in \bigcup_{n=1}^{3} P_n$ if and only if $|\ell| = 1$.*

Proof

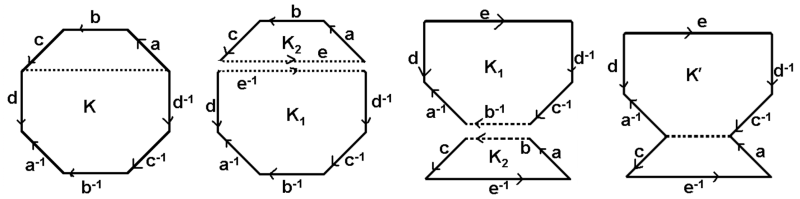


Figure 3.0.3: This diagram shows how a polygonal region K with a labeling scheme of $abcda^{-1}b^{-1}c^{-1}d^{-1}$ can be cut and then pasted back together at a different pair of edges to form the region K' with the labeling scheme $ac^{-1}d^{-1}eda^{-1}ce^{-1}$.

Let $L' \in \bigcup_{n=1}^{3} P_n$. The edges of L' will all be of length one since L' is created by eight unit vectors. L' is the linkage describing the edges of the polygonal region K' which was obtained by cutting-and-pasting the region K, described by L. Since K was cut along the line ℓ and has a different labeling scheme than K', ℓ must appear in the labeling scheme of K'. So K' has an edge labeled ℓ which corresponds to a vector in L' and so ℓ must be of length one.

Suppose that ℓ is the line segment in K which was cut along such that K' is obtained. Thus when K was cut apart along ℓ, the resulting two polygonal regions had unit length sides since the edges of K can be described by $L \in \bigcup_{n=1}^{3} P_n$. When these two regions are pasted together to form K', the edges which comprise K' will be of unit length since cutting-and-pasting has no effect on the length of existing edges and so L' will consist of unit vectors. Furthermore, the two regions obtained from cutting K can be described by linkages with an interior distance between distinct vertices greater than or equal to one since $L \in \bigcup_{n=1}^{3} P_n$. Pasting these two regions together will preserve this property, so L' will have interior distances between distinct vertices which are no less than one. The action of cutting-and-pasting is geometrical and will not affect the topological space, so K' will also paste together to be a genus 2 surface. This means that L' is the minimal vector decomposition of a well-rounded translation surface of genus 2, so $L' \in \bigcup_{n=1}^{3} P_n$.

□

By Proposition 3.1, there exists a mapping via cutting-and-pasting which sends any linkage $L \in \bigcup_{n=1}^{3} P_n$ to a linkage $L' \neq L$ also in $\bigcup_{n=1}^{3} P_n$, shown in Figure 3.0.4. The same notion of equivalence between two polygonal regions K and K' which paste together to be the same topological space, can be extended to the linkages which describe K and K'.

Proposition 3.2 *Given the linkages $L, L' \in \bigcup_{n=1}^{3} P_n$ describing the polygonal regions K, K'. The cutting-and-pasting techniques described above can be regarded as an equivalence relation, where $L \sim L'$ if K and K' are equivalent.*

Proof Let the polygonal regions K, K', K'' described by the linkages L, L', L''

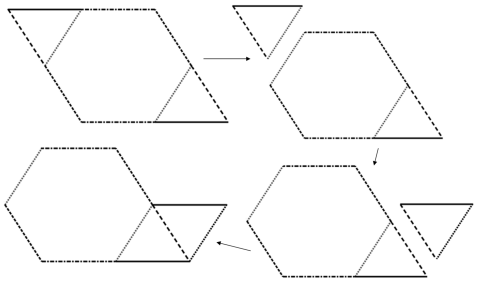


Figure 3.0.4: This is an example of how to get a linkage on $F_{1_2} \in P_2$ from a linkage on $F_{1_0} \in P_1$.

be equivalent.

Reflexive: $L \sim L \ \forall \ L \in \bigcup_{n=1}^{3} P_n$.
K is equivalent to K since K can be cut along some line and then pasted together upon that same line. It follows that L will describe K, so $L \sim L$. Symmetric: $L \sim L'$ implies $L' \sim L$ for all $L, L' \in \bigcup_{n=1}^{3} P_n$.
$L \sim L'$ implies K is equivalent K'. Then K' is equivalent to K, so $L' \sim L$

Transitive: $L \sim L'$ and $L' \sim L''$ implies $L \sim L''$ for all $L, L', L'' \in \bigcup_{n=1}^{3} P_n$.
$L \sim L'$ implies K is equivalent K' and $L' \sim L''$ implies K' is equivalent to K''. Then there exists a cutting-and-pasting of K to K' and then to K'', so K is equivalent to K''. Thus $L \sim L''$.

□

3.1 Structure of the CW-Complex

n-cell	P
0	3
1	8
2	10
3	3

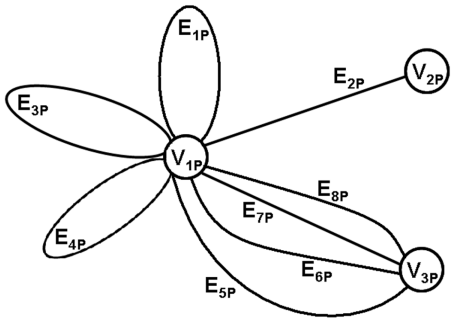


Figure 3.1.1: This figure shows the number of n-cells the CW-complex P has for each n. The 1-skeleton for P is also given.

To understand the space that is $P = \bigcup_{n=1}^{3} P_n$ with the equivalence relation given in Proposition 3.2, consider the equivalence classes for the linkages on the boundary of P. These are given in Figure 3.1.2. By the equivalence relation of

Proposition 3.2, it appears that P consists of 2 verticies, 8 edges, and 10 faces coming from 3 polyhedron. However, the linkages which form edges in P in the equivalence class E_{2_P} are equivalent to themselves except for the linkage which is the midpoint of the edge. That is if L is a linkage on some edge E in the equivalence class E_{2_P}, then there exists a linkage L' also on E with $L \sim L'$. There is one linkage W on E such that if $W \sim W'$ for W' on E, then W' is a relabeling of W and so W and W' are the same linkage, up to a relabeling. It turns out that W will be the midpoint of E in the configuration space and can be regarded as a third vertex of P.

It follows that P is a CW-complex with the n-cell breakdown as given in Figure 3.1.1. An n-cell is defined as being homeomorphic to a n-dimensional ball. The n-skeleton is the union of all of the m-cells for $m \leq n$. The 1-skeleton of P is given in Figure 3.1.1.

Equiv. Class	Memebers
F_{1_P}	F_{5_0}
F_{2_P}	$F_{1_0}, F_{1_2}, F_{2_2}$
F_{3_P}	F_{9_0}, F_{3_2}, F_{4_2}, F_{5_2}
F_{4_P}	F_{6_2}, F_{2_3}, F_{3_3}, F_{4_3}, F_{5_3}
F_{5_P}	F_{7_2}, F_{8_2}, F_{1_3}, F_{6_3}
F_{6_P}	faces in P_1 formed by: $\{V_{8_0}, V_{18_0}, V_{3_P}\}$, $\{V_{5_0}, V_{17_0}, V_{3_P}\}$
F_{7_P}	faces in P_1 formed by: $\{V_{1_0}, V_{16_0}, V_{3_P}\}$, $\{V_{4_0}, V_{13_0}, V_{3_P}\}$
F_{8_P}	faces in P_1 formed by: $\{V_{1_0}, V_{18_0}, V_{3_P}\}$, $\{V_{2_0}, V_{17_0}, V_{3_P}\}$
F_{9_P}	faces in P_1 formed by: $\{V_{8_0}, V_{13_0}, V_{3_P}\}$, $\{V_{5_0}, V_{16_0}, V_{3_P}\}$
F_{10_P}	faces in P_1 formed by: $\{V_{2_0}, V_{4_0}, V_{3_P}\}$
E_{1_P}	$E_{1_0} - E_{8_0}$, $E_{11_0} - E_{14_0}$, $E_{23_0} - E_{30_0}$, E_{3_2}, E_{4_2}, $E_{7_2} - E_{10_2}$
E_{2_P}	E_{9_0}, E_{10_0}, E_{5_2}, E_{6_2}, $E_{5_3} - E_{8_3}$
E_{3_P}	$E_{15_0} - E_{18_0}$, E_{2_2}, E_{11_2}, E_{14_2}, E_{15_2}, E_{16_2}, E_{1_3}, E_{2_3}, E_{10_3}, E_{12_3}
E_{4_P}	$E_{19_0} - E_{22_0}$, E_{1_2}, E_{12_2}, E_{13_2}, E_{17_2}, E_{18_2}, E_{3_3}, E_{4_3}, E_{9_3}, E_{11_3}
E_{5_P}	edges in P_1 between: $\{V_{1_0}, V_{3_P}\}$, $\{V_{2_0}, V_{3_P}\}$, $\{V_{4_0}, V_{3_P}\}$
E_{6_P}	edges in P_1 between: $\{V_{13_0}, V_{3_P}\}$, $\{V_{16_0}, V_{3_P}\}$
E_{7_P}	edges in P_1 between: $\{V_{17_0}, V_{3_P}\}$, $\{V_{18_0}, V_{3_P}\}$
E_{8_P}	edges in P_1 between: $\{V_{5_0}, V_{3_P}\}$, $\{V_{8_0}, V_{3_P}\}$
V_{1_P}	$V_{1_0} - V_{20_0}$, $V_{1_2} - V_{12_2}$, $V_{1_3} - V_{8_3}$
V_{2_P}	midpoint of edges in the equivalence class E_{2_P}
V_{3_P}	point $(\frac{3\pi}{4}, \frac{3\pi}{4}, \frac{3\pi}{4}) \in P_1$= regular octagon

Figure 3.1.2: These are the equivalence classes for the linkages on the boundary of P given for each n-cell ($n = 0, 1, 2$) and the linkages which belong to them.

4 Properties of the CW-Complex

To understand P beyond how it is constructed, some of its topological properties are calculated. The next three subsections explore the Euler characteristic, homology groups, and fundamental group of P.

4.1 Euler Characteristic

For a finite CW complex X, the Euler characteristic $\chi(X)$ is defined to be the alternating sum $\sum_n (-1)^n c_n$ where c_n is the number of n-cells of X [4]. For a surface S, this is the familiar formula: $\chi(S) = F - E + V = 2 - 2g$, where F is the number of faces, E the number of edges, V the number of vertices, and g the genus of S. For the space P,

$$\chi(P) = \sum_{n=0}^{3} (-1)^n c_n = 3 - 8 + 10 - 3 = 2.$$

A motivation for calculating the Euler characteristic is that it is a topological invariant, meaning that it is preserved under homeomorphism and thus can be used to show that other spaces are not related to P.

4.2 Homology Groups

The homology groups of a space are also topological invariants. The n^{th} (integer) homology group of X is defined to be the abelian group

$$H_n(X) = ker(\partial_n)/Im(\partial_{n+1}),$$

where $\partial_n : C_n(X) \to C_{n-1}(X)$ is a homomorphism on the n-chains of X. The n-chains are the abelian groups whose elements are integer linear combinations of oriented n-cells. It follows that the set of n-cells will generate the n-chains and so $rank(C_n) = c_n$. ∂_n is the boundary operator acting on the n-chains.

The chain complex for P is defined as:

$$(0) \xrightarrow{\partial_4} C_3(X) \xrightarrow{\partial_3} C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \xrightarrow{\partial_0} (0)$$

The matrix representation of ∂_n, for each n, can be found by determining the map of each unique n-cell onto the $(n-1)$-cells which comprise its boundary. The boundary of each n-cell of P was found using the following process:

∂_0: The boundary of a 0-cell is trivial.

∂_1: The boundary of an oriented 1-cell, E_{i_P}, is $V_{j_P} - V_{k_P} \in C_0(P)$, where E_{i_P} is oriented toward V_{j_P} and away from V_{k_P}. The orientation of each 1-cell with respect to P was determined by the cutting-and-pasting mapping discussed in Section 3. For example, for P, $E_{2_P} = V_{2_P} - V_{1_P}$.

∂_2: The boundary of an oriented 2-cell, F_{i_P}, is $a_j E_{j_P} + a_k E_{k_P} + ... + a_m E_{m_P} \in C_1(P)$, where F_{i_P} is oriented in either the clockwise or counter-clockwise direction and the sign of the coefficient of each 1-cell is determined

by whether it is oriented to go with the orientation of F_{i_P} $(+)$ or if it goes against the orientation of F_{i_P} $(-)$. For example, for P, $F_{2_P} = E_{1_P} + E_{4_P} - E_{1_P} - E_{3_P} + E_{1_P} + E_{2_P} - E_{2_P} = E_{1_P} - E_{3_P} + E_{4_P}$

∂_3: The boundary of an oriented 3-cell, P_i, is $b_j F_{j_P} + b_k F_{k_P} + ... + b_m F_{m_P} \in C_2(P)$, where the sign of the coefficient of each 2-cell is determined by the 2-cells orientation. The orientation of each 2-cell can be described by the set $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ of the canonical basis, where $\vec{e}_3$ is the outward facing normal vector of the 2-cell and so $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$ denotes a 2-cell oriented counter-clockwise while $(\vec{e}_2, \vec{e}_1, \vec{e}_3)$ denotes a 2-cell oriented clockwise. Taking the determinant of the matrix representation of $(\vec{e}_i, \vec{e}_j, \vec{e}_3)$ results in 2-cells oriented counter-clockwise having a positive sign and those oriented clockwise have a negative sign. For example, for P, $P_3 = F_{5_P} - F_{5_P} - F_{4_P} + F_{4_P} - F_{4_P} - F_{5_P} = -2F_{5_P}$. Notice that a 2-cell is not required to always have the same orientation with respect to the 3-cell since its orientation is determined by how the 1-cells around it are placed.

For the boundary operators of P, let $M(\partial_n)$ denote the matrix representation of ∂_n. Then:

$$M(\partial_3) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \\ -2 & 0 & 0 \\ 2 & 0 & 0 \\ -2 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad M(\partial_1) = \begin{pmatrix} 0 & -1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \end{pmatrix}$$

$$M(\partial_2) = \begin{pmatrix} 5 & 1 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \end{pmatrix} \quad M(\partial_0) = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$$

Since $H_n(P)$ is an abelian group, we can apply the Fundamental Theorem for Finitely Generated Abelian Groups.

Theorem 4.1 *[6] Let G be a finitely generated abelian group. Let T be its torsion subgroup.*

(a) There is a free abelian subgroup H of G having finite rank β such that

$G = H \oplus T$.

(*b*) *There are finite cyclic groups* $T_1, ..., T_k$ *where* T_i *has order* $t_i > 1$*, such that* $t_i | t_2 | ... | t_k$ *and* $T = T_1 \oplus ... \oplus T_k$.

(*c*) *The numbers* β *and* $t_1, ..., t_k$ *are uniquely determined by* G.

In particular, for each n,

$$H_n(P) \cong \mathbb{Z}^{\beta_n} \oplus \mathbb{Z}/t_{1_n} \oplus ... \oplus \mathbb{Z}/t_{k_n}.$$

The number β_n is called the n^{th} *Betti* number of P. The Smith Normal form of the matrix representation of the boundary operators of P gives all of these numbers. The torsion coefficients of $H_n(P)$ are given by the entries which are greater than one in $M(\partial_{n+1})$. Furthermore, $rank(ker(\partial_{n+1}))$ is the number of zero columns of $M(\partial_{n+1})$. Additionally, the number of non-zero rows of $M(\partial_{n+1})$ is equal to $rank(Im(\partial_{n+1}))$. It follows that

$$\beta_n = rank(ker(\partial_n)) - rank(Im(\partial_{n+1})) = rank(H_n).$$

Let $SNF(\partial_n)$ denote the Smith Normal form of $M(\partial_n)$, then:

$$SNF(\partial_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad SNF(\partial_1) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$SNF(\partial_2) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad SNF(\partial_0) = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$$

Figure 4.2.1 shows the homology groups of P found from the Smith Normal Form of the $M(\partial_n)$.

The homology groups of a space keep track of the number of n-dimensional holes in the space. For example, $rank(H_0(P)) = 1$, so P has one 0-dimensional hole (an open point) and so P is path-connected since it will have a single path.

n	β_n	t_{i_n}	$H_n(P)$
0	3-2=1	-	$\mathbb{Z}$
1	6-6=0	5	$\mathbb{Z}/5\mathbb{Z}$
2	4-3=1	2	$\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$
3	0-0=0	-	$\{0\}$

Figure 4.2.1: This table shows the homology groups and their rank for P.

Similarly, P has no 1-dimensional holes (a circle, S^1) nor 3-dimensional holes (S^3), but P does have a 2-dimensional hole (S^2).

The homology groups are related to the Euler characteristic in that for a finite CW-complex X, $\chi(X) = \sum_n (-1)^n rank(H_n(X))$ [4]. This is just the alternating sum of the Betti numbers for each homology group; so

$$\chi(P) = 1 - 0 + 1 - 0 = 2.$$

This agrees with the calculation of $\chi(P)$ given in Section 4.1.

4.3 The Fundamental Group

The group of loops in a space X starting and ending at a basepoint $x_0 \in X$ is the fundamental group of X, denoted $\pi_1(X)$. Two loops are considered to be the same if one can be deformed into the other within X [4]. The space P is path-connected; allowing for the use of Van Kampen's Theorem to calculate $\pi_1(P)$.

Theorem 4.1 (Van Kampen) *[4] If X is the union of path-connected open sets A_α, A_β each containing the basepoint $x_0 \in X$ and if the intersection $A_\alpha \cap A_\beta$ is path-connected, then the homomorphism $\phi : \pi_1(A_\alpha) * \pi_1(A_\beta) \to \pi_1(X)$ is surjective. If in addition the intersection $A_\alpha \cap A_\beta$ is path-connected, then the kernel of ϕ is the normal subgroup N generated by all elements of the form $i_{\alpha\beta}(w) i_{\beta\alpha}(w)^{-1}$ for $w \in \pi(A_\alpha \cap A_\beta)$, and hence ϕ induces an isomorphism $\pi_1(X) \cong (\pi_1(A_\alpha) * \pi_1(A_\beta))/N$.*

For calculating $\pi_1(P)$, let

$$P = \mathbb{R}^2 \cup \left(\bigcup_{k=1}^{10} U_k \right)$$

where U_k is an open set around the 1-skeleton for each face of P. $\pi_1(\mathbb{R}^n) = \{0\}$ for all n, so $\pi_1(P) = *_k \pi_1(U_k)/ker(\phi)$. (Note: It is sufficient to consider only the 2-skeletons of P. Taking into account the n-skeletons of P for $n > 2$ will add no new information to $\pi_1(P)$ [4].) Let p_0 denote the basepoint for finding $\pi_1(P)$ where p_0 is the vertex in P which corresponds to the vertex that was used to plot the configuration space for each linkage type. The loops around each F_k

will generate $ker(\phi)$ since each will start and end at p_0 and are distinct from one another since the U_k are distinct (due to the faces of P being distinct). The generators of $*_k\pi_1(U_k)$ are the distinct paths taken along the edges of P from p_0 to p_0. Figure 4.3.1 shows each face of of P with their corresponding edges and vertices.

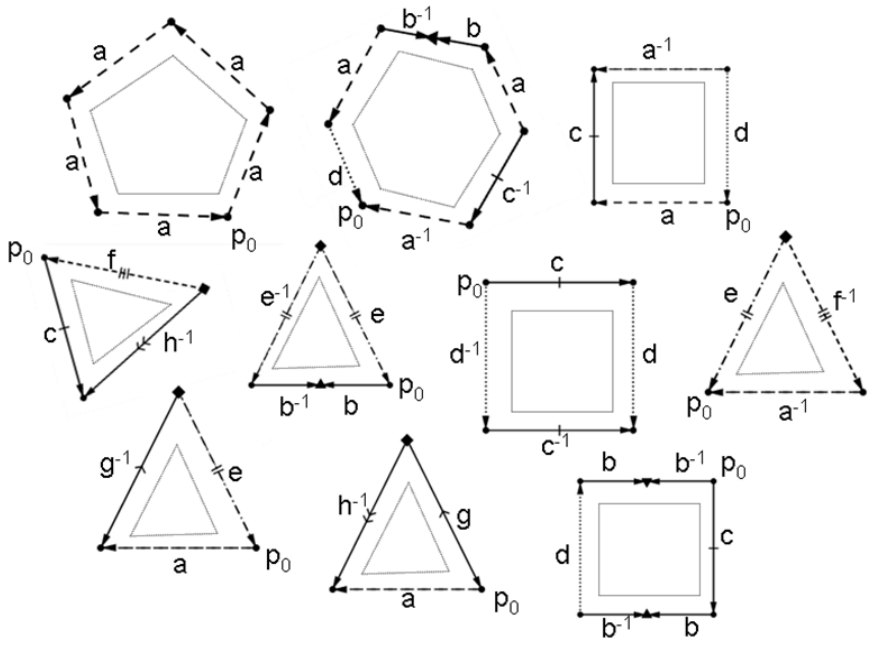


Figure 4.3.1: The labels and orientations used for each edge in P are shown with respect to the faces which they lie on. The orientation of each face is implied by the power of each edge. The interior line indicates the boundary of the open set around the 1-skeleton, U_k, defining each face. For example, the 2-cell which has five edges (F_{1_P}) is generated by edge a and $aaaaa \in ker(\phi)$. The basepoint p_0 is also indicated.

Calculating $\pi_1(P)$ yields:

$$\pi_1(P) = \langle a, bb^{-1}, c, d, h^{-1}g, f^{-1}e, g^{-1}e, h^{-1}f \rangle /$$
$$\langle a^5, a^{-1}c^{-1}abb^{-1}ad, aca^{-1}d, cbb^{-1}dbb^{-1}, cdc^{-1}d^{-1}, ah^{-1}g, a^{-1}f^{-1}e, ag^{-1}e, ch^{-1}f, bb^{-1}e^{-1}e \rangle.$$

For notation simplicity, a, b, c, d, e, f, g, h was used to replace E_{1_P}, E_{2_P}, E_{3_P}, E_{4_P}, E_{5_P}, E_{6_P}, E_{7_P}, E_{8_P}, respectively.

This can be reduced slightly since the loop $bb^{-1}e^{-1}e$ is trivial, as well as bb^{-1}. Additionally, the generator $h^{-1}f$ can be generated by three other generators of $*_k\pi_1(F_k)$ and thus is unnecessary. Then:

$$\pi_1(P) = \langle a, c, d, h^{-1}g, f^{-1}e, g^{-1}e \rangle /$$

$$\langle a^5, a^{-1}c^{-1}a^2d, aca^{-1}d, cd, cdc^{-1}d^{-1}, ah^{-1}g, a^{-1}f^{-1}e, ag^{-1}e, ch^{-1}f\rangle.$$

To simplify the notation of the generators, let $h^{-1}g = w$, $f^{-1}e = u$, and $g^{-1}e = v$, yielding:

$$\pi_1(P) = \langle a, c, d, w, u, v\rangle/$$
$$\langle a^5, a^{-1}c^{-1}a^2d, aca^{-1}d, cd, cdc^{-1}d^{-1}, aw, a^{-1}u, av, cwvu^{-1}\rangle.$$

The fundamental group is related to the first homology group by the following theorem.

Theorem 4.2 *[4] By regarding loops as singular 1 cycles, we obtain a homomorphism $h : \pi_1(X, x_0) \to H_1(X)$. If X is path-connected, then h is surjective and has kernel the commutator subgroup of $\pi_1(X)$, so h induces an isomorphism from the abelianization of $\pi_1(X)$ onto $H_1(X)$.*

Calculating the abelianization of $\pi_1(P)$ does indeed yield an outcome isomorphic to $H_1(P)$.

$$Abel(\pi_1(P)) = \pi_1(P)/[\pi_1(P), \pi_1(P)]$$
$$= \langle a, c, d, w, u, v\rangle/\langle a^5, ac^{-1}d, cd, aw, a^{-1}u, av, cwvu^{-1}\rangle.$$

Since $ker(\phi) = \{a^5, ac^{-1}d, cd, aw, a^{-1}u, av, cwvu^{-1}\}$, it follows that each element of the set is equal to the identity element and so $a = w^{-1} = u = v^{-1}$ and $c = d^{-1}$. This leads to $c = uv^{-1}w^{-1} = a^3$ and $c = ad = ac^{-1}$ so $c^2 = a$. Then $c^5 = id$. $Abel(\pi_1(P))$ can then be reduced to:

$$Abel(\pi_1(P)) = \langle a\rangle/\langle a^5\rangle \cong \mathbb{Z}/5\mathbb{Z} \cong H_1(P).$$

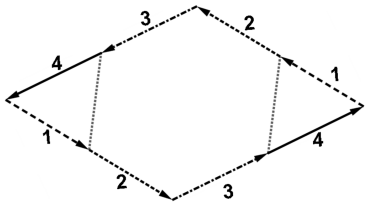
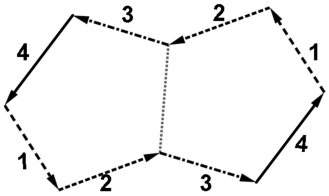
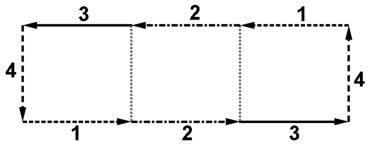
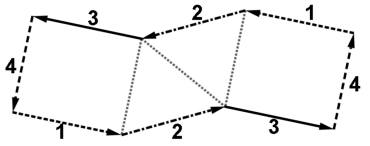
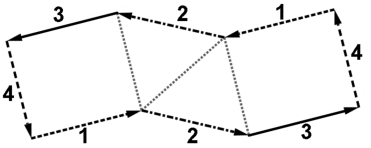
Acknowledgments

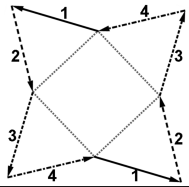
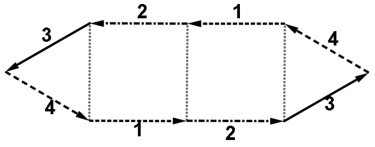
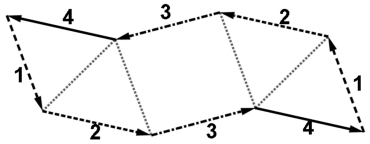
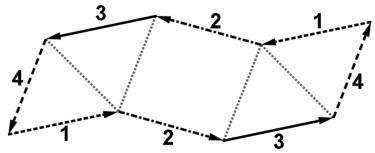
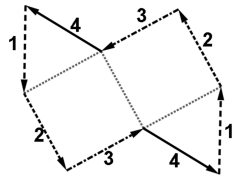
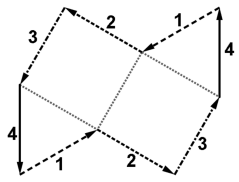
It is a pleasure to thank my fellow researcher Kathryn Marsh of Purdue University with whom I collaborated on this project. I would also like to thank my mentor, Prof. Christopher Judge of Indiana University, the NSF for their funding and the REU program at Indiana University.

Tables

The following are the linkages of type $(3, 3, 3, 3), (3, 2, 2, 1), (1, 2, 2, 1)$ which can be found on the boundaries of P_0, P_2, and P_3, respectively. The dashed lines within the linkages indicate an interior distance of one between vertices.

$(3, 3, 3, 3)$		
Linkage	Name	Formation/Properties

$(3,3,3,3)$		
Faces		
	F_{1_0}	Vectors $\vec{v_1}$ and $\vec{v_4}$ are brought together such that $\alpha_{41} = \frac{\pi}{3}$. F_{i_0} $(i = 2,3,4)$ can be obtained by doing a cyclic permutation of the labels of F_{1_0} which sends 1 to i.
	F_{5_0}	The vertex between vectors $\vec{v_2}$ and $\vec{v_3}$ are brought together. F_{i_0} $(i = 6,7,8)$ can be obtained by doing a cyclic permutation of the labels of F_{5_0} which sends 1 to 2, 3, or 4.
	F_{9_0}	Vector $\vec{v_2}$ is brought close to its parallel pair. F_{i_0} $(i = 10,11,12)$ can be obtained by doing a cyclic permutation of the labels of F_{9_0} which sends 1 to 2, 3, or 4.
Edges		
	E_{1_0}	The edge between faces F_{5_0} and F_{9_0}. E_{i_0} $(i = 2,3,4)$ can be obtained by doing a cyclic permutation of the labels of E_{1_0} which sends 1 to 2, 3, or 4.
	E_{5_0}	The edge between faces F_{8_0} and F_{9_0}. E_{i_0} $(i = 6,7,8)$ can be obtained by doing a cyclic permutation of the labels of E_{5_0} which sends 1 to 2, 3, or 4.

$(3,3,3,3)$		
	E_{9_0}	The edge between faces F_{2_0} and F_{4_0}. E_{10_0} can be obtained by doing a cyclic permutation of the labels of E_{9_0} which sends 1 to 2.
	E_{11_0}	The edge between faces F_{4_0} and F_{8_0}. E_{i_0} $(i = 12, 13, 14)$ can be obtained by doing a cyclic permutation of the labels of E_{11_0} which sends 1 to 2, 3, or 4.
	E_{15_0}	The edge between faces F_{1_0} and F_{10_0}. E_{i_0} $(i = 16, 17, 18)$ can be obtained by doing a cyclic permutation of the labels of E_{15_0} which sends 1 to 2, 3, or 4.
	E_{19_0}	The edge between faces F_{1_0} and F_{9_0}. E_{i_0} $(i = 20, 21, 22)$ can be obtained by doing a cyclic permutation of the labels of E_{19_0} which sends 1 to 2, 3, or 4.
	E_{23_0}	The edge between faces F_{1_0} and F_{6_0}. E_{i_0} $(i = 24, 25, 26)$ can be obtained by doing a cyclic permutation of the labels of E_{23_0} which sends 1 to 2, 3, or 4.
	E_{27_0}	The edge between faces F_{1_0} and F_{8_0}. E_{i_0} $(i = 28, 29, 30)$ can be obtained by doing a cyclic permutation of the labels of E_{27_0} which sends 1 to 2, 3, or 4.
Verticies		

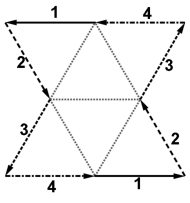
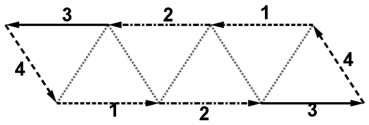
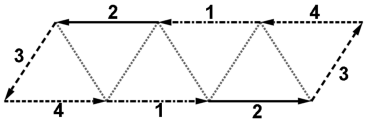
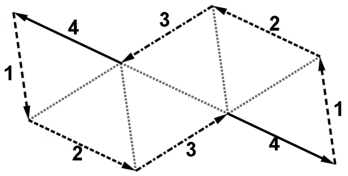
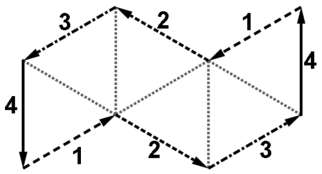
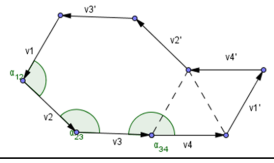
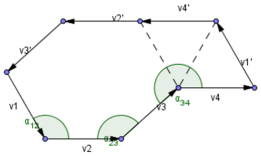
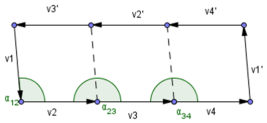
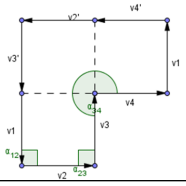
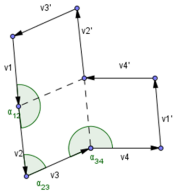
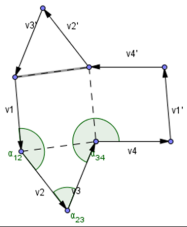
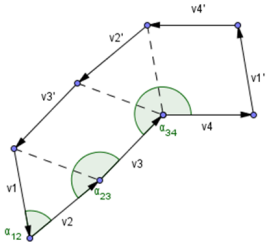
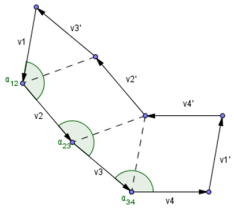
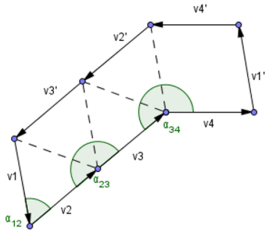
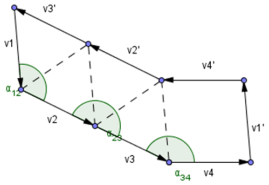
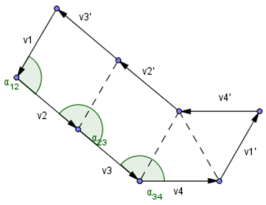
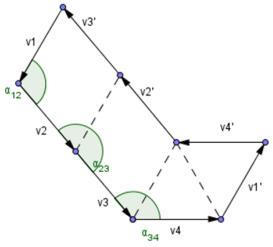
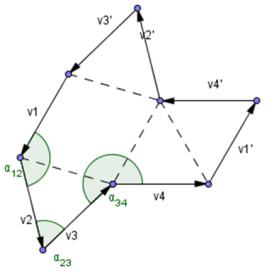
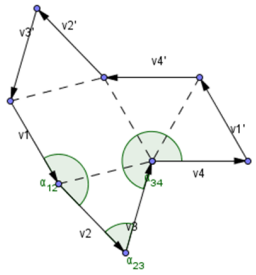
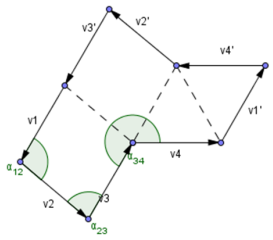
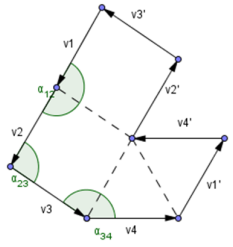
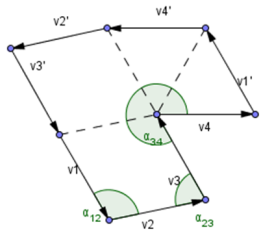
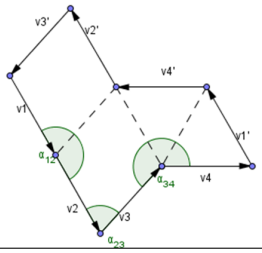
$(3,3,3,3)$		
	V_{1_0}	Vertex between edges E_{9_0}, E_{26_0}, and E_{28_0}. V_{i_0} $(i = 2,3,4)$ can be obtained by doing a cyclic permutation of the labels of V_{1_0} which sends 1 to 2, 3, or 4.
	V_{5_0}	Vertex between edges E_{5_0}, E_{11_0}, and E_{18_0}. V_{i_0} $(i = 6,7,8)$ can be obtained by doing a cyclic permutation of the labels of V_{5_0} which sends 1 to 2, 3, or 4.
	V_{9_0}	Vertex between edges E_{2_0}, E_{11_0}, and E_{22_0}. V_{i_0} $(i = 10,11,12)$ can be obtained by doing a cyclic permutation of the labels of V_{9_0} which sends 1 to 2, 3, or 4.
	V_{13_0}	Vertex between edges E_{4_0}, E_{15_0}, and E_{23_0}. V_{i_0} $(i = 14,15,16)$ can be obtained by doing a cyclic permutation of the labels of V_{13_0} which sends 1 to 2, 3, or 4.
	V_{17_0}	Vertex between edges E_{5_0}, E_{19_0}, and E_{27_0}. V_{i_0} $(i = 18,19,20)$ can be obtained by doing a cyclic permutation of the labels of V_{17_0} which sends 1 to 2, 3, or 4.

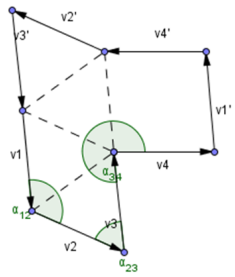
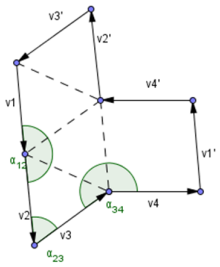
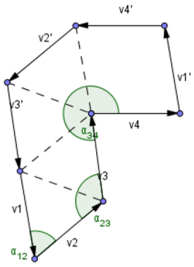
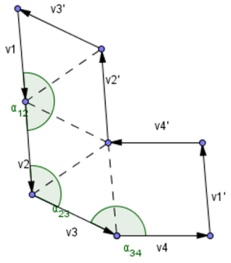
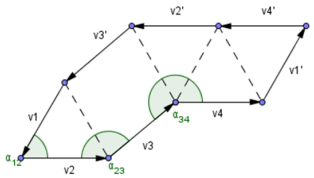
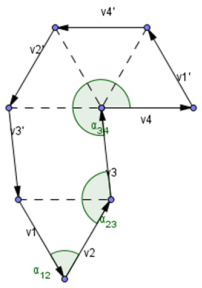
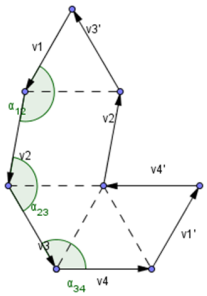
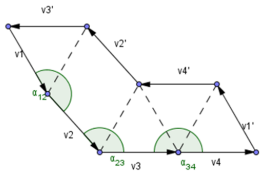
Table 1: This table shows the structure of the linkages of type $(3,3,3,3)$ on the boundary of P_0.

$(3,2,2,1)$			
Linkage	Name/ Connections	Linkage	Name/ Connections

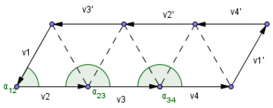
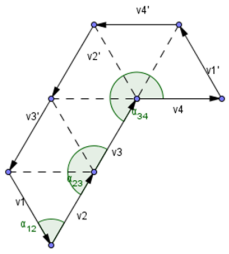
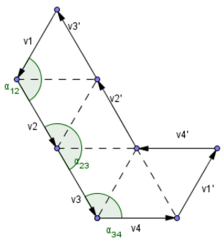
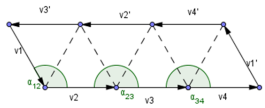
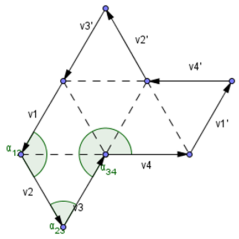
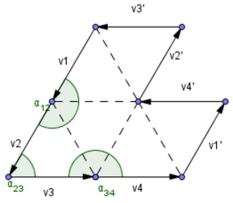
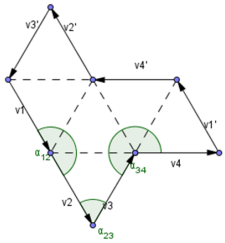
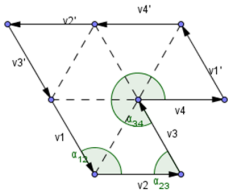
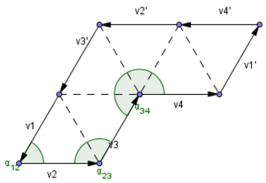
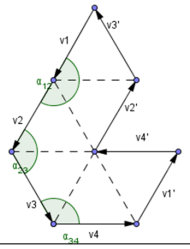
$(3, 2, 2, 1)$

Faces			
	F_{1_2} edges: E_{3_2} E_{5_2} E_{7_2} E_{8_2} E_{15_2} E_{17_2}		F_{2_2} edges: E_{4_2} E_{6_2} E_{9_2} E_{10_2} E_{16_2} E_{18_2}
	F_{3_2} edges: E_{1_2} E_{2_2} E_{3_2} E_{4_2}		F_{4_2} edges: E_{7_2} E_{9_2} E_{11_2} E_{13_2}
	F_{5_2} edges: E_{8_2} E_{10_2} E_{12_2} E_{14_2}		F_{6_2} edges: E_{5_2} E_{6_2} E_{11_2} E_{12_2}
	F_{7_2} edges: E_{1_2} E_{13_2} E_{15_2} E_{16_2}		F_{8_2} edges: E_{2_2} E_{14_2} E_{17_2} E_{18_2}
Edges			
	E_{1_2} vertices: V_{1_2} V_{2_2}		E_{2_2} vertices: V_{3_2} V_{4_2}

$(3,2,2,1)$			
	E_{3_2} vertices: V_{1_2} V_{3_2}		E_{4_2} vertices: V_{2_2} V_{4_2}
	E_{5_2} vertices: V_{5_2} V_{6_2}		E_{6_2} vertices: V_{7_2} V_{8_2}
	E_{7_2} vertices: V_{5_2} V_{9_2}		E_{8_2} vertices: V_{6_2} V_{10_2}
	E_{9_2} vertices: V_{8_2} V_{11_2}		E_{10_2} vertices: V_{7_2} V_{12_2}

$(3,2,2,1)$			
	E_{11_2} vertices: V_{5_2} V_{8_2}		E_{12_2} vertices: V_{6_2} V_{7_2}
	E_{13_2} vertices: V_{9_2} V_{11_2}		E_{14_2} vertices: V_{10_2} V_{12_2}
	E_{15_2} vertices: V_{1_2} V_{9_2}		E_{16_2} vertices: V_{2_2} V_{11_2}
	E_{17_2} vertices: V_{3_2} V_{10_2}		E_{18_2} vertices: V_{4_2} V_{12_2}
Verticies			

$(3, 2, 2, 1)$

	V_{1_2} edges: E_{1_2} E_{3_2} E_{15_2}		V_{2_2} edges: E_{1_2} E_{4_2} E_{16_2}
	V_{3_2} edges: E_{2_2} E_{3_2} E_{17_2}		V_{4_2} edges: E_{2_2} E_{4_2} E_{18_2}
	V_{5_2} edges: E_{5_2} E_{7_2} E_{11_2}		V_{6_2} edges: E_{5_2} E_{8_2} E_{12_2}
	V_{7_2} edges: E_{6_2} E_{10_2} E_{12_2}		V_{8_2} edges: E_{6_2} E_{9_2} E_{11_2}
	V_{9_2} edges: E_{7_2} E_{13_2} E_{15_2}		V_{10_2} edges: E_{8_2} E_{14_2} E_{17_2}

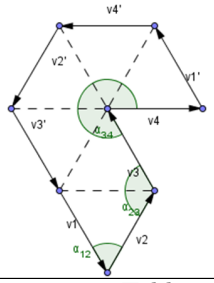
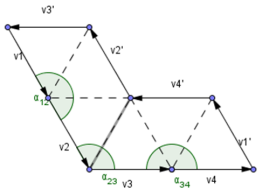
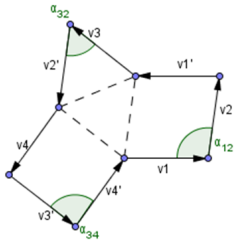
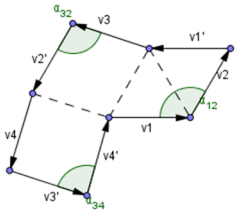
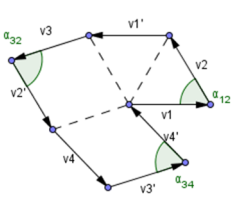
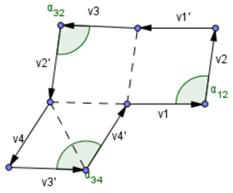
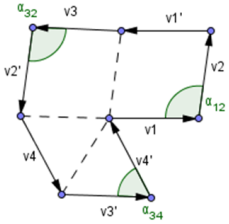
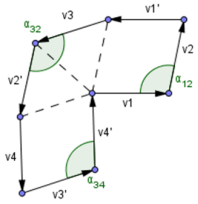
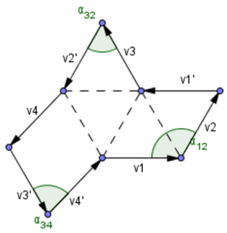
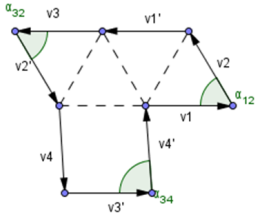
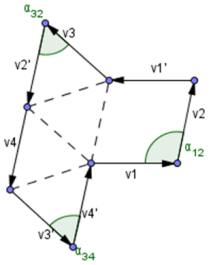
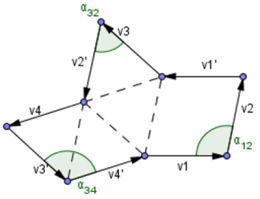
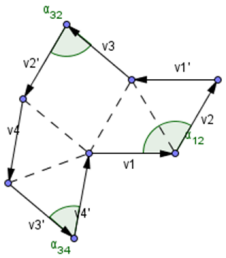
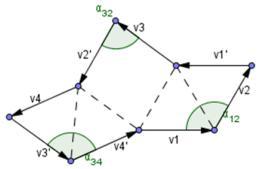
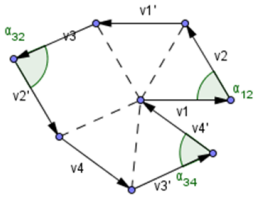
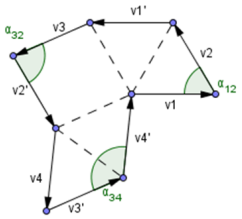
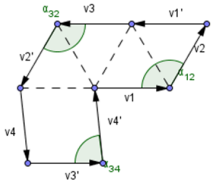
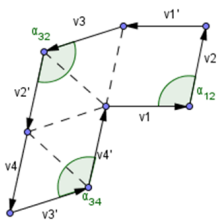
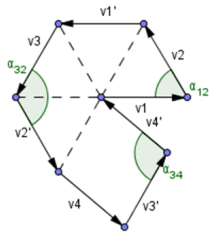
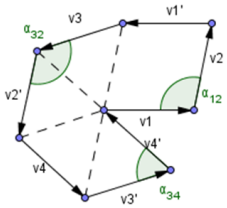
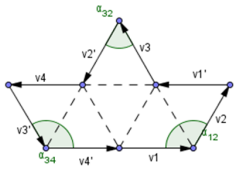
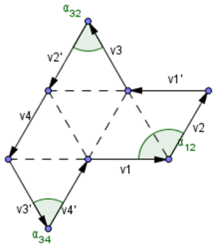
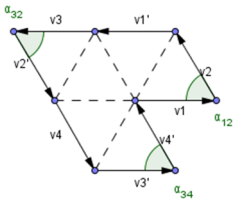
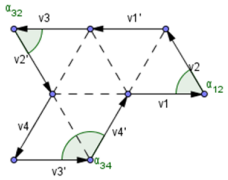
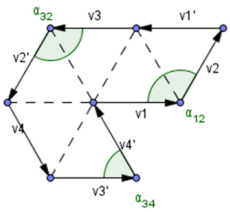
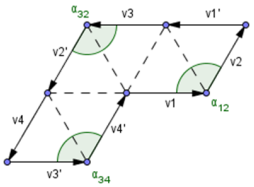
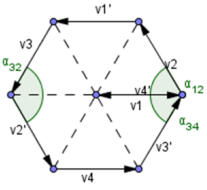
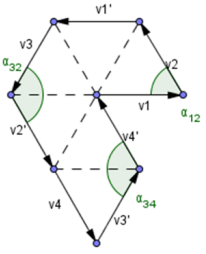
$(3,2,2,1)$			
	V_{11_2} edges: E_{9_2} E_{13_2} E_{16_2}		V_{12_2} edges: E_{10_2} E_{14_2} E_{18_2}

Table 2: This table shows the structure of the linkages of type $(3,2,2,1)$ on the boundary of P_2. (Image help from Kathryn Marsh.)

$(1,2,2,1)$			
Linkage	Name/ Connections	Linkage	Name/ Connections
Faces			
	F_{1_3} edges: E_{1_3} E_{2_3} E_{3_3} E_{4_3}		F_{2_3} edges: E_{1_3} E_{5_3} E_{6_3} E_{9_3}
	F_{3_3} edges: E_{2_3} E_{7_3} E_{8_3} E_{11_3}		F_{4_3} edges: E_{4_3} E_{6_3} E_{8_3} E_{10_3}
	F_{5_3} edges: E_{3_3} E_{5_3} E_{7_3} E_{12_3}		F_{6_3} edges: E_{9_3} E_{10_3} E_{11_3} E_{12_3}

$(1, 2, 2, 1)$

Edges			
	E_{1_3} vertices: V_{1_3} V_{2_3}		E_{2_3} vertices: V_{3_3} V_{4_3}
	E_{3_3} vertices: V_{2_3} V_{3_3}		E_{4_3} vertices: V_{1_3} V_{4_3}
	E_{5_3} vertices: V_{2_3} V_{5_3}		E_{6_3} vertices: V_{1_3} V_{6_3}
	E_{7_3} vertices: V_{3_3} V_{7_3}		E_{8_3} vertices: V_{4_3} V_{8_3}
	E_{9_3} vertices: V_{5_3} V_{6_3}		E_{10_3} vertices: V_{6_3} V_{8_3}

$(1,2,2,1)$			
	E_{11_3} vertices: V_{7_3} V_{8_3}		E_{12_3} vertices: V_{5_3} V_{7_3}
Verticies			
	V_{1_3} edges: E_{1_3} E_{4_3} E_{6_3}		V_{2_3} edges: E_{1_3} E_{3_3} E_{5_3}
	V_{3_3} edges: E_{2_3} E_{3_3} E_{7_3}		V_{4_3} edges: E_{2_3} E_{4_3} E_{8_3}
	V_{5_3} edges: E_{5_3} E_{9_3} E_{12_3}		V_{6_3} edges: E_{6_3} E_{9_3} E_{10_3}
	V_{7_3} edges: E_{7_3} E_{11_3} E_{12_3}		V_{8_3} edges: E_{8_3} E_{10_3} E_{11_3}

$(1, 2, 2, 1)$	

Table 3: This table shows the structure of the linkages of type $(1, 2, 2, 1)$ on the boundary of P_3. (Image help from Kathryn Marsh.)

References

[1] M. A. Armstrong: *Basic Topology.* Springer-Verlag.New York, 1983.

[2] M. Genduphe and C. Judge, *Well-rounded holomorphic 1-forms* 2014 (in preparation)

[3] R. Ghrist: *Elementary Applied Topology*, 2014 (in preparation)

[4] A. Hatcher: *Algebraic Topology.* Cambridge University Press, Cambridge, 2002.

[5] H. Masur, *Ergodic Theory of Translation Surfaces.*

[6] J. R. Munkres: *Elements of Algebraic Topology.*Benjamin/Cummings Publishing, Menlo Park, 1984.

[7] J. R. Munkres: *Topology.* 2nd Ed., Princeton Hall, Upper Saddle River, 2000.

Images made by Elizabeth J. Winkelman, unless noted as otherwise on the figure.

Homotopy Type Theory and $\pi_4(S^3)$

Richard Wong

Abstract

Homotopy Type Theory is a new branch of mathematics that unifies ideas from various fields in mathematics, such as Algebra, Topology, and Logic, in a foundational way. It emphasizes the importance of constructive mathematics, and allows for computer-checked proofs using programming languages such as Agda. This makes HoTT a powerful tool. For example, it is difficult to calculate higher homotopy groups of spheres in the conventional setting, but HoTT makes such computations accessible. We are particularly interested in computing $\pi_4(S^3)$. It has been calculated to be $\mathbb{Z}/2\mathbb{Z}$. Guillaume Brunerie has proven in HoTT that there is some n such that $\pi_4(S^3)$ is $\mathbb{Z}/n\mathbb{Z}$. We seek to compute this n by formalizing the proof through an implementation in Agda.

1 Introduction

Homotopy Type Theory (HoTT) is an extremely new and interesting field of mathematics that interprets the ideas of *type theory*, a field of logic and theoretical computer science, through the lens of *homotopy theory*, a concept that comes from algebraic topology. The combination of the two yields deep connections and offers new insight into various topics, such as higher homotopy groups of spheres, and there is still much to explore.

One central idea of HoTT as an approach to mathematics is the emphasis on constructive mathematics. For example, if one were to give a proof of an existence theorem in HoTT, one would need to give an explicit example or construction. Similarly, a proof that two structures are isomorphic should construct an isomorphism between the two structures.

The other central idea of HoTT is the ability to formalize mathematical arguments type-theoretically. This formalization can then be implemented through proof assistants such as Agda or Coq. This allows for computer-assisted and computer-verified proofs. In fact, many of the arguments and proofs given in the HoTT book were first proven formally through a proof assistant, and then translated back into an informal argument.

These two ideas make HoTT a powerful tool, and offers a new and useful approach to mathematics. For example, in algebraic topology, one area of interest is the study of homotopy groups of spaces. The first homotopy group is the fundamental group, which captures information about loops in the space, and provides global information about the structure of the space. Similarly, higher homotopy groups yield further information about the structure of the

space. The study of the k^{th} homotopy group of S^n, $\pi_k(S^n)$, is a central question in algebraic topology, and has led to the development of many techniques and powerful theorems. However, it is surprisingly difficult to calculate higher homotopy groups in the conventional setting, even for simple spaces such as S^n.

However, because HoTT is constructive and is well-suited for using of computer proof assistants, such computations can be made more accessible. Various conventional results about $\pi_k(S^n)$ have already been proven in HoTT, and it has been used to compute specific homotopy groups, as well as prove theorems about special cases, such as $\pi_n(S^n)$. We are particularly interested in computing $\pi_4(S^3)$, as it is the first stable homotopy group of the form $\pi_{n+1}(S^n)$. In the conventional setting, it has been computed to be $\mathbb{Z}/2\mathbb{Z}$. It has been proven informally in HoTT that there is some n such that $\pi_4(S^3)$ is $\mathbb{Z}/n\mathbb{Z}$. We seek to compute this n by formalizing the proof through an implementation in Agda.

2 Background

2.1 Type Theory and Agda

Type theory is a formal system in theoretical computer science and logic. It is an alternative to set theory as a foundational language for mathematics. The advantages of using type theory include the ability to formalize arguments and run them through a proof assistant such as Agda, which can pattern match and type-check arguments to verify a proof.

In set theory, there are two major components. There is the deductive system of first-order logic, and within the deductive system are the axioms of a particular theory, such as ZFC. As a result, in set theory, there are the two basic notions of propositions and sets, and they exist on different levels. However, type theory is its own deductive system, and has only the one basic notion of types.

In type theory, propositions and sets are on the same level. So it is possible to interpret the type $A : \mathcal{U}$ as either a set or as a proposition, or both. If A is inhabited by $a : A$, then if A is viewed as a set, one can roughly think of a as an element of A. Note, however, that while $a \in A$ is a proposition, $a : A$ is not. It is instead an atomic statement, and cannot be proved or disproved. If A is viewed as a proposition, then one considers $a : A$ to be a proof of the proposition, or a witness to the truth of A. In this way, the proof of a proposition in type theory can be thought of as constructing an inhabitant of a type.

Because of this fact, in type theory, one needs to consider two types of equality: propositional equality and definitional equality. The former, *propositional equality*, is the traditional set-theoretic notion of equality as a proposition. It exists as a type, and can be inhabited or uninhabited. For example, if $a, b : A$, then the type of the equality proposition is $a =_A b$, which may or may not be inhabited. However, there is also the notion of *definitional equality*, which is not a type, but an atomic statement at the same level as $a : A$. For example, if we have a function $f : \mathbb{N} \to \mathbb{N}$ by $f(x) = x + 1$, then $f(2)$ is definitionally equal

to $2 + 1$. Definitional equality is written as $a \equiv b : A$. Unlike propositional equality, it does not make sense to prove or disprove definitional equality. It is instead a matter of checking the definitions, which is typically decidable.

Remark 2.1 When we say that A is a type, we write $A : \mathcal{U}$. Here, $\mathcal{U}$ is called a **universe**, a type whose elements are types. There are many subtleties in the precise definition of a universe to avoid paradoxes, but we are not concerned with this issue here.

Using basic type theory, one can construct familiar types such as the Boolean data type or the type of natural numbers. To define the Booleans as a type Bool $: \mathcal{U}$, one simply needs to define the inhabitants true and false : Bool, and the boolean negation function not : Bool $\to$ Bool that takes true to false, and false to true.

One can also define the natural numbers as a type Nat : $\mathcal{U}$ recursively using rules that are similar to the Peano Axioms. One defines the inhabitant zero : Nat, and a function suc : Nat $\to$ Nat. Note that this is enough to define the natural numbers. It is not necessary to require the suc zero $\neq$ zero, as there are no constructors. In other words, there is no means to construct an inhabitant of the type suc zero $=$ zero. The only tool to prove propositional equality that is available for this type is reflexivity - $\text{refl}_a : a =_A a$, which says that an inhabitant is equal to itself.

The following is code in Agda that formalizes this construction of the Booleans and the natural numbers, as well as a recursive definition of addition on the natural numbers.

```
module example1 where

-Definition of a Boolean set and its operations.

data Bool : Set where
  true  : Bool
  false : Bool

not : Bool → Bool
not true  = false
not false = true

-Definition of the natural numbers and addition.

data Nat : Set where
  zero : Nat
  suc  : Nat → Nat

_+_ : Nat → Nat → Nat
zero + m = m
suc n + m = suc (n + m)
```

The following code is an example of a formalized proof in Agda. Using the definition of natural numbers in example1, we first define the ordering "leq" ($\leq$) on the natural numbers, and then prove that it is transitive.

```
module example2 where

open example1

-A proof that leq is transitive

data _leq_ : Nat → Nat → Set where
  leq-zero  : {n : Nat} → zero leq n
  leq-suc   : {m n : Nat} → n leq m → (suc n) leq (suc m)

leq-trans : {l m leq : Nat} → l leq m → m leq n → l leq n
leq-trans leq-zero _              = leq-zero
leq-trans (leq-suc p) (leq-suc q) = leq-suc (leq-trans p q)
```

Our definition of "leq" is done inductively. The first case says that zero is less than or equal to any n : Nat. The second case says that if n is less than or equal to m, then the successor function preserves the ordering.

Our proof of the transitivity can be thought of as a function in the following way: If it is given l, m, n : Nat, and two proofs that l leq m and m leq n, then it will return a proof that l leq n. In the first case, if the first proof is from the first case of our definition, leq-zero, then l must be zero, and so we need a proof that zero leq n. However, we can prove that using leq-zero.

In the second case, if we have a proof leq-suc p that (suc l) leq (suc n) and a proof leq-suc q that (suc n) leq (suc m), we can unwind the definition to obtain the proof p that l leq m, and the proof q that m leq n. We can then reduce the size of l, m, and n, and so we can iterate this process until we reduce to the first case.

2.2 Homotopy and Homotopy Type Theory

First, let us first recall the notions of paths and homotopies in a topological space.

Definition 2.1 Let X be a topological space. A **path** between x and y in X is a continuous map $p : [0, 1] \times X \to X$ such that $p(0) = x$ and $p(1) = y$.

We would like to know when two paths p, q are the same. We note that pointwise equality of paths is too strict of a notion, since $p^{-1} \circ p$ is not pointwise equivalent to the identity path. There is, however, a homotopy between $p^{-1} \circ p$ and the identity path. So we consider two paths to be the same if there is a homotopy between them.

Definition 2.2 Let X be a topological space. Let p, q be paths in X from x to y. A **homotopy of paths** between p and q in X is a continuous map $H(s,t) : [0,1] \times [0,1] \to X$ such that $H(z,0) = p(z)$ and $H(z,1) = q(z)$. We say that $H(s,t)$ is **endpoint-preserving** if $H(0,z) = x$ and $H(1,z) = y$

Theorem 2.3 *Homotopy of paths is an equivalence relation, and is preserved by operations such as inverses and concatenation.*

HoTT applies the lens of homotopy to type theory in the following way. Recall that the type $A : \mathcal{U}$ can be viewed as either a type or a set. Through the lens of homotopy theory, one can also view A as a topological space. In this way, one should consider an inhabitant of the equality type, $p : x =_A y$, as a path p from x to y in A. Then, if one wanted to show that x is equal to y in the type A, one should construct a path p between them.

Furthermore, if there were two proofs $p, q : A$, one would want to know whether these proofs are, in some sense, the same. To do so, one can consider $A : \mathcal{U}$ from the homotopy perspective and determine if p and q are homotopic. For example, if $p, q : x =_A y$ are considered as paths from x to y in A, if they are homotopic as paths, then they are in some sense the same proof.

Applying the lens of homotopy to type theory makes HoTT both concrete and abstract. It is concrete in the sense that the mathematics is constructive, and arguments and proofs refer to concrete points, paths, inhabitants, and types. However, HoTT is also abstract in the sense that it sees these specific objects only up to homotopy. So it is not necessary to painstakingly construct an explicit homotopy between paths, for instance.

It turns out that this notion is extremely useful for dealing with higher inductive types and higher homotopy groups.

We will need the following notions and definitions to discuss and prove theorems about higher homotopy groups.

Definition 2.4 A **pointed type** (A, a) is a type $A : \mathcal{U}$, along with a point $a : A$, the **base point**.

Definition 2.5 A type $A : \mathcal{U}$ is **contractible** if there is $a : A$ such that $a =_A x$ for all $x : A$ Then $A = \mathbf{1}$, where 1 is the **unit type**.

Definition 2.6 The **loop space** of a pointed type (A, a) is defined to be the pointed type $\Omega(A, a) :\equiv (a =_A a, \mathrm{refl}_a)$. An element of $\Omega(A, a)$ is called a **loop** at a.

Definition 2.7 Let $n : \mathbb{N}$. The **n-fold iterated loop space** of a pointed type (A, a) is a pointed type defined recursively:

$$\Omega^0(A, a) :\equiv (A, a)$$
$$\Omega^{n+1}(A, a) :\equiv \Omega^n(\Omega(A, a))$$

The elements of $\Omega^n(A, a)$ are called the n-dimensional loops at a.

The loop space is an example of a higher inductive type. A higher inductive type is a type that is generated by "constructors", which generate not just points, but may also generate paths and higher paths. For example, one can construct S^n as a higher inductive type generated by a point $base : S^n$, and the n-dimensional loop $loop_n : \Omega^n(S^n, base)$.

Definition 2.8 Let $A : \mathcal{U}$ be a type. The **suspension** of A, ΣA, is defined by the following generators:

- A point $N : \Sigma A$
- A point $S : \Sigma A$
- A function $merid : A \to N =_{\Sigma A} S$

The suspension is another higher inductive type that can be created from A. For each point in A, there is a path from N to S. The naming scheme is suggestive of a globe, and in fact, one can define S^n as the suspension of S^{n-1}.

Remark 2.9 It is possible to show that the two definitions of S^n as a loop space and of S^n as a suspension are, in fact, equivalent.

Definition 2.10 Given a type $A : \mathcal{U}$, and n such that $n \geq -1$, we define the **n-truncation of A**, $\|A\|_n$, as a higher inductive type generated by:

- A function $| - |_n : A \to \|A\|_n$
- For each $t : S^{n+1} \to \|A\|_n$, a point $h(t) : \|A\|_n$.
- For each $t : S^{n+1} \to \|A\|_n$ and for each $x : S^{n+1}$, a path $s_t(x) : t(x) = h(t)$

In homotopy theory, an n-truncated space has no interesting information above dimension n. In other words, with an n-truncated type, one forgets about paths of dimension n and higher. The 0-truncation of a type $A : \mathcal{U}$ can be thought of as simply a set with no non-trivial paths between its inhabitants. The (-1)-truncation of a type is a mere proposition - the interesting information is whether the type is inhabited or not. Though it is not covered in this definition, we define the (-2)-truncation of a type to be a contractible type.

The idea of forgetting about higher dimensions serves as the intuition for the following remarks:

Remark 2.11 Let $n, k \geq -2$, with $k \leq n$. Then for a type $A : \mathcal{U}$,

$$\| \|A\|_n \|_k = \|A\|_k$$

Remark 2.12 For all $n \geq -2$, $k \geq 0$,

$$\|\Omega^k(S^n)\|_n = \Omega^k(\|S^n\|_{n+k})$$

In fact, this is true for all pointed types (A, a).

Definition 2.13 A type $A : \mathcal{U}$ is **n-connected** if $\|A\|_n$ is contractible.

The notion of n-connectedness is the opposite of n-truncation. An n-connected type can be thought of as having no interesting information below dimension n.

Remark 2.14 For all $n : \mathbb{N}$, the n-sphere is $(n-1)$-connected.

Definition 2.15 Given a span of types and functions:

$$\begin{array}{ccc} C & \xrightarrow{g} & B \\ \downarrow{\scriptstyle f} & & \\ A & & \end{array}$$

The **pushout** $A \sqcup^C B$ is the higher inductive type generated by:

- A function $inl : A \to A \sqcup^C B$
- A function $inr : B \to A \sqcup^C B$
- for each $c : C$ a path $glue(c) : (inl(f(c)) = inr(g(c)))$

So we have the following diagram:

$$\begin{array}{ccc} C & \xrightarrow{g} & B \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle inr} \\ A & \xrightarrow{inl} & A \sqcup^C B \end{array}$$

The pushout of $A \sqcup^C B$ can be thought of as the disjoint union of A and B, with the condition that for every $c : C$ there is a witness that $f(c)$ and $g(c)$ are equal.

2.3 Higher Homotopy Groups

The study of higher homotopy groups is a central question in algebraic topology, and has led to the development of many modern techniques in algebraic topology. There are many results known about higher homotopy groups of spheres, and while many have been calculated, they have not yet been fully classified. In addition, such calculations are difficult in the conventional setting, though HoTT offers ways to make such calculations more accessible.

The k^{th} homotopy group of a pointed type (A, a), $\pi_k(A, a)$, can be thought of as the set of mappings up to homotopy of the pointed k-sphere (S^k, b) into the space A such that base points are preserved. However, it is equivalent and more convenient to use the following definition using n-fold loop spaces and truncations:

Definition 2.1 Given $k \geq 1$, the **kth homotopy group** of a pointed type (A, a), $\pi_k(A, a)$, is defined to be $\|\Omega^k(A, a)\|_0$.

Recall that the 0-truncation makes $\Omega^n(A, a)$ a set, and the concatenation and inverse operations of paths in $\Omega^n(A, a)$ make $\|\Omega^n(A, a)\|_0$ a group. When considering higher homotopy groups of spheres, we will be a bit lazy in our notation and assume the base point to be implicit.

Remark 2.2 If $k > 2$, then $\pi_k(S^n)$ is abelian. This is a result of Eckmann-Hilton argument.

Theorem 2.3 *If $k < n$, then $\pi_k(S^n)$ is trivial.*

In homotopy theory, one can intuitively see that for $k < n$, a k-dimensional sphere mapped onto an n-dimensional sphere can be contracted down to a point.
Proof

$$\begin{aligned}
\pi_k(S^n) &= \|\Omega^k(S^n)\|_0 \\
&= \Omega^k(\|S^n\|_k) && \text{(by Remark 2.12)} \\
&= \Omega^k(\| \, \|S^n\|_{n-1} \, \|_k) && \text{(by Remark 2.13)} \\
&= \Omega^k(\|\mathbf{1}\|_k) && \text{(by Remark 2.14)} \\
&= \Omega^k(\mathbf{1}) \\
&= \mathbf{1}
\end{aligned}$$

Recall that by remark 2.14, S^n is $(n-1)$-connected, so the $(n-1)$-truncation is contractible, so $\|S^n\|_{n-1} = \mathbf{1}$. □

Before we can prove our next result about higher homotopy groups of spheres, we will need to state a key theorem in algebraic topology, the Freudenthal suspension theorem, which we will state here in the language of HoTT:

Theorem 2.4 (Freudenthal Suspension Theorem) *Suppose X is a pointed, n-connected type, with $n \geq 0$. Then $\|X\|_{2n} \simeq \|\Omega(\Sigma(X))\|_{2n}$*

The following theorem is extremely useful in classifying and computing higher homotopy groups. As a consequence of the theorem, we are particularly interested in $\pi_4(S^3)$ because it is the first stable homotopy group of the form $\pi_{n+1}(S^n)$.

Theorem 2.5 (Stability of Homotopy Groups of Spheres) *If $k < 2n - 2$, then $\pi_{k+1}(S^{n+1}) = \pi_k(S^n)$.*

Proof

$$\begin{aligned}
\pi_{k+1}(S^{n+1}) &= \|\Omega^{k+1}(S^{n+1})\|_0 \\
&= \|\Omega^k(\Omega(S^{n+1}))\|_0 \\
&= \Omega^k(\|\Omega(S^{n+1})\|_k) \\
&= \Omega^k(\|\Omega(\Sigma(S^n))\|_k) \\
&= \Omega^k(\|\,\|\Omega(\Sigma(S^n))\|_{2n}\,\|_k) \\
&= \Omega^k(\|\,\|S^n\|_{2n}\,\|_k) && \text{(by Freudenthal Suspension Theorem)} \\
&= \Omega^k(\|S^n\|_k) \\
&= \|\Omega^k(S^n)\|_0 \\
&= \pi_k(S^n)
\end{aligned}$$

□

Note that the key step is the use of the Freudenthal Suspension theorem, while the rest of the steps are the unfolding of definitions and the use of remarks 2.12 and 2.13.

The following table can be found in the Homotopy Type Theory book, and lists some of the computed higher homotopy groups of spheres. One important computation that has already been done in HoTT is the computation that $\pi_1(S^1) = \mathbb{Z}$. This result leads to a proof in HoTT that $\pi_n(S^n) = \mathbb{Z}$.

	S^0	S^1	S^2	S^3	S^4	S^5	S^6	S^7	S^8
π_1	0	$\mathbb{Z}$	0	0	0	0	0	0	0
π_2	0	0	$\mathbb{Z}$	0	0	0	0	0	0
π_3	0	0	$\mathbb{Z}$	$\mathbb{Z}$	0	0	0	0	0
π_4	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	0
π_5	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0
π_6	0	0	$\mathbb{Z}_{12}$	$\mathbb{Z}_{12}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0
π_7	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}\times\mathbb{Z}_{12}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0
π_8	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2^2$	$\mathbb{Z}_{24}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$
π_9	0	0	$\mathbb{Z}_3$	$\mathbb{Z}_3$	$\mathbb{Z}_2^2$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$
π_{10}	0	0	$\mathbb{Z}_{15}$	$\mathbb{Z}_{15}$	$\mathbb{Z}_{24}\times\mathbb{Z}_3$	$\mathbb{Z}_2$	0	$\mathbb{Z}_{24}$	$\mathbb{Z}_2$
π_{11}	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{15}$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	$\mathbb{Z}_{24}$
π_{12}	0	0	$\mathbb{Z}_2^2$	$\mathbb{Z}_2^2$	$\mathbb{Z}_2$	$\mathbb{Z}_{30}$	$\mathbb{Z}_2$	0	0
π_{13}	0	0	$\mathbb{Z}_{12}\times\mathbb{Z}_2$	$\mathbb{Z}_{12}\times\mathbb{Z}_2$	$\mathbb{Z}_2^3$	$\mathbb{Z}_2$	$\mathbb{Z}_{60}$	$\mathbb{Z}_2$	0

Figure 2.3.1: The higher homotopy groups of spheres. The colored diagonals indicate the stable homotopy groups.

3 Guillaume Brunerie's Proof that $\pi_4(S^3) = \mathbb{Z}/n\mathbb{Z}$

While $\pi_4(S^3)$ has been computed to be $\mathbb{Z}/2\mathbb{Z}$, it is not possible to simply import the argument into HoTT, as the conventional proof makes use of arguments that are not homotopy invariant. Brunerie instead uses the approach of the James Construction, which is useful because the James Construction of a space A, denoted $J(A)$, has the same type as the loop space of the suspension of the space A, $\Omega(\Sigma A)$.

Let us first obtain some intuition about the James Construction. Let A be a pointed type that is k-connected. Then we will define types $J_2(A)$, $J_3(A)$ such that we have the following sequence of types and maps, with certain conditions on the connectedness of the maps:

$$1 \to A \to J_2(A) \to J_3(A) \to J_4(A) \to ...$$

Then we define the James Construction of A, $J(A)$ to be the following colimit:

$$\lim_{m\to} J_m(A) :\equiv J(A)$$

Theorem 3.1 *Given a pointed type A,* $J(A) = \Omega(\Sigma A)$

One can intuitively think of $J_n(A)$ as successive approximations of $\Omega(\Sigma A)$. We will now inductively define $J_n(A)$.

Definition 3.2 Let (A, a_0) be a pointed type. Then there is a map $a_0 : \mathbf{1} \to A$ sending $1 \to a_0$. Then we define $J_n(A)$ recursively:
Let $J_0(A) := \mathbf{1}$. Then $J_{n+1}(A)$ is a type with:

- An inclusion map $i_n : J_n(A) \to J_{n+1}(A)$
- A constructor map $\alpha_n : J_0(A) \times J_n(A) \to J_{n+1}(A)$

Given $x : J_n(A)$, these maps satisfy the condition $\alpha_n(a_0, x) = i_n(x)$. Furthermore, they must satisfy the following relation:

$$\alpha_{n+1}(a, i_n(x)) = i_{n+1}(\alpha_n(a, x)) = i_{n+1}(a_n x)$$

So we can see J_{n+2} as the pushout of the following diagram:

$$\begin{array}{ccc} J_{n+1}(A) \sqcup^{J_n(A)} (A \times J_n(A)) & \xrightarrow{g} & A \times J_{n+1}(A) \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle \alpha_{n+1}} \\ J_{n+1}(A) & \xrightarrow{i_{n+1}} & J_{n+2}(A) \end{array}$$

Now let us consider the case $A = S^2$. Note that by Theorem 3.1, $J_1(S^2) = S^2$, and so we have the following sequence:

$$1 \to S^2 \to J_2(S^2) \to J_3(S^2) \to ...$$

And we have that

$$\lim_{m\to} J_m(S^2) = \Omega(\Sigma S^2) = \Omega(S^3)$$

By a condition on the connectedness of the map from $J_2(S^2) \to \Omega(S^3)$ (it is 4-connected), we have that $\pi_3(J_2(S^2)) = \pi_3(\Omega(S^3))$. However, we see that $\pi_3(\Omega(S^3)) = \pi_4(S^3)$ from the definition of the 4th homotopy group of S^3. So now, rather than computing $\pi_4(S^3)$, we are instead interested in computing $\pi_3(J_2(S^2))$.

When we consider the diagram associated to $J_{n+2}(A)$, with $A = S^2$, and $n = 0$, we have the following:

$$\begin{array}{ccc} S^2 \sqcup^{1\times 1} (S^2 \times \mathbf{1}) & \longrightarrow & S^2 \times S^2 \\ \downarrow & & \downarrow{\scriptstyle \alpha_{n+1}} \\ S^2 & \xrightarrow{i_{n+1}} & J_2(S^2) \end{array}$$

We see that we can view $J_2(S^2)$ as the following pushout:

$$J_2(S^2) = S^2 \sqcup^{(S^2 \vee S^2)} (S^2 \times S^2)$$

Note that $(S^2 \vee S^2)$ is shorthand for the pushout $S^2 \sqcup^1 S^2$.

$$\begin{array}{ccc} \mathbf{1} & \xrightarrow{g} & S^2 \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle inr} \\ S^2 & \xrightarrow{inl} & S^2 \sqcup^1 S^2 \end{array}$$

We would like to simplify our definition of $J_2(S^2)$ as a pushout. To do so, we claim that $S^2 \times S^2 = (S^2 \vee S^2) \sqcup^{S^3} \mathbf{1}$, and omit the proof. This allows us to make the following simplifications:

$$\begin{aligned} J_2(S^2) &= S^2 \sqcup^{(S^2 \vee S^2)} (S^2 \times S^2) \\ &= S^2 \sqcup^{(S^2 \vee S^2)} ((S^2 \vee S^2) \sqcup^{S^3} \mathbf{1}) \\ &= (S^2 \sqcup^{(S^2 \vee S^2)} (S^2 \vee S^2)) \sqcup^{S^3} \mathbf{1} \\ &= S^2 \sqcup^{S^3} \mathbf{1} \end{aligned}$$

We can now view $J_2(S^2)$ as the following pushout for some $f : S^3 \to S^2$.

$$\begin{array}{ccc} S^3 & \xrightarrow{g} & \mathbf{1} \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle \alpha_{n+1}} \\ S^2 & \xrightarrow{i_{n+1}} & J_2(S^2) \end{array}$$

Again, by a condition on the connectedness of the maps, we have a map $\pi_3(f) : \pi_3(S^3) \to \pi_3(S^2)$. However, the higher homotopy groups $\pi_3(S^3)$ and $\pi_3(S^2)$ have already been computed to be $\mathbb{Z}$ both in the conventional setting and in HoTT. Then the map $\pi_3(f) : \mathbb{Z} \to \mathbb{Z}$ is just multiplication by some number n. It can be shown using a few more tools and high-powered arguments from homotopy theory, such as the Blakers-Massey Theorem, that $\pi_3(J_2(S^2)) = \mathbb{Z}/n\mathbb{Z}$, where this n is exactly the n from the map $\pi_3(f)$.

4 Proof in Agda

```
module jamesconstructpi4s3 where

open import Function
open import Data.Nat

--------------------------------------
- Identity types, from Amr's circle code

- Our own version of propositional equality that makes 'a' explicit

data _≡_ {ℓ} {A : Set ℓ} : (a b : A) → Set ℓ where
  refl : (a : A) → (a ≡ a)

pathInd : ∀ {ℓ ℓ'} → {A : Set ℓ} →
          (C : {x y : A} → (x ≡ y) → Set ℓ') →
          (c : (x : A) → C (refl x)) →
          ({x y : A} (p : x ≡ y) → C p)
pathInd C c (refl x) = c x

- Transport

- Lemma 2.3.1

transport : ∀ {ℓ ℓ'} → {A : Set ℓ} {x y : A} →
  (P : A → Set ℓ') → (p : x ≡ y) → P x → P y
transport {ℓ} {ℓ'} {A} {x} {y} P p =
  pathInd
    (λ {x} {y} p → (P x → P y))
    (λ _ → id)
    {x} {y} p
```

```
--------------------------------------

-let's define S^3.  we need to define S^1 and then loops.

module circle where
  private
    data S¹* : Set where
      base* : S¹*

  S¹ : Set
  S¹ = S¹*

  base : S¹
  base = base*

  postulate
    loop : base ≡ base

  recS¹ : {C : Set} → (cbase : C) → (cloop : cbase ≡ cbase) → S¹ → C
  recS¹ cbase cloop base = cbase

  indS¹ : {P : S¹ → Set} →
    (pbase : P base) → (ploop : transport P loop pbase ≡ pbase) →
    (point : S¹) → P point
  indS¹ pbase ploop base* = pbase

open circle public
```

5 Further Steps

While we were not able to finish the formalization of Brunerie's proof, we have made progress in understanding his proof and have taken steps to determine what the n is. From here, we should do finish formalizing the definition of $J_n(A)$, and then formalize Brunerie's argument that $J_2(S^2) = S^2 \sqcup^{S^3} \mathbf{1}$.

Furthermore, once the calculation of this n has been completed, further steps to take include computing other stable and unstable homotopy groups of spheres. Currently HoTT has not yet computed any previously-unknown homotopy groups of spheres, though it is expected that HoTT will be able to do so. One hopes that through the tools of HoTT, one will be able to completely classify higher homotopy groups of spheres, and/or provide a general method to calculate higher homotopy groups of spheres.

Another possible direction for future research would be to expand HoTT as a general framework for mathematics. In addition to homotopy theory, category theory, set theory, and the real numbers have all been constructed and/or formalized through HoTT. It would be interesting to develop further areas of

mathematics in HoTT.

References

[HoTT] The Univalent Foundations Program: Homotopy Type Theory: Univalent Foundations of Mathematics. Institute for Advanced Study, 2013.

[Bru] G. Brunerie: The James Construction and $\pi_4(S^3)$. University Lecture. Institute for Advanced Study, 2013.

[Lic] D. Licata: Pi4S3.agda. Github, Github Repository, `https://github.com/dlicata335/hott-agda/blob/master/homotopy/in-progress/Pi4S3.agda`, 2013.