# Research Experience for Undergraduates Program Research Reports

Indiana University, Bloomington

Summer 2019

# Contents

<b>Unusual Topographical Features of Laplacian Eigenfunctions</b> <i>Kelly Chen and Olti Myrtaj</i>	5
Extremal Length of Winding Curves	15
Mixing of Random Walks on Punctured Graphs	32
Cauchy-Fueter Formulas for Universal Clifford Algebras of In- definite Signature	43
<b>On the Large Time Diffusion of Isotropic and Anisotropic Particles</b> <i>Mikhail Sweeney</i>	72
Bernoulli Percolation on Hyperbolic Graphs	€9
Spectral Properties of Reversible Indel Models	)5

# Preface

During the summer of 2019 eight students participated in the Research Experience for Undergraduates program in Mathematics at Indiana University. This program was sponsored by the National Science Foundation through the Research Experience for Undergraduates grant DMS-1461061 and the Department of Mathematics at Indiana University, Bloomington. The program ran for eight weeks, from June 3 through July 26, 2019. Eight faculty served as research advisers to the students from Indiana University:

- Kelly Chen and Olti Myrtaj were advised by Chris Judge.
- Max Newman was advised by Dylan Thurston.
- Chung Kyong Nguen was advised by Graham White.
- Ely Sandine was advised by Matvei Libine.
- Mikhail Sweeney and Linden Yuan was advised by Louis Fan.
- Christine Sullivan was advised by Chris Connell.

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 in a series of workshops. Other opportunities included the option to participate in a GRE and subject test preparation seminar. Additional educational activities included tours of the library, the Low Energy Neutron Source facility, the Slocum puzzle collection at the Lilly Library, 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 statewide Indiana Summer Undergraduate Research conference which was hosted at Indiana University - Purdue University at Indianapolis.

On the lighter side, students were treated to weekly board game nights as well as the opportunity to do some local hiking. They also enjoyed a night of "laser tag" courtesy of the Department of Mathematics. 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 and the Indiana University Bloomington Department of Mathematics without whose support this program could not exist. We especially thank our staff member Mandie McCarty for coordinating the complex logistical arrangments (housing, paychecks, information packets, meal plans, frequent shopping for snacks). Additional logistical support was provided by the Department of Mathematics and our chair, Elizabeth Housworth. We are in particular thankful to Jeff Taylor for the computer support he provided. Thanks also go to those faculty who served as mentors and those who gave lectures. We thank David Baxter of the Center for Exploration of Energy and Matter (nee IU cyclotron facility) for past personal tours of the LENS facility and his informative lectures. Thanks to Andrew Rhoda for his tour of the Slocum Puzzle Collection.

Chris Connell September, 2019



Figure 1: REU Participants, from left to right: Chris Connell, Kelly Chen, Mikhail Sweeney, Ely Sandine, Max Newman, Linden Yuan, Olti Myrtaj, Christine Sullivan and Chung Kyong Nguen (not pictured).

## Unusual Topographical Features of Laplacian Eigenfunctions

Kelly Chen<sup>\*</sup> Olti Myrtaj<sup>†</sup> kjchen@mit.edu olti@vt.edu

Summer 2019

#### Abstract

Solutions to the Laplace eigenvalue problem satisfying Dirichlet boundary conditions on a rectangular region of the plane are well known. In this work, we study interesting features of the level sets of such eigenfunctions. We consider the possibility of degenerate critical points with Poincaré-Hopf index 0, which correspond to cusp-like points on the level sets. In particular, we construct a one-parameter family of eigenfunctions and prove the existence of an index-0 critical point.<sup>1</sup>

#### 1 Introduction

Let  $\Omega$  be a domain with piecewise smooth boundary. A smooth function  $\psi : \Omega \to \mathbb{R}$  is called a Dirichlet eigenfunction of the Laplacian if and only if there exists a real number  $\lambda$  such that

$$\begin{array}{rcl}
-\Delta\psi(x,y) &=& \lambda \cdot \psi(x,y) & \text{ for } (x,y) \in \Omega, \\
\psi(x,y) &=& 0 & \text{ for } (x,y) \in \partial\Omega.
\end{array} \tag{1}$$

The number  $\lambda$  is called the *eigenvalue* associated with the eigenfunction  $\psi$ . It is known [Ulb76] that for a generic choice of domain, the critical points of each eigenfunction are nondegenerate. In particular, there are no zeros of the gradient vector field  $\nabla \psi$  that have Poincaré-Hopf index equal to zero. (Lemma 2.5 describes the connection between the index and the nondegeneracy of a critical point.)

We restrict our attention to the case where  $\Omega$  is the square  $[0, \pi] \times [0, \pi]$ , which is not a generic domain. We show that in this case, there are in fact critical points of index 0:

**Theorem 1.1.** There exists a Dirichlet function of the Laplacian on  $[0, \pi] \times [0, \pi]$  with an index-0 critical point in  $(0, \pi) \times (0, \pi)$ .

On the square, one may verify that if m and n are integers, then the function

$$\psi_{m,n}(x,y) = \sin(mx)\sin(ny)$$

is a solution to (1) with  $\lambda = \lambda_{m,n} = m^2 + n^2$ . In fact, it is is well-known [Crn-Hlb, pp. 300-301] that each eigenvalue has the form  $m^2 + n^2$ , where m and n are integers, and a function  $\psi$  is a Dirichlet eigenfunction of the Laplacian with eigenvalue  $\lambda$  if and only if

$$\psi = \sum_{m^2+n^2=\lambda} a_{m,n} \psi_{m,n},$$

where  $a_{m,n}$  are real numbers.

In this paper, we consider  $\lambda = 65$ , the smallest  $\lambda$  that generates a four-dimensional eigenspace. We observe that  $1^2 + 8^2 = 4^2 + 7^2 = 65$ , so we construct the linear combination

$$f(x,y) = a_{1,8} \cdot \psi_{1,8}(x,y) + a_{4,7} \cdot \psi_{4,7}(x,y) + a_{8,1} \cdot \psi_{8,1}(x,y) + a_{7,4} \cdot \psi_{7,4}(x,y).$$

<sup>\*</sup>Massachusetts Institute of Technology

<sup>&</sup>lt;sup>†</sup>Virginia Tech

<sup>&</sup>lt;sup>1</sup>This material is based upon work supported by the National Science Foundation under Grant No. DMS-1461061.

We study the one-parameter family of eigenfunctions

$$f^{t}(x,y) = \cos(t)\sin(x)\sin(8y) + \sin(t)\sin(4x)\sin(7y), \ t \in [0,\pi],$$

which contains all possible f with  $a_{8,1} = a_{7,4} = 0$ , up to scalar multiplication. To prove Theorem 1.1, we show that there exists a t such that  $f^t$  has an index-0 critical point.

#### 2 Preliminaries

Recall the following definitions:

**Definition 2.1.** A function  $g: \mathbb{R}^2 \to \mathbb{R}$  has a **critical point** at  $(x_0, y_0)$  if and only if the gradient  $\nabla g(x, y) = (g_x(x, y), g_y(x, y))$  is the zero vector at  $(x_0, y_0)$ .

**Definition 2.2.** A critical point  $(x_0, y_0)$  of g is called **degenerate** if the Hessian

$$\begin{pmatrix} g_{xx}(x,y) & g_{xy}(x,y) \\ g_{yx}(x,y) & g_{yy}(x,y) \end{pmatrix}$$

has determinant equal to 0 at  $(x_0, y_0)$ .

**Definition 2.3.** Let  $\mathbf{v}: \mathbb{R}^2 \to \mathbb{R}^2$ , where  $\mathbf{v} = (v_1, v_2)$ , be a smooth vector field, and let  $\gamma$  be a simple closed curve <sup>2</sup> in  $\mathbb{R}^2$ . If  $\mathbf{v}$  does not vanish on  $\gamma$ , then the **Poincarè-Hopf index** (or simply index) of  $\mathbf{v}$  along  $\gamma$  is given by

$$\operatorname{ind}(\mathbf{v},\gamma) = \frac{1}{2\pi} \int_{\gamma} \frac{v_1 dv_2 - v_2 dv_1}{v_1^2 + v_2^2}.$$
(2)

Note that in order for  $\operatorname{ind}(\mathbf{v}, \gamma)$  to be well-defined,  $v_1^2 + v_2^2$  must be nonzero on  $\gamma$ . If  $\mathbf{v}$  is the gradient vector field of a function  $g: \mathbb{R}^2 \to \mathbb{R}$ , then this condition is equivalent to g having no critical points on  $\gamma$ .

Next, we consider some properties of the index and some relevant examples.

ı

**Proposition 2.4.** Let  $\mathbf{v}$  be a smooth vector field and let  $\gamma$  be a smooth simple closed curve. Then  $\operatorname{ind}(\mathbf{v}, \gamma)$  is an integer. Proof. Abusing notation slightly, we let  $\gamma : [a, b] \to \mathbb{R}^2$  be a parameterization of  $\gamma$ . If we define

$$\omega := \frac{x\,dy - y\,dx}{x^2 + y^2}$$

then

$$v^*(\omega) = \frac{v_1 dv_2 - v_2 dv_1}{v_1^2 + v_2^2}.$$

Therefore, by a standard fact concerning differential forms we find that

$$\int_{\gamma} \frac{v_1 dv_2 - v_2 dv_1}{v_1^2 + v_2^2} = \int_{\gamma} v^*(\omega) = \int_{v(\gamma)} \omega.$$

The curve  $v(\gamma)$  is a closed curve—not necessarily simple—that does not contain the origin since  $\mathbf{v} \neq 0$  on  $\gamma$ . Define  $\alpha : [a, b] \to \mathbb{R}^2$  by

$$\alpha(t) = \frac{v \circ \gamma(t)}{|v \circ \gamma(t)|}$$

Then the image of  $\alpha$  lies in the unit circle. Moreover, a straightforward argument shows that  $(v \circ \gamma)^* \omega = \alpha^* \omega$ , and so

$$\int_{v(\gamma)} \omega = \int_{\alpha} \omega.$$

<sup>&</sup>lt;sup>2</sup>To be more precise,  $\gamma$  is an oriented curve—that is, an equivalence class of  $C^1$  paths where two paths  $p: [a, b] \to \mathbb{R}^2$  and  $p': [a', b'] \to \mathbb{R}^2$  are equivalent if and only if there exists a strictly increasing  $C^1$  map  $\phi: [a, b] \to [a', b']$  such that  $p = p' \circ \phi$ . For simplicity, we will not explicitly state this again in the rest of the paper.

Define the map p from  $\mathbb{R}$  to the unit circle by  $p(t) = (\cos(t), \sin(t))$ . A straightforward calculation shows that  $p^*(\omega) = dt$ . Since [a, b] is contractible, the path lifting lemma implies that there exists a path  $\tilde{\alpha} : [a, b] \to \mathbb{R}$  so that  $\alpha(t) = p \circ \tilde{\alpha}(t)$ . Therefore,

$$\int_{\alpha} \omega = \int_{\widetilde{\alpha}} p^*(\omega) = \int_{\widetilde{\alpha}} dt = \int_{a}^{b} \widetilde{\alpha}'(s) \, ds = \widetilde{\alpha}(b) - \widetilde{\alpha}(a)$$

Since  $\alpha$  is a closed curve, we have

$$p \circ \widetilde{\alpha}(b) = \alpha(b) = \alpha(a) = p \circ \widetilde{\alpha}(a)$$

In other words,  $\cos(\tilde{\alpha}(b)) = \cos(\tilde{\alpha}(a))$  and  $\sin(\tilde{\alpha}(b)) = \cos(\tilde{\alpha}(a))$ . Therefore,  $\tilde{\alpha}(b) - \tilde{\alpha}(a)$  is a multiple of  $2\pi$  and so the integral  $\int_{\alpha} \omega$  is a multiple of  $2\pi$ .

The quantity  $(2\pi)^{-1} \int_{\alpha} \omega$  in the proof of Proposition 2.4 is called the *winding number* of the curve  $\alpha$ . Geometrically, it is the number of times that the curve  $\alpha$  winds around the origin in the counter-clockwise direction. In this way, we may regard the index of **v** as the number of times the curve  $v \circ \gamma$  winds around the origin.

We provide some examples of computing the index from this perspective: see Figures 1 and 2 below.



Figure 1: Contour plot of  $f(x, y) = x^2 + y^2$  generated by Mathematica. The origin is a minimum. We normalize and then translate the vectors of the gradient vector field that lie on the curve  $\gamma$  to a separate picture so that their tails line up. In this case, the vector field makes exactly one turn around the origin in the counterclockwise direction and so the index equals +1.



Figure 2: Contour plot of  $f(x, y) = x^2 - y^2$  generated by Mathematica. The origin is a saddle point. We normalize and then translate the vectors of the gradient vector field that lie on the curve  $\gamma$  to a separate picture so that their tails line up. In this case, the vector field makes exactly one turn around the origin, but in the clockwise direction, and so the index equals -1.

As a final example, let  $f : \mathbb{R}^2 \to \mathbb{R}$  be defined by  $f(x, y) = x^3 - y^2$ . At the origin, this function has a degenerate critical point. We can compute the index of this critical point using (9) and find that the index is equal to 0. The following figure shows the level sets of f near the origin.



Figure 3: Contour plot of  $f(x, y) = x^3 - y^2$  generated by Mathematica. The origin is an index-0 critical point. We observe that it has a "cusp-like" feature.

The following lemma is well-known and establishes a connection between the index<sup>3</sup> of a critical point and its degeneracy:

**Lemma 2.5** (Morse Lemma). Consider a smooth function  $f : \mathbb{R}^2 \to \mathbb{R}$  with nondegenerate critical point (0,0). Then there exist neighborhoods U and U' of (0,0) and a diffeomorphism  $\phi : U \to U'$  so that for each  $(x,y) \in U$  we have

$$(f \circ \phi)(x, y) = x^2 \pm y^2.$$

A consequence of the Morse lemma is that if a critical point is nondegenerate, then it has either index +1 (as in Figure 1) or index -1 (as in Figure 2). This implies that if a critical point has index 0, then it is nondegenerate.

Now we present some useful properties of the index:

**Lemma 2.6.** Let  $\gamma$  be a simple closed curve in  $\mathbb{R}^2$  that admits a piecewise smooth parameterization, and let S be the bounded component<sup>4</sup> of the complement of  $\gamma$ . Let  $\mathbf{v} : S \cup \gamma \to \mathbb{R}^2$  be a vector field that has no zeros. Then the index of v around  $\gamma$  equals 0.

*Proof.* Recall that for  $\mathbf{v} = (v_1, v_2)$ , the index is given by

$$\operatorname{ind}(\mathbf{v},\gamma) = \frac{1}{2\pi} \int_{\gamma} \frac{v_1 dv_2 - v_2 dv_1}{(v_1)^2 + (v_2)^2}.$$
(2)

Because there are no zeros of **v** on  $S \cup \gamma$ , we can apply Stokes' theorem to (2):

$$ind(\mathbf{v},\gamma) = \frac{1}{2\pi} \int_{\gamma} \frac{v_1 dv_2 - v_2 dv_1}{v_1^2 + v_2^2} = \frac{1}{2\pi} \int_{S} d\left(\frac{v_1 dv_2 - v_2 dv_1}{v_1^2 + v_2^2}\right) = \frac{1}{2\pi} \int_{S} d\left(\frac{v_1 dv_2}{v_1^2 + v_2^2}\right) - d\left(\frac{v_2 dv_1}{v_1^2 + v_2^2}\right).$$
(3)

Since **v** has no zeros on  $\gamma$ , (3) is well-defined.

<sup>&</sup>lt;sup>3</sup>Note that the Poincarè-Hopf index, with which we are concerned, is not the same as the Morse index. The Poincarè-Hopf index gives information about a vector field's zeros, while the Morse index gives information about the appearance of a function around a critical point.

<sup>&</sup>lt;sup>4</sup>This is well-defined by the Jordan curve theorem.

Let f, g and h denote differential 0-forms and let  $\omega$  and  $\alpha$  denote differential 1-forms. We make use of the following properties of differential forms to evaluate (3):

$$\begin{array}{rcl} d(fg) &=& gdf + fdg\\ d(f\omega) &=& df \wedge \omega\\ d(h \circ f) &=& h'(f)df\\ d(f^2) &=& 2fdf\\ \omega \wedge \alpha &=& -\alpha \wedge \omega\\ \omega \wedge \omega &=& 0. \end{array}$$

First, we observe that

$$d\left(\frac{v_1}{v_1^2 + v_2^2}\right) = d(v_1(v_1^2 + v_2^2)^{-1})$$
  
=  $(v_1^2 + v_2^2)^{-1}dv_1 + v_1d(v_1^2 + v_2^2)^{-1}$   
=  $(v_1^2 + v_2^2)^{-1}dv_1 - v_1(v_1^2 + v_2^2)^{-2}d(v_1^2 + v_2^2)$   
=  $(v_1^2 + v_2^2)^{-1}dv_1 - v_1(v_1^2 + v_2^2)^{-2}(2v_1dv_1 + 2v_2dv_2).$  (4)

By (4), we have

$$d\left(\frac{v_1}{v_1^2 + v_2^2} dv_2\right) = d\left(\frac{v_1}{v_1^2 + v_2^2}\right) \wedge dv_2$$
  
=  $((v_1^2 + v_2^2)^{-1} dv_1 - v_1(v_1^2 + v_2^2)^{-2} (2v_1 dv_1 + 2v_2 dv_2)) \wedge dv_2,$   
=  $(v_1^2 + v_2^2)^{-1} dv_1 \wedge dv_2 - v_1(v_1^2 + v_2^2)^{-2} (2v_1 dv_1 + 2v_2 dv_2) \wedge dv_2$   
=  $\frac{dv_1 \wedge dv_2}{v_1^2 + v_2^2} - \frac{2v_1^2 dv_1 \wedge dv_2}{(v_1^2 + v_2^2)^2}.$  (5)

Then by a symmetric argument, we also have

$$d\left(\frac{v_2}{v_1^2 + v_2^2} dv_1\right) = \frac{dv_2 \wedge dv_1}{v_2^2 + v_1^2} - \frac{2v_2^2 dv_2 \wedge dv_1}{(v_2^2 + v_2^2)^2}.$$
(6)

Using (5) and (6), (3) becomes

$$\begin{aligned} \operatorname{ind}(\mathbf{v},\gamma) &= \frac{1}{2\pi} \int_{S} \frac{dv_{1} \wedge dv_{2}}{v_{1}^{2} + v_{2}^{2}} - \frac{2v_{1}^{2}dv_{1} \wedge dv_{2}}{(v_{1}^{2} + v_{2}^{2})^{2}} - \frac{dv_{2} \wedge dv_{1}}{v_{2}^{2} + v_{1}^{2}} + \frac{2v_{2}^{2}dv_{2} \wedge dv_{1}}{(v_{2}^{2} + v_{2}^{2})^{2}} \\ &= \frac{1}{2\pi} \int_{S} \left( \frac{1}{v_{1}^{2} + v_{2}^{2}} - \frac{2v_{1}^{2}}{(v_{1}^{2} + v_{2}^{2})^{2}} + \frac{1}{v_{2}^{2} + v_{1}^{2}} - \frac{2v_{2}^{2}}{(v_{2}^{2} + v_{2}^{2})^{2}} \right) dv_{1} \wedge dv_{2} \\ &= \frac{1}{2\pi} \int_{S} \left( \frac{2}{v_{1}^{2} + v_{2}^{2}} - \frac{2}{v_{1}^{2} + v_{2}^{2}} \right) dv_{1} \wedge dv_{2} \\ &= \frac{1}{2\pi} \int_{S} 0 \ dv_{1} \wedge dv_{2} \\ &= 0. \end{aligned}$$

**Lemma 2.7.** Let  $\mathbf{v} : U \subset \mathbb{R}^3 \to \mathbb{R}^2$  be a continuously differentiable vector field, and define for each t the vector field  $\mathbf{v}^t(x,y) = \mathbf{v}(x,y,t)$ . Then given a simple closed curve  $\gamma \subset \mathbb{R}^2$ , the function  $t \mapsto \operatorname{ind}(\mathbf{v}^t,\gamma)$  is continuous.

*Proof.* Let  $\mathbf{v}^t = (v_1^t, v_2^t)$ , and recall that the index is given by

$$\operatorname{ind}(\mathbf{v}^{t},\gamma) = \frac{1}{2\pi} \int_{\gamma} \frac{v_{1}^{t} dv_{2}^{t} - v_{2}^{t} dv_{1}^{t}}{(v_{1}^{t})^{2} + (v_{2}^{t})^{2}}.$$
(2)

Abusing notation slightly, we will use  $v_1$  to denote  $v_1^t$  and  $v_2$  to denote  $v_2^t$ . We observe that

$$dv_2 = \frac{\partial v_2}{\partial x}dx + \frac{\partial v_2}{\partial y}dy = (v_2)_x dx + (v_2)_y dy,\tag{7}$$

$$dv_1 = \frac{\partial v_1}{\partial x}dx + \frac{\partial v_1}{\partial y}dy = (v_1)_x dx + (v_1)_y dy.$$
(8)

Using (7) and (8), (2) becomes

$$\operatorname{ind}(\mathbf{v}^{t},\gamma) = \frac{1}{2\pi} \int_{\gamma} \frac{v_{1}(v_{2})_{x} - v_{2}(v_{1})_{x}}{v_{1}^{2} + v_{2}^{2}} dx + \frac{v_{1}(v_{2})_{y} - v_{2}(v_{1})_{y}}{v_{1}^{2} + v_{2}^{2}} dy.$$
(9)

There exists a continuous parameterization  $\gamma^* \colon [a, b] \to \mathbb{R}^2$  of  $\gamma$  and a continuous function  $H \colon [a, b] \times [c, d] \to \mathbb{R}$  such that (9) is equivalent to  $\int_a^b H(s,t)ds$ . By Lemma 2.8,  $t \mapsto \operatorname{ind}(\mathbf{v}^t,\gamma)$  is continuous. 

**Lemma 2.8.** Suppose that a function  $H: [a, b] \times [c, d] \to \mathbb{R}$  is continuous, and define  $h: [c, d] \to \mathbb{R}^2$  by  $h(t) = \int_a^b H(s, t) ds$ . Then h is continuous.

*Proof.* We want to show that for all  $\epsilon > 0$  and for all  $t \in [c, d]$ , there exists a  $\delta > 0$  such that if  $|t - t'| < \delta$ , then

 $\begin{aligned} |\int_{a}^{b} H(s,t)ds - \int_{a}^{b} H(s,t')ds| &< \epsilon. \\ \text{We observe that } H \text{ is a continuous function on the compact set } [a,b] \times [c,d], \text{ so it is uniformly continuous. It follows that for all } \frac{\epsilon}{b-a} > 0, \text{ there exists a } \delta > 0 \text{ such that for all } s \in [a,b] \text{ and } t \in [c,d], \text{ if } |t-t'| < \delta \text{ then } |H(s,t)-H(s,t')| < \frac{\epsilon}{b-a}. \end{aligned}$ Then we observe that

$$\left| \int_{a}^{b} H(s,t)ds - \int_{a}^{b} H(s,t')ds \right| = \left| \int_{a}^{b} [H(s,t) - H(s,t')]ds \right|$$
$$\leq \int_{a}^{b} |H(s,t) - H(s,t')|ds$$
$$< \int_{a}^{b} \frac{\epsilon}{b-a}ds$$
$$= \epsilon,$$

which is the condition for uniform continuity of h. The continuity of h follows.

We will say that a zero (x, y) of a vector field is *isolated* if there exists a neighborhood of (x, y) that contains no other zeros. The zeros of a gradient vector field  $\nabla f$  are exactly the critical points of f.

**Proposition 2.9.** Let  $\Omega \subset \mathbb{R}^2$  be an open set, let  $\varphi : \Omega \to \mathbb{R}$  be an eigenfunction of the Laplacian, and let  $\gamma$  be a simple closed curve in  $\mathbb{R}^2$ . Let S be the component of  $\mathbb{R}^2 - \gamma$  that is bounded. If the index of  $\nabla \varphi$  around  $\gamma$  equals zero, then either S contains no critical points or S contains at least one isolated critical point.

*Proof.* It follows from Proposition 2.5 in [Jdg-Mnd19] that the set C of critical points of  $\varphi$  is a disjoint union of finitely many simple closed curves and finitely many isolated points. Suppose that  $\alpha$  is a simple closed curve that is a connected component of C.

Because  $\nabla \varphi$  vanishes along  $\alpha$ , the restriction of  $\varphi$  to  $\alpha$  is constant c. As in the proof of Proposition 2.5 in [Jdg-Mnd19], because  $\varphi$  is an eigenfunction of the Laplacian, one finds that  $c \neq 0$  and the Hessian of  $\varphi$  does not vanish along  $\alpha$ . Using the inverse function theorem and the compactness of  $\alpha$ , one finds a neighborhood U of  $\alpha$ , a number  $\epsilon > 0$ , and a diffeomorphism  $F: \mathbb{R}/\mathbb{Z} \times (c - \epsilon, c + \epsilon) \to U$  so that for each t the function  $s \mapsto \varphi \circ F(s, t)$  is constant and for each s the only critical point of  $t \mapsto \varphi \circ F(s, t)$  occurs at t = c.

In particular, each critical point of  $\varphi$  that lies in U also lies in  $\alpha$ . If  $\beta$  is a simple closed curve that is a component of a level set of a function and contains no critical points, then the index of the gradient vector field around  $\beta$  equals 1. In particular, for each  $t \neq 0$ , the index of  $\nabla \varphi$  about the curve  $\beta_t = F(\mathbb{R}/\mathbb{Z} \times \{t\})$  equals 1. If C were to contain no isolated points but were to contain simple closed curve components, then one could homotope  $\gamma$ —without passing through critical points—to a concatenation of simple closed curves  $\beta_t$  about the simple closed curve components of C together with arcs  $\eta_i$  joining the  $\beta_t$ . Using the additivity of the integral, and the cancellation of the integrals over the various  $\eta_i$ , one would find that the index around  $\gamma$  would be positive.

But this contradicts our assumption, and so C is either empty or contains at least one isolated critical point. 

_	_	-	-

We will also cite the following well-known theorems to prove our result:

**Theorem 2.10** (Intermediate Value Theorem). Let  $F: [a, b] \to \mathbb{R}$  be continuous. Then for each  $u \in f([a, b])$ , there exists a number  $c \in (a, b)$  such that f(c) = u.

**Theorem 2.11** (Implicit Function Theorem). Consider a continuously differentiable function  $F \colon \mathbb{R}^2 \to \mathbb{R}$  and a point  $(x_0, y_0) \in \mathbb{R}^2$  such that  $F(x_0, y_0) = c$ . If  $F_y(x_0, y_0) \neq 0$ , then there exists a neighborhood  $U \times V \subset \mathbb{R}^2$  of  $(x_0, y_0)$  and a unique continuous function  $g \colon U \to V$  such that F(x, g(x)) = c for all  $x \in U$ .

#### **3** Proof of Main Result

Here we prove our main result, that there exists an eigenfunction on the square that has an index-0 degenerate critical point. More specifically, we will show that the there exists a  $t \in (0, \frac{\pi}{16})$  such that function  $f^t$  defined by

 $f^{t}(x,y) = 8\cos(t)\sin(x)\sin(8y) + 7\sin(t)\sin(4x)\sin(7y)$ 

has an index-0 critical point that lies in  $(\frac{5\pi}{8}, \frac{7\pi}{8}) \times (\frac{13\pi}{14}, \pi)$ . For the convenience of the reader, we define the following constants and sets:

- $x_{\min} = \frac{5\pi}{8}, x_{\max} = \frac{7\pi}{8}$
- $y_{\min} = \frac{13\pi}{14}, \ y_{\max} = \pi$
- $t_{\min} = 0, t_{\max} = \frac{\pi}{16}$
- $I = (x_{\min}, x_{\max}), J = (y_{\min}, y_{\max})$
- $B = I \times J, \ \partial B = \overline{B} B$

(In this paper, given  $X \subset \mathbb{R}^2$ ,  $\overline{X}$  denotes the closure of X in  $\mathbb{R}^2$ .) We begin by proving the following three lemmas:

**Lemma 3.1.** There are no critical points of  $f^{t_{\min}}$  in B.

*Proof.* Recall that

$$f^{t_{\min}}(x,y) = f^0(x,y) = \cos(0)\sin(x)\sin(8y) + \sin(0)\sin(4x)\sin(7y)$$
  
= sin(x) sin(8y).

We want to check that for all  $(x, y) \in B$ ,  $f_x^{t_{\min}}(x, y)$  and  $f_y^{t_{\min}}(x, y)$  are not both 0; it suffices to show that  $f_x^{t_{\min}}(x, y) \neq 0$ . Observe that

$$f_x^{t_{\min}}(x,y) = \cos(x)\sin(8y),$$

which is 0 if and only if  $\cos(x) = 0$  or  $\sin(8y) = 0$ . This occurs at  $x = \frac{\pi}{2} + k\pi$  and  $y = \frac{k'\pi}{8}$   $(k, k' \in \mathbb{Z})$  respectively. By inspection, there are no such  $x \in I$  or  $y \in J$ , so  $f_x^{t_{\min}}(x, y) \neq 0$  for all  $(x, y) \in B$ . It follows that there are no critical points of  $f^{t_{\min}}$  in B.

**Lemma 3.2.** For all  $t \in [t_{\min}, t_{\max}]$ , there are no critical points of  $f^t$  on  $\partial B$ .

*Proof.* Let  $t \in [t_{\min}, t_{\max}]$ , and define the following:

- $\partial B_{\text{bottom}} = \overline{I} \times \{y_{\min}\}$
- $\partial B_{\text{top}} = \overline{I} \times \{y_{\text{max}}\}$
- $\partial B_{\text{left}} = \{x_{\min}\} \times J$
- $\partial B_{\text{right}} = \{x_{\text{max}}\} \times J$

We begin by showing that  $f_y^t(x,y) \neq 0$  for all  $(x,y) \in \partial B_{\text{bottom}} \cup \partial B_{\text{top}}$  and  $f_x^t(x,y) \neq 0$  for all  $(x,y) \in \partial B_{\text{left}} \cup \partial B_{\text{right}}$ .

 $f_y^t \neq 0$  on  $\partial B_{\text{bottom}}$ : Let  $x \in \overline{I}$ ; we claim that  $f_y^t(x, y_{\min}) \neq 0$ . Recall that

$$\begin{aligned} f_y^t(x, y_{\min}) &= f_y^t(x, \frac{13\pi}{14}) = 8\cos\left(t\right)\sin\left(x\right)\cos\left(8\frac{13\pi}{14}\right) + 7\sin\left(t\right)\sin\left(4x\right)\cos\left(7\frac{13\pi}{14}\right) \\ &= 8\cos\left(t\right)\sin\left(x\right)\cos\left(8\frac{13\pi}{14}\right). \end{aligned}$$

Observe that  $\cos(t) > 0$  on  $[t_{\min}, t_{\max}]$ , that  $\sin(x) > 0$  on  $\overline{I}$ , and that  $\cos\left(8\frac{13\pi}{14}\right) = \cos\left(\frac{10\pi}{7}\right) < 0$ . Then  $f_u^t(x, y_{\min}) < 0$ ; the claim follows.

 $f_y^t \neq 0$  on  $\partial B_{\text{top}}$ : Again let  $x \in \overline{I}$ ; now we show that  $f_y^t(x, y_{\text{max}}) > 0$ . Recall that

$$f_y^t(x, y_{\max}) = f_y^t(x, \pi) = 8\cos(t)\sin(x)\cos(8\pi) + 7\sin(t)\sin(4x)\cos(7\pi)$$
  
= 8\cos(t)\sin(x) - 7\sin(t)\sin(4x).

It suffices to show that

$$8\cos(t)\sin(x) > 7\sin(t)\sin(4x). \tag{10}$$

Since  $\cos(t)$  and  $\sin(x)$  are positive and strictly decreasing on  $[t_{\min}, t_{\max}]$  and  $\overline{I}$  respectively, we have that  $8\cos(t)\sin(x) \ge 1$  $8\cos(t_{\max})\sin(x_{\max}) = 8\cos(\frac{\pi}{16})\sin(\frac{7\pi}{8})$ . Similarly, since  $\sin(t)$  is positive and strictly increasing on  $[t_{\min}, t_{\max}]$  and  $\sin(4x) \leq 1$  everywhere, we have that  $7\sin(t)\sin(4x) \leq 7\sin(\frac{\pi}{16})$ .

We observe that

$$8\cos(\frac{\pi}{16})\sin(\frac{7\pi}{8}) > 7\sin(\frac{\pi}{16})$$

(all such inequalities can be verified using half-angle and sum formulas for sine and cosine); then (10) holds.

 $f_x^t \neq 0$  on  $\partial B_{\text{left}}$ : Let  $y \in J$ ; we show that  $f_x^t(x_{\min}, y) > 0$ . Recall that

$$f_x^t(x_{\min}, y) = f_x^t(\frac{5\pi}{8}, y) = \cos(t)\cos(\frac{5\pi}{8})\sin(8y) + 4\sin(t)\cos(4\frac{5\pi}{8})\sin(7y)$$
$$= \cos(t)\cos(\frac{5\pi}{8})\sin(8y).$$

Observe that  $\cos(t) > 0$  on  $[t_{\min}, t_{\max}]$ , that  $\sin(8y) < 0$  on J, and that  $\cos(\frac{5\pi}{8}) < 0$ . Then  $f_x^t(x_{\min}, y) > 0$ .

 $f_x^t \neq 0 \text{ on } \partial B_{\text{right}}$ : Again let  $y \in J$ ; now we show that  $f_x^t(x_{\max}, y) > 0$ . Recall that

$$f_x^t(x_{\max}, y) = f_x^t(\frac{7\pi}{8}, y) = \cos(t)\cos(\frac{7\pi}{8})\sin(8y) + 4\sin(t)\cos(4\frac{7\pi}{8})\sin(7y)$$
$$= \cos(t)\cos(\frac{7\pi}{8})\sin(8y).$$

Observe that  $\cos(t) > 0$  on  $[t_{\min}, t_{\max}]$ , that  $\sin(8y) < 0$  on J, and that  $\cos(\frac{7\pi}{8}) < 0$ ; then  $f_x^t(x_{\max}, y) > 0$ .

Because  $\partial B = \partial B_{\text{bottom}} \cup \partial B_{\text{top}} \cup \partial B_{\text{left}} \cup \partial B_{\text{right}}$ , for all  $(x, y) \in \partial B$  we have that  $f_x^t(x, y) \neq 0$  or  $f_y^t(x, y) \neq 0$ . Then for all  $t \in [t_{\min}, t_{\max}]$ ,  $f^t$  has no critical points on  $\partial B$ . 

**Lemma 3.3.** There exists at least one critical point of  $f^{t_{\text{max}}}$  in B.

*Proof.* Let  $x^* = \frac{23\pi}{32}$  and  $y^* = \frac{121\pi}{128}$ . In this proof, we abuse notation slightly and refer to  $f^{t_{\max}}$  as simply f. We begin by showing that for all  $x_0 \in (x_{\min}, x_{\max})$ , there exists exactly one  $y_0 \in (y_{\min}, y^*)$  such that  $f_y(x_0, y_0) = 0$ . First, recall from the proof of Lemma 3.2 that  $f_y < 0$  on  $\overline{I} \times \{y_{\min}\}$ . Now we show that  $f_y > 0$  on  $\overline{I} \times \{y^*\} = 0$  $[x_{\min}, x_{\max}] \times \frac{121\pi}{128}.$ 

Recall that  $f_y(x,y) = 8\cos(t_{\max})\sin(x)\cos(8y) + 7\sin(t_{\max})\sin(4x)\cos(7y)$ . It suffices to show that

$$8\cos(\frac{\pi}{16})\sin(x)\cos(8\frac{121\pi}{128}) > 7\sin(\frac{\pi}{16})\sin(4x)(-\cos(7\frac{121\pi}{128})).$$
(11)

for all  $x \in \overline{I}$ .

We observe that  $\sin(x) \ge \sin(x_{\max}) = \sin(\frac{5\pi}{8})$  and  $\sin(4x) \le 1$  on  $\overline{I}$ . It follows that  $8\cos(\frac{\pi}{16})\sin(x)\cos(8\frac{121\pi}{128}) \ge 8\cos(\frac{\pi}{16})\sin(\frac{5\pi}{8})\cos(8\frac{121\pi}{128})$  and  $7\sin(\frac{\pi}{16})\sin(4x)(-\cos(7\frac{121\pi}{128})) \le 7\sin(\frac{\pi}{16})(-\cos(7\frac{121\pi}{128}))$ , so (11) holds.

For each  $x_0 \in \overline{I}$ ,  $f_y(x_0, y_{\min}) < 0$  and  $f_y(x_0, y^*) > 0$ , so we can apply the Intermediate Value Theorem to  $y \mapsto f_y(x_0, y)$  to get that  $f_y(x_0, y_0) = 0$  for some  $y_0 \in (y_{\min}, y^*)$ .

To verify that there is only one such  $y_0$  for each  $x_0$ , it suffices to show that  $y \mapsto f_y(x_0, y)$  is strictly increasing on  $(y_{\min}, y^*)$ . Equivalently, we will show that  $f_{yy}(x, y) > 0$  for all  $(x, y) \in \overline{I} \times (y_{\min}, y^*)$ .

Observe that

$$f_{yy}(x,y) = -64\cos(t_{\max})\sin(x)\sin(8y) - 49\sin(t_{\max})\sin(4x)\sin(7y).$$

Using the same approach as before, we show that

$$64\cos(\frac{\pi}{16})\sin(x)(-\sin(8y)) > 49\sin(\frac{\pi}{16})\sin(4x)\sin(7y).$$
(12)

Recall that  $\sin(x) \ge \sin(\frac{5\pi}{8})$  and  $\sin(4x) \le 1$  on  $\overline{I}$ , and observe that  $-\sin(8y) > -\sin(8y_{\min}) = -\sin(8\frac{13\pi}{14})$  and  $\sin(7y) < 1$  on  $(y_{\min}, y^*)$ . It follows that (12) holds on  $\overline{I} \times (y_{\min}, y^*)$ .

Because  $f_{yy} > 0$  on  $\overline{I} \times (y_{\min}, y^*)$ , it follows that  $f_{yy} \neq 0$  on  $\overline{I} \times (y_{\min}, y^*)$ . Then the Implicit Function Theorem can be applied to each  $(x_0, y_0) \in \overline{I} \times (y_{\min}, y^*)$  such that  $f_y(x_0, y_0) = 0$ .

Recall that for each  $x_0 \in \overline{I}$ , there exists exactly one  $y_0 \in (y_{\min}, y^*)$  such that  $f_y(x_0, y_0) = 0$ . Then by the Implicit Function Theorem, for each  $x_0$  there exists a neighborhood  $U_{x_0} \subset \mathbb{R}$  of  $x_0$  and a unique continuous function  $g_{x_0}: U_{x_0} \to (y_{\min}, y^*)$  such that  $f_y(x, g_{x_0}(x)) = 0$  for all  $x \in U_{x_0}$ .

By the uniqueness of  $g_{x_0}$ , we have for all  $x_0, x'_0 \in \overline{I}$  that whenever  $U_{x_0} \cap U_{x'_0} \neq \emptyset$ ,  $g_{x_0}(x) = g_{x'_0}(x)$  for all  $x \in U_{x_0} \cap U_{x'_0}$ . Then because the  $U_{x_0}$  cover  $\overline{I}$ , there exists a unique continuous  $G: \overline{I} \to (y_{\min}, y^*)$ , defined piecewise on each of the  $U_{x_0}$ , such that  $f_y(x, G(x)) = 0$  for all  $x \in \overline{I}$ . (In fact, because  $\overline{I}$  is compact and the  $U_{x_0}$  form an open cover, we can define G using only finitely many  $U_{x_0}$ .) In other words, on  $\overline{I} \times (y_{\min}, y^*)$ ,  $f_y(x, y) = 0$  if and only if y = G(x).

Finally, let  $h: \overline{I} \to \mathbb{R}$  be such that  $h(x) = f_x(x, G(x))$ . Observe that h is continuous by composition; we will prove the existence of a critical point inside B by applying the Intermediate Value Theorem to h. First we show that  $h(x^*) = f_x(x^*, G(x^*)) < 0$ :

Recall that  $G(x^*) \in (y_{\min}, y^*)$ . Therefore it suffices to show that

$$f_x(x,y) = \cos(t_{\max})\cos(x)\sin(8y) + 4\sin(t_{\max})\cos(4x)\sin(7y) < 0$$

for all  $(x, y) \in \{x^*\} \times (y_{\min}, y^*)$ , or equivalently that

$$\cos(\frac{\pi}{16})\cos(\frac{23\pi}{32})\sin(8y) < -4\sin(\frac{\pi}{16})\cos(4\frac{23\pi}{32})\sin(7y) \tag{13}$$

for all  $y \in (y_{\min}, y^*)$ .

We observe that  $\sin(8y) \ge -1$  and  $\sin(7y) > \sin(7y^*) = \sin(7\frac{121\pi}{128})$  on  $(y_{\min}, y^*)$ . Then since  $\cos(\frac{\pi}{16})\cos(\frac{23\pi}{32})\sin(8y) \le -\cos(\frac{\pi}{16})\cos(\frac{23\pi}{32})$  and  $-4\sin(\frac{\pi}{16})\cos(4\frac{23\pi}{32})\sin(7y) > -4\sin(\frac{\pi}{16})\cos(4\frac{23\pi}{32})\sin(7\frac{121\pi}{128})$ , (13) holds on  $(y_{\min}, y^*)$ . It follows that  $f_x(x^*, G(x^*)) = h(x^*) < 0$ . Next, we observe that  $f_x(x_{\min}, G(x_{\min})) = h(x_{\min}) > 0$ : recall from the

It follows that  $f_x(x^*, G(x^*)) = h(x^*) < 0$ . Next, we observe that  $f_x(x_{\min}, G(x_{\min})) = h(x_{\min}) > 0$ : recall from the proof of Lemma 3.2 that  $f_x(x, y) > 0$  on  $\{x_{\min}\} \times J$ . Because  $(x_{\min}, G(x_{\min})) \in \{x_{\min}\} \times (y_{\min}, y^*) \subset \{x_{\min}\} \times J$ , we have that  $f_x(x_{\min}, G(x_{\min})) > 0$ .

Now we can apply the Intermediate Value Theorem to h(x) on  $[x_{\min}, x^*]$  to conclude that there exists some  $x_0 \in (x_{\min}, x^*)$  such that  $h(x_0) = f_x(x_0, G(x_0)) = 0$ . By construction,  $f_y(x_0, G(x_0)) = 0$ , so  $(x_0, G(x_0))$  is a critical point of f in B.

(There must in fact be at least two distinct critical points of  $f^{t_{\max}}$  in *B* because an analogous argument applies to  $[x^*, x_{\max}]$ : recall from the proof of Lemma 3.2 that  $f_x^{t_{\max}}(x, y) > 0$  on  $\{x_{\max}\} \times J$  as well, so  $h(x_{\max}) > 0$ . Then apply the Intermediate Value Theorem to h on  $[x^*, x_{\max}]$  to find a critical point  $(x'_0, G(x'_0))$  for some  $x'_0 \in (x^*, x_{\max})$ .)

Now we proceed to our main result, as stated in Section 1:

**Theorem 1.1.** There exists a Dirichlet function of the Laplacian on  $[0, \pi] \times [0, \pi]$  with an index-0 critical point in  $(0, \pi) \times (0, \pi)$ .

*Proof.* Let

 $\tilde{t} = \sup\{t \in [t_{\min}, t_{\max}] | f^t \text{ has no critical points in } B\}.$ 

It follows from Lemma 3.1 that  $\tilde{t} \in (t_{\min}, t_{\max}]$ .

We claim that  $f^{\tilde{t}}$  has at least one critical point in B. To show this, suppose not: then  $f^{\tilde{t}}$  has no critical points in B. By definition of the supremum, for all  $\epsilon > 0$ , there exists a  $t \in (\tilde{t}, \tilde{t} + \epsilon)$  such that  $f^t$  has a critical point in B. In particular, we can set  $\epsilon = \frac{1}{n}$  for each positive integer n to produce a sequence  $\{t_n\}$  converging to  $\tilde{t}$ , where  $f^{t_n}$  has a critical point  $(x_n, y_n) \in B$  for all n.

Now we observe that  $\{(x_n, y_n)\}$  is a sequence in the compact set  $\overline{B}$ , so it has a convergent subsequence  $\{(x_{n_k}, y_{n_k})\}$ ; say that  $\{(x_{n_k}, y_{n_k})\}$  converges to some  $(\tilde{x}, \tilde{y}) \in \overline{B}$ . It follows that the corresponding subsequence  $\{t_{n_k}\}$  also converges to  $\tilde{t}$ , so  $f^{t_{n_k}} \to f^{\tilde{t}}$ .

Because the function  $(x, y, t) \mapsto f^t(x, y)$  is smooth, the function

$$(x, y, t) \mapsto \nabla f^t(x, y)$$

is continuous. It follows that  $\nabla f^{t_{n_k}}(x_{n_k}, y_{n_k})$  converges to  $\nabla f^{\tilde{t}}(\tilde{x}, \tilde{y})$ . But each of the  $(x_{n_k}, y_{n_k})$  is a critical point of  $f^{t_{n_k}}$ , so  $\nabla f^{t_{n_k}}(x_{n_k}, y_{n_k}) = (0, 0)$  for all  $n_k$ . Then  $\nabla f^{\tilde{t}}(\tilde{x}, \tilde{y}) = (0, 0)$ .

Now we have that  $(\tilde{x}, \tilde{y}) \in \overline{B}$  is a critical point of  $f^{\tilde{t}}$ . In fact, by Lemma 3.2, we have that  $(\tilde{x}, \tilde{y}) \in B$ , as claimed.

By Lemmas 2.6 and 3.1,  $\operatorname{ind}(\nabla f^{t_{\min}}, \partial B) = 0$ . Furthermore, by Proposition 2.4 and Lemmas 2.7 and 3.2,  $t \mapsto \operatorname{ind}(\nabla f^t, \partial B)$  is integer-valued and continuous on  $[t_{\min}, t_{\max}]$ , so it is constant. Therefore  $\operatorname{ind}(\nabla f^{\tilde{t}}, \partial B) = 0$ . Then by Proposition 2.9, B contains either no critical points or at least one isolated critical point of  $f^{\tilde{t}}$ . But since  $(\tilde{x}, \tilde{y}) \in B$ , we can say that B contains at least one isolated critical point  $(\tilde{x}', \tilde{y}')$  of  $f^{\tilde{t}}$ . Then there exists a neighborhood U of  $(\tilde{x}', \tilde{y}')$  containing no other critical points of  $f^{\tilde{t}}$ .

Now choose any simple closed curve  $\gamma \subset U$  such that the bounded component of  $\mathbb{R}^2 - \gamma$ , S, contains  $(\tilde{x}', \tilde{y}')$ , and let  $\iota: [t_{\min}, \tilde{t}] \to \mathbb{R}$  be defined by

$$\iota(t) = \operatorname{ind}(\nabla f^t, \gamma).$$

(We restrict t to the interval  $[t, \tilde{t}]$  to ensure that  $f^t$  has no critical points on  $\gamma$ ; otherwise,  $\iota(t)$  might not be well-defined.)

Observe that  $\iota$  is constant. In particular, by Lemma 2.6,  $\iota(t_{\min}) = 0$ , so  $\iota(\tilde{t}) = 0$  as well. Then since  $(\tilde{x}', \tilde{y}')$  is the only critical point of  $f^{\tilde{t}}$  in S, we have that the index of  $(\tilde{x}', \tilde{y}')$  is 0.

#### Acknowledgments

This project was supported by the REU program run by the Department of Mathematics at Indiana University Bloomington. We thank our advisor, Chris Judge, for overseeing the project and for his many insightful suggestions on the paper. We also thank Chris Connell for directing the REU program and for his help with Mathematica.

#### References

[Crn-Hlb] Courant, R. and Hilbert, D. Methods of Mathematical Physics. Vol. I. Interscience Publishers, Inc., New York, N.Y., 1953. xv+561 pp.

[Jdg-Mnd19] Judge, C. and Mondal, S. Euclidean Triangles Have No Hot Spots. To appear in Ann. of Math.

[Ulb76] Uhlenbeck, K. <u>Generic Properties of Eigenfunctions</u>. American Journal of Mathematics, Vol. 98, No. 4 (Winter, 1976), pp. 1059-1078.

#### EXTREMAL LENGTH OF WINDING CURVES

ABSTRACT. For a given regular polygonal tesselation of the plane, there are interesting collections of curves and weighted multicurves which which minimize area for a fixed lower bound of the lengths of curves under a flat metric. Curves around two of the punctures of the twice-punctured sphere with winding numbers 2 and -1 are shown to have the same area-minimizing metric as the hexagon with straight lines moving edge to opposite edge and opposite edges glued together. For a fixed lower bound of the lengths of curves in any given hexagonally punctured hexagonal torus, it is found that the minimal area under these constraints is approximated very closely by a quadratic equation related to the size of the puncture. This extremal metric also engenders regions of positive and flat curvature.

#### 1. INTRODUCTION

Suppose we are given a Riemannian *m*-manifold *M* with some metric associated metric *g*. *M* then, given a choice of basepoint  $x_0$ , has a fundamental group  $\pi_1(M, x_0)$ . The fundamental group has a natural interpretation as the collection of all families of loops through the point  $x_0$  up to homotopy relative to  $x_0$ . As we have a metric *g*, there is a natural way of measuring the length of a loop, and so, given a family of loops  $[\gamma]$  in  $\pi_1(M, x_0)$ , it is a natural question to ask what the length of the smallest loop is in  $[\gamma]$ . And recognizing that the choice of  $[\gamma]$  and  $x_0$  is non-arbitrary, it makes sense to consider the smallest length of all non-trivial curves in *M*. This is the systole problem for a fixed metric *g*. Once one considers this problem for a fixed *g*, it is natural to consider variations of metrics, over all Riemannian metrics for a general Riemannian m-manifold or over conformally equivalent metrics in the case of a Riemann surface, for example. But it is apparent that two metrics which are equivalent may not have the systolic value as, given a metric *g*, we could simply look at  $\frac{1}{2}g$ . To avoid issues with rescalings such as this, one looks instead at the smallest isosystolic ratio over a class of metrics, where the isosystolic ratio is given by

(1.1) 
$$\frac{(Sys(M,g))^m}{Vol(M,g)}$$
 [2].

Notice that the volume scales as a power of m, and the systole scales linearly. So the isosystolic ratio functions as an invariant amongst equivalent metrics. Notice too that, if one chooses a fixed lower bound for systolic length, this problem is exactly a minimal volume problem with the systolic constraint. Explicit metrics satisfying the isosystolic ratio were found by Pu [8], Loewner (unpublished, see [8]) and Bavard [1] for  $\mathbb{RP}^n$ , the torus, and the klein bottle respectively. And for manifolds of a special type, Gromov [4] gave an upper bound for the isosystolic ratio dependent only on the dimension of the manifold. With that said, the problem, in general, is very difficult to solve, and so it makes sense to add more structure to make the problem more amenable to solution.

A simpler, though certainly non-trivial, case of this sort involves fixing a Riemann surface *S* instead of a general Riemann *m*-manifold *M* and fixing a conformal class of metrics on *S*, usually a class with all metrics of the form  $\rho |dz|$  where  $\rho$  is a Borel-measurable map on *S*. Calabi [2] reduced this problem to a variational problem, requiring solution to a particular non-linear partial differential equation. Zwiebach and Headrick [5] worked along a very similar homotopy problem and generalized some of Calabi's notions to convert the problem away from a variational problem to a convex optimization problem, one treatable by simple computational methods.

Extremal length can be viewed as a simpler version of the problem Calabi, Zwiebach, and Headrick address. In particular, beyond fixing a Riemann surface and a conformal class of metrics, the extremal length problem fixes the family of curves considered to be some given family  $\Gamma$ , which in general can be much simpler than the collection of all non-trivial curves or collections of homotopy or homology classes. Again, in order for extremal length to be invariant in conformally equivalent spaces, issues of rescaling need to be addressed and so we define the extremal length of the family of curves  $\Gamma$ , denoted  $\lambda(\Gamma)$  by

(1.2) 
$$\lambda(\Gamma) := \sup_{\rho} \frac{\inf_{\gamma \in \Gamma} L^2(\gamma, \rho)}{A(S, \rho)}$$

where  $L(\gamma, \rho)$  is the length of the curve  $\gamma$  under the metric  $\rho$  and  $A(S, \rho)$  is the area of the surface under the metric  $\rho$ . And, as with the isosystolic ratio problem, one can fix the length requirement for the family of curves and consider only the minimal area problem.

In this paper, we calculate the extremal length of families of curves for which the flat metric is extremal in all regular polygonal tesselations of the plane and the 30-60-90 tesselation of the plane. In doing so, we find that the extremal metric for a 2 and -1 winding curve about two punctures in the thrice punctured sphere identical to the hexagon of Calabi [2], which was given full numerical treatment in a recent work of Zwiebach and Naseer [7].

The resolution of this winding curve's extremal metric inspired the investigation of the moduli space of the hexagonally punctured hexagonal tori with curves moving edge to opposite edge. This extends the results of Zwiebach and Naseer in the hexagonal torus to cover more of the moduli space of tori in the manner of Zwiebach and Headrick, who investigated the moduli space of square-punctured square tori in [6]. We fid an exoression for extremal length for elements of the moduli space parametrized by the ratio of the apothem of the puncture and the apothem of the outer hexagon. We find regions of positive curvature apparently associated to the intersection of three families of geodesic curves and regions of flat curvature apparently associated to the intersection of two families of geodesic curves, in addition to some spikes of negative curvature..

#### 2. Tesselations of the Plane

2.1. **30-60-90 Tesselation.** Consider the tesselation of the plane given by 30-60-90 triangles, as seen in Figure 2.1, and see that it requires two 30-60-90 triangles attached at their



FIGURE 2.1. On the left, we see the tesselation of the plane by 30-60-90 triangles. And one element of both the green and pink homotopy glasses, as well as the singularity which separates the classes. On the right is the green path on the two-cover of the space. We will investigate the green case in particular later.

longer legs in the natural way to form an equilateral triangle with the two lower edges identified together and the two hypotenuses identified together. Forming a sphere is then given by placing a point *a* at the left and right base corners, the point at infinity at the base's intersection with the two triangle's shared edge, and *b* at the upper vertex of the equilateral triangle. We look at the horizontal paths, as in the figure. This partitions the region into 2 homotopy classes separated by critical points, one in region *A* (green) and one in region *B* (pink). We consider the weighted multi-curve *C* of the form  $a_1\gamma_1 + a_2\gamma_2$  with  $\gamma_1$  an *A* path and  $\gamma_2$  a *B* path. Suppose  $\pi_1(\mathbb{S}^2 \setminus \{a, b, \infty\})$  is generated by  $\alpha$  and  $\beta$ . Then  $\gamma_1$  and  $\gamma_2$  are given by  $\beta^{-1}\alpha\beta\beta\beta\alpha$  and  $\beta^{-1}\beta^{-1}\alpha\beta\alpha\beta^{-1}$  respectively. We must have that, if  $\rho_0$  is to be extremal that the multicurve has associated weights 1 and  $\frac{1}{2}$ . This ansatz, suggested by an application of Beurling's criterion to the flat metric.

It suffices now simply to calculate the extremal metric with respect to  $\rho_0$  in the plane. When we normalize the tall leg of the triangle to be of length one, then we get that the square of the length of each path is  $(2(2 \cdot \frac{1}{\sqrt{3}} + \frac{1}{2\sqrt{3}} \cdot 2))^2$  and the area is  $2 \cdot 2(\frac{1}{\sqrt{3}} + \frac{1}{2\sqrt{3}})$ , yielding  $\lambda(\Gamma) = 2\sqrt{3}$ .

2.2. Hexagonal Tesselation. We look now at rectangles formed by connecting any two central vertices of two disjoint pentagons in the hexagonal tesselation of the plane as seen in Figure 2.2. Observe that if we connect the vertices of two hexagons along a ray emanating from both vertices that we are simply at the 30-60-90 problem again. Consider instead the shortest path from one vertex  $x_0$  to a vertex  $x_1$  not along one of the rays of  $x_0$ . Then there are six homotopy classes of equal weights A, B, C, D, E, F. Suppose  $\pi_1(S^2 \setminus \{a, b, \infty\})$  is generated by  $\alpha$  and  $\beta$ . Then we have that A is given



FIGURE 2.2. On the left, we see the tesselation of the plane by regular hexagons. On the right we see the hexagonal torus formed by the gluing operations induced by the tesselation.

by  $\beta\beta\alpha\beta\alpha\beta^{-1}\beta^{-1}\alpha^{-1}\beta$ , B by  $\beta^{-1}\alpha^{-1}\beta\beta\beta\alpha\beta\alpha\beta^{-1}$ , C by  $\alpha^{-1}\alpha^{-1}\beta^{-1}\beta^{-1}\alpha\beta\beta\alpha^{-1}\alpha^{-1}\beta$ , D by  $\beta^{-1}\alpha^{-1}\beta\beta\beta\alpha\beta\alpha\beta^{-1}$ , E by  $\beta^{-1}\alpha\beta\beta\alpha^{-1}\beta^{-1}\alpha^{-1}\beta^{-1}\beta^{-1}$ , ad F by  $\alpha\beta\alpha\beta^{-1}\beta^{-1}\alpha^{-1}\beta^{-1}\beta^{-1}$ . So the extremal length of this set of multicurves is simply  $\sqrt{5-2\sqrt{3}}$ , given by simply calculating the length between vertices when normalizing the height of the rectangle as 1.

2.3. Equilateral triangles. Now consider the tesselation 2.3 of the plane given by equilateral triangles of side lengths 1, forming quadrilaterals each with opposite edges identified in a parallel manner. And so forming a torus. The vertical paths in this tesselation correspond exactly the homotopy class of the meridian, and the horizontal paths in this tesselation correspond exactly to the longitudinal paths. There are corresponding weights one and one. So the extremal length of the longitudinal paths is  $\frac{2}{\sqrt{3}}$  and the extremal length of the latitudinal paths is  $\frac{\sqrt{3}}{2}$ . And the extremal length of the 1 – 1 multicurve is given by 1.

2.4. **Square Tesselation.** As for the square tesselation of the plane, consider two squares placed left to right as in Figure 2.4. Tesselation requires the left and right edge to be identified, the top left edge with the bottom right edge, and the bottom left edge with the top right edge. This forms then a torus with a half twist. The horizontal paths correspond simply to the longitudes, and so the extremal length is simply  $\frac{2^2}{2} = 2$ . The vertical paths consists of paths homotopic to the twisted meridian i.e. a chosen representative path which corresponds in the half-glued two-cover to two parallel horizontal lines. They have an extremal length  $\frac{1}{2}$ . There are associated weights 1 and 1, so we get that that extremal length of their weighted multicurve is  $\frac{4^2}{4} = 4$ .



FIGURE 2.3. On the left, we see the tesselation of the plane by equilateral triangles, as well as two different families of curves: one in orange, and one in pink and purple. Both are portrayed on the two-cover.



FIGURE 2.4. On the left, we see the tesselation of the plane by squares, as well as two homotopy classes: one in blue and one in orange. Both are demonstrated firstly on the two-cover, in the center, and then on the half-glued two-cover, on the right. See that there is indeed a half twist to be done in the last gluing operation in the half-glued two-cover.

#### 3. The 30-60-90 Lift and the Work of Headrick and Zwiebach

Consider again the 30-60-90 tesselation of the plane shown in Figure 2.1 and recall the A and B curves. The A (green) and B (pink) curves are as in the figure. We see straightforwardly that the pink curves cover the surface evenly and so by Beurling's criterion it must be that the metric is simply flat. However, the green curves do not cover the surface evenly and so we choose to investigate the extremal metric of the B paths alone. Consider the process of tracing out a single B path, beginning at the base. Moving upward one hits the hypotenuse of the triangle. But instead of "bouncing off" the hypotenuse as

EXTREMAL LENGTH OF WINDING CURVES



FIGURE 3.1. On the left, we see the original unfolding of the path of the green and pink paths, with appropriate identifications for the large cover. And on the right we identify the central hexagon foliated by all three homotopy classes, and then the equilateral triangles completing hexagram each foliated by two homotopy classes, and then finally the remaining triangles foliated simply by a single homotopy class.

in the previous diagram, we should imagine the line continuing straight into another 30-60-90 triangle glued hypotenuse to hypotenuse with the starting hypotenuse, continuing in this manner until one returns back to the base, which coincides precisely with the upper edge of a half hexagon. Considering the rest of the curves gives a full hexagon, with the symmetries and gluing operations of the triangles forming it requiring opposite edges to be identified. In Figure 3.1 is the full hexagon, with both A and B paths shown in their respective colors. With this we can see that the extremal length problem for the family of B curves reduces precisely to the case of curves moving from one edge to another on the hexagon.

This case was originally studied by Calabi [2] and given numerical resolution in a recent paper of Zwiebach and Naseer [7]. The extremal area A is bounded by .8400 < A < .8414. As predicted by Calabi, Zwiebach and Naseer find that there is a central region  $U_3$  foliated by three families of geodesic curves and a region  $U_2$  foliated by two orthogonal bands of geodesics. However, there are neither regions foliated by a single family of geodesic curves nor a region unfoliated by a family of geodesics. And further, beyond  $U_2$  having flat curvature, it is found that  $U_3$  has the curvature of a half sphere. The pattern of geodesics from their paper is shown below.

EXTREMAL LENGTH OF WINDING CURVES



FIGURE 3.2. This is the image of the geodesic paths of Calabi's Hexagon which Zwiebach and Naseer produce. The regions foliated by three families of geodesics with half-spherical curvature is the central hexagonal region, while the triangles surrounding each vertex are flat and foliated by two families of geodesics [7].

#### 4. The Hexagonally Punctured Torus

4.1. The Method of Headrick and Zwiebach. Recall that the generalized homotopy (homology) systole problem involves, for a collection of non-trivial homotopy (homology) classes  $D_{\alpha}$  ( $C_{\alpha}$ ) in  $\pi_1(M^m)$  ( $H_1(M)$ ), for an *m*-manifold *M*, indexed by some set *A*, a collection of lengths  $l_{\alpha}$  similarly indexed so that for each  $D_{\alpha}$  ( $C_{\alpha}$ ) we have that for the extremal metric  $\Omega$  that

$$l_{\alpha} - length(\gamma, \Omega) \leq 0,$$

for each  $\alpha$ , and for any path  $\gamma \in D_{\alpha}$  ( $C_{\alpha}$ ).  $length(\gamma, \Omega)$  denotes the length of the path  $\gamma$  under the metric  $\Omega$ . This reads simply that the length of each element of each class  $D_{\alpha}$  ( $C_{\alpha}$ ) has some lower bound for their lengths which the extremal metric  $\Omega$  must satisfy. The problem then is to minimize the area of the manifold while maintaining these length constraints.

Zwiebach and Headrick [5] transform the homotopy problem over a subcollection  $S \subset \pi_1(M)$  into a homology problem by means of a lift to a suitable covering space  $\tilde{M}$ . We sketch, quite roughly, Zwiebach and Headrick's explanation for why this is reasonable, given in section 8 of their paper. Consider the quotient map  $\pi : \pi_1(M) \to H_1(M)$ , and notice that *S* is a collection consisting of non-trivial homotopy classes which are mapped to non-trivial or trivial homology classes, denoted  $S_1$  and  $S_0$  respectively. For the non-trivially mapped elements  $S_1$  there is no real issue as each element of  $\pi(S_1)$  is

homologically non-trivial and so constrained in the homology problem in the appropriate manner. And Zwiebach and Headrick show there is no added constraint as each element of  $\pi(S_1)$  is non-trivial and so contains at least one homotopically non-trivial curve, which ensures an adequate lower bound which is not an added constraint. It suffices to rectify the issue of  $S_0$ .

Zwiebach and Headrick find an appropriate lift by means of a simple surgery argument which takes a simple closed curve  $\gamma_0$  in  $S_0$  and lifts it to a non-trivial  $\tilde{\gamma}_0$ . As M has non-trivial homotopy, it has genus at least one. So  $\gamma$  is the boundary of two regions of M, one with at least one handle about some hole. Cutting about this handle, creating a copy with the same cut of M and gluing these manifolds together forms a double cover of M where  $\tilde{\gamma}_0$  is the addition of two homologically non-trivial curves. So Zwiebach and Headrick demonstrate that each simple closed curve has a lift to a two-fold cover and a non-trivial homology class. Then by the Galois correspondence to an index two, and so normal, subgroup of  $\pi_1(M)$  there is the natural quotient map  $p: \pi_1(M) \to \mathbb{Z}/2\mathbb{Z}$ . See then that p is an element of  $H^1(M, \mathbb{Z}/2\mathbb{Z})$ . And so, in general, there is a  $dim(H^1(M, \mathbb{Z}/2\mathbb{Z}))$ -fold cover of M lifting the elements of  $S_0$  to non-trivial homology classes. Zwiebach and Headrick then demonstrate that this lifting operation adds no new non-trivial homological constraints.

Once the lift operation to a suitable covering space is completed, Zwiebach's analysis provides the means by which to translate the homology problem now established into a convex optimization problem. The conformal class of metrics is constrained, without loss of generality, to metrics  $\Omega g_0$  for  $\Omega$  a measurable function on M sometimes called the *Weyl factor*, and  $g^0$  a choice of *fiduciary metric*. The homology problem then is to minimize the area under all metrics  $\Omega g^0$  while maintaining length constraints as before. This is a convex optimization problem, but it is practically unapproachable as it requires optimizing over a very large infinite dimensional space.

To make the optimization more tractable, it is reformulated in terms of calibrations by Zwiebach and Headrick. A calibration is a closed and unitarily bounded 1–form. By closedness and boundedness of u it is shown that the length constraint on curves in  $C_{\alpha}$  corresponds precisely to the existence of a calibration  $u^{\alpha}$  which *calibrates*  $C_{\alpha}$  i.e. for an arbitrary representative  $m_{\alpha}$  of  $C_{\alpha}$ ,

(4.2) 
$$\int_{m_{\alpha}} u^{\alpha} = l_{\alpha}.$$

This yields the primal optimization program written fully as

(4.3) Minimize  $Area(M, \Omega)$  over  $\Omega, u^{\alpha}$ 

with 
$$\Omega - |u^{\alpha}|_0^2 \ge 0$$

where  $u^{\alpha}$  calibrates  $C_{\alpha}$  and  $|\cdot|_0$  denotes the fiducial metric norm. A much more detailed analysis is provided in [5].

4.2. Setup for the Convex Optimization Problem. Consider a hexagon centered at the origin which bulges from  $-\frac{1}{\sqrt{3}}$  to  $\frac{1}{\sqrt{3}}$  along the *x*-axis and which has 2n edges thereon

(without multiplicity) formed by equilateral triangles, as seen in Figure 4.1, and let the lattice points (x, y) be the vertices of these triangles. Consider the vertical level sets formed by the rows of these triangles. We denote their *y*-components by y[i, j] with i = 0 corresponding to the level set on the *x*-axis, i = 1 corresponding to that on the first level above the *x*-axis and so on. Notice that there are 2n + 1 such level sets and that by our choice of the placement of i = 0 we have that *i* ranges over  $[-n, n] \cap \mathbb{Z}$ . Similarly for the horizontal aspects of a lattice points, which we shall index by x[i, j] as its value depends both on the *i* value which determines why and some chosen *j* value. We can view each x[i, j] as horizontal values shifted first to the left edge of the hexagon by a constant of  $\frac{-1}{\sqrt{3}}$  and then to the right by  $\frac{|i|}{2\sqrt{3}n}$  as the height moves away from the *x*-axis to end up on the left edge of the hexagon at the height y[i, j], and then shifted right by *j* triangular edges, giving a term of  $\frac{j}{\sqrt{3}n}$ . We see then that, for a fixed *i*, *j* must range over  $[0, 2n + 1 - |i|] \cap \mathbb{Z}$ . To be explicit, we have

(4.4) 
$$y[i,j] := \frac{i}{2n} ; x[i,j] := \frac{-1}{\sqrt{3}} + \frac{|i|}{2\sqrt{3}n} + \frac{j}{\sqrt{3}n}$$



FIGURE 4.1. This is the discretization for n = 2 and k = 1. In this diagram the sharp purple objects correspond to ones, reds to twos, and the magentas to threes. The short arrows correspond to calibrations, the long ones to local coordinate forms, and the highlighted edges to the representative paths from the different homology classes. As examples, the red edge corresponds to  $m^2$  and the magenta short arrow corresponds to  $u^3$ .

We look now at the  $u^{\alpha}$  and determine that  $u^1$  points in the direction of  $\hat{x^1}$ ,  $u^2$  in the direction of  $\hat{x^3}$  and  $u^3$  in the direction of  $\hat{x^2}$ . See that under the action of rotation by 60 degrees about the origin that we have  $u^1 \mapsto u^2 \mapsto u^3$ . So for  $u^1 = a_1 dx^1 + a_2 dx^2 + a_3 dx^3 + d\phi^1$ , we have by Hodge decomposition theorem that  $d\phi^2 \circ (R_{60}^{\rightarrow})^{-1}(x,y) = d\phi^1(x,y)$  and that  $u^2 = -a_1 dx^3 - a_2 dx^1 - a_2 dx^2 + d\phi^2$ . Similarly,  $d\phi^2 \circ (R_{60}^{\rightarrow})(x,y) = d\phi^3(x,y)$  and  $u^3 = a_1 dx^2 + a_2 dx^3 + a_3 dx^2 + d\phi^3$ . As we assume the systolic length is 1, integrating along  $m^2$  we find that  $\frac{1}{2}a_1 + \frac{1}{2}a_3 - a_2 = 1$ . And seeing as  $u^1$  is invariant under reflection across the ray from the origin in the direction of  $\hat{x^1}$ , we must have that  $a_3 = a_1$ . Combining these two results, and that  $dx^1 + dx^2 + dx^3 = 0$ , we get that

(4.5)  
$$u^{1} = dx^{2} + d\phi^{1}$$
$$u^{2} = dx^{1} + dx^{2} + d\phi^{2}$$
$$u^{3} = dx^{2} + d\phi^{3}.$$

We wish to work with the  $\phi^2$  function we investigated in setting up the program. We define our function on the lattice points of the hexagon and so define  $\phi^2[i, j] :=$   $\phi^2(x[i, j], y[i, j])$ . The metric and derivatives however will be defined on the centers of the equilateral triangles, with each center indexed by the three vertices which generate it [[i, j], [k, l], [m, n]]. In fact, by symmetry across the x-axis, we need only define the function on those centers above the x-axis. We consider the upward pointing triangles and downward pointing triangles separately. As we are only considering the triangles in the upper half plane, it is straightforward to see that all upward pointing triangles are of the form [[i, j], [i - 1, j], [i - 1, j + 1]] with the second and third elements obviously forming the basis of the triangle and the first being the peak. Similarly downward facing triangles are of the form [[i, j], [i, j + 1], [i - 1, j + 1]].

For an upward (respectively, downward) facing triangle we have that the partial derivatives in the direction of  $\hat{x^1}$  are given by the average of the average of the rightmost (leftmost)  $\phi^2$  values taken with a signed difference with the other  $\phi^2$  value divided by the distance by the sole point and the average location of the other two points. Similarly for the partial derivative in the direction of  $\hat{x^2}$ . More explicitly we have for an upward triangle and for  $\phi^l$  partial derivatives given by

$$\begin{pmatrix} \frac{\partial \phi^l}{\partial x^1} \end{pmatrix}_{x^2} [[i,j], [i-1,j], [i-1,j+1]] := \\ \frac{1}{\frac{1}{2n}} \cdot \left( \frac{1}{2} (\phi^l[i,j] + \phi^l[i-1,j+1]) - \phi^l[i-1,j] \right)$$

$$\begin{pmatrix} \frac{\partial \phi^{i}}{\partial x^{2}} \end{pmatrix}_{x^{1}} [[i, j], [i - 1, j], [i - 1, j + 1]] := \\ \frac{1}{\frac{1}{2n}} \Big( \frac{1}{2} (\phi^{l}[i, j] + \phi^{l}[i - 1, j]) - \phi^{l}[i - 1, j + 1] \Big).$$

Obviously we have similarly for downward facing triangles.

What one would very much like to do in executing this program is limit that you are calculating this program repeatedly for each  $\phi^l$  instead of simply for  $\phi^2$ , as the derivative expressions are quite cumbersome and would seem to require a bit of calculation. which is too tedious. Let us denote by  $R_{60}^{\rightarrow}[i, j]$  rotation of the vertex [i, j] by 60° clockwise about the origin. And let  $R_{60}^{\rightarrow}(T)$  be the obvious rotation of a triangle *T* about the origin. We see then that

$$(4.6) \qquad \begin{pmatrix} \frac{\partial \phi^3}{\partial x^1} \end{pmatrix}_{x^2} T = -\left(\frac{\partial \phi^2}{\partial x^2}\right)_{x^1} R_{60}^{\rightarrow} T - \left(\frac{\partial \phi^2}{\partial x^1}\right)_{x^2} R_{60}^{\rightarrow} T \\ \left(\frac{\partial \phi^3}{\partial x^2}\right)_{x^1} T = -\left(\frac{\partial \phi^2}{\partial x^1}\right)_{x^2} \circ R_{60}^{\rightarrow} T \\ \left(\frac{\partial \phi^1}{\partial x^2}\right)_{x^1} T = -\left(\frac{\partial \phi^2}{\partial x^2}\right)_{x^1} (R_{60}^{\rightarrow})^{-1} T - \left(\frac{\partial \phi^2}{\partial x^1}\right)_{x^2} (R_{60}^{\rightarrow})^{-1} T \\ \left(\frac{\partial \phi^1}{\partial x^1}\right)_{x^2} T = -\left(\frac{\partial \phi^2}{\partial x^2}\right)_{x^1} \circ (R_{60}^{\rightarrow})^{-1} T$$

It suffices to give an expression for these rotation terms, but these are readily calculable by examining the action of a rotation matrix on the vector  $(x[i,j], y[i])^T$ , which results in the calculation that for  $R_{60}^{\rightarrow}[i,j] = [i',j']$  that

(4.7)  
$$i' = \frac{2n - |i| - 2j + i}{2},$$
$$j' = \frac{2n + |i| + 2j + 3i - 2|i'|}{4}$$

Similarly there is an expression for the inverse  $(R_60^{\rightarrow})^{-1}[i, j]$ , achieved by the same methods.

Then, as we have our calibrations sorted, and as we have discretized our functions and hexagon in an appropriate manner, it suffices simply to constrain the Weyl factor  $\Omega$  in the requisite manner. We have a fiducial metric given by the quadratic form

$$Q(x^{1}, x^{2}) = (dx^{1})^{2} + (dx^{2})^{2} - dx^{1}dx^{2}.$$

So we have the constrains

(4.8)  $\begin{aligned} \Omega T &\geq \left[ (1 - \partial_2^1)^2 + (-\partial_1^2 - \partial_2^1)^2 - (1 - \partial_2^1)(-\partial_1^2 - \partial_2^1) \right] \circ \phi^2 \circ (R_{60}^{\rightarrow})^{-1} T \\ \Omega T &\geq \left( (1 + \partial_1^2)^2 + (1 + \partial_2^1)^2 - (1 + \partial_1^2)(1 + \partial_2^1) \right) \circ \phi^2 T \\ \Omega T &\geq \left[ (-\partial_1^2 - \partial_2^1)^2 + (1 - \partial_1^2)^2 - (-\partial_1^2 - \partial_2^1)(1 - \partial_1^2) \right] \circ \phi^2 \circ R_{60}^{\rightarrow} T \end{aligned}$ 

The program now is explicitly calculable and gives  $\Omega$  as the maximum of these constraints on each equilateral triangle.

The issue now is to define the curvature *K* on the surface, given by  $\frac{-1}{2\Omega}\nabla^2 \ln \Omega$ . Doing so requires taking second derivatives of  $\ln \Omega$  along the surface. The natural place where we can take partial derivatives is at the barycenter when the values of the function to be differentiated are defined on the lattice points, as we did with  $\phi^2$ . When the function is defined on the triangles though, there is no such natural operation. To fix this issue, we redefine omega to be on the lattice points by averaging the values of  $\Omega$  in each triangle surrounding a given lattice point. And then differentiation of this in the  $x^1, x^2$  frame is as before. But then when taking the second derivative again we run into the issue of the natural differentiation in this lattice is against a function defined on the lattice points. As such, we perform the the same operation with the derivative of  $\Omega$  as we did with  $\Omega$ itself. And so we gain the partial derivatives defined on the barycenters. This method works, though it will smooth out and somewhat distort the metric at hand. The extent to which this is the case is unclear, though, as we shall see, there is still interesting behavior of the curvature which conforms with what one might expect to occur.

4.3. Numerical Results and Metric Description. Here we assume the systole length is 1, as in the previous sections, and we present the qualities of the extremal metric. We plot the whole domain. The highest resolution we managed to obtain was at n = 32, where we have the following table of values for the area of the hexagonally punctured hexagonal torus.

k	Area
1	0.74
8	0.70
16	0.57
24	0.36
30	0.13

We pay close attention to the area and metric of the hexagonally punctured torus, with special attention to the case with k = 16. The graphics here are taken at resolution n = 32, which generates the estimate extremal area A as





FIGURE 4.2. Two views of the  $\phi^2$  values



FIGURE 4.3. Here we can see the pronounced spikes of  $d\phi$  on the x-axis and especially at the vertices thereon, both on the puncture, and on the outer hexagon.

In Figure 4.2 is the surface of phi for k = 16 at the highest resolution. We see that  $\phi^2$  appears to increase linearly as it moves from the lower edge to the upper edge, excepting

some pronounced spikes at all four vertices. So one might expect that the geodesics are given by a unique curve of constant derivative which originates at some point on the lower edge and travels upwards along a unique path fully determined by, for a point p, a function H of something like the form  $\phi^2(p) + \int_{-\frac{1}{2}}^{y} d\phi^2$ . And then the two possible x values for the two possible geodesics might be given by two points in the preimage of H. However, the degree of the spikes, and that the spikes diverge in both positive and negative directions, is perhaps indicative of the inadequacy of the maximal resolution. It is not clear precisely how to read off directly geodesic values from the  $\phi^2$  values as the differential of the  $\phi^2$  values about the *x*-axis spikes as in Figure 4.3, though it should in principle be possible.

We look now at the extremal metric  $\Omega$  and  $\ln \Omega$ , as shown in Figures 4.4 and 4.5. We see it is almost flat for much of the surface with slight bulges about the puncture. And it blows up around the vertices. This is the expected behavior, in line with the results of Zwiebach and Headrick's results for the swiss cross. As Zwiebach and Headrick saw in their investigation of the swiss cross [6], the "minimal-area problem wants the boundary to be smooth" (30) and so, for us, blowing up at the vertices may correspond to a smoothing of the metric around the vertex, but not at the vertex itself.



FIGURE 4.4. Two views of  $\ln \Omega$ . Notice the smootheness of this metric as opposed to simply  $\Omega$ .



Figure 4.5. Two views of  $\Omega$ . 14

It is straightforward to look at the curvature now, and it is shown in Figure 4.6. The result is that the metric  $\Omega$  appears to have regions of flat and of positive and negative curvature, but has an assortment of spikes around the hexagon, especially between vertices of the hexagon and punctures and at the vertices on the x-axis. The spikes between vertices are positive while the spike at the vertices of the x-axis is extremely negative, seemingly a point-mass. The positive curvature between vertices is explainable as the line where geodesics from one curve first meets its adjacent curve, and where they are met by geodesics of the third family of curves. However, if this area had wholly positive curvature one would expect that to be the case except in a few small regions foliated only by two families of geodesics. Instead we see alternating positive and negative curvature which changes fairly rapidly, though in regions bounded by two edges, one from the puncture and one from the outer hexagon, there are noticealy fewer spikes with negative curvature. As for the regions of flat curvature, they appear to occur very closely to the edges of the puncture, and would seem to correspond with regions foliated simply by two families of geodesics. However, it should be said that the curvature seen here is also a residue both of the manner in which the curvature was calculated and of the accuracy of the  $\phi^2$  values, as discussed earlier. However the operation of averaging values across triangles is much more suggestive of smoothing out the metric, and so the curvature. But this is precisely the opposite of what we see here.



FIGURE 4.6. Two views of the curvature *K*.

4.4. Moduli Space of the Hexagonally Punctured Hexagonal Torus. The parameter h which is given by the ratio of the hexagonal puncture's apothem and the outer hexagon's apothem fully determines the hexagonally punctured hexagon, and so makes sense as the modulus. We can of course take then h in [0,1]. The case of h = 1 is trivially 0 and the case of h = 0 is precisely the case studied by Zwiebach and Naseer [7]. It is straightforward to see that the doubly connected region of the hexagonally punctured torus has different annular modulus, and so different extremal length for differing values of h. Obviously the family  $\Gamma$  here consists exactly of the homology of curves described above. And as the shortest lengths of curves here have length one in the extremal metrics

found above, it is the case that

(4.9) 
$$\lambda(h) := \frac{1}{A(h)}.$$

That ism the extremal length is simply a function of the modulus h. The relation we find is give y a quadratic with great precision, an r-value of 0.9999296489.

Quadratic regression reveals that the relation is given by

(4.10) 
$$\lambda(h) = \frac{1}{0.737673 + 0.0013617h - 0.000703976h^2}.$$

4.5. **Convergence.** We will now briefly look at the convergence of the optimization program for the case k = 16 or, rather, the case where  $k = \frac{n}{2}$ . For this case, the following table gives values

n	Area	$\Delta(Area)$
20	0.609313728	0.01886890848
24	0.59507735	0.01423637808
28	0.584459574	0.01061777534
32	0.578880912	0.0055786623

In the case swiss cross of Zwiebach and Headrick, the convergence of the program was proportional to  $(resolution)^{-3}$ . We would expect something similar to occur here so we model the difference as a function of the power of the resolution. In doing so, we find that the convergence rate is well approximated (r = -0.960313538) by the expression

(4.11) 
$$\Delta(A) = 34.65141n^{-2.47750982}.$$

The primal program strictly decreases as the resolution increases as each  $\Omega$  value is feasible and maintains the length constraints. So to get an approximate bound on the expected value we should consider our most accurate numerical estimate shaved off by a factor of the infinite sum of values after 32 of the change in area as the resolution increases i.e.  $0.578880912 - \frac{1}{4} \int_{3} 2^{\infty} \Delta(A) dn = 0.56$ . We expect then that the area will be be  $0.56 \pm .02$ , where 0.56 is the difference of our best predicted area and half the integral expression for  $\Delta(A)$  and .02 is half the integral the integral expression for  $\Delta(A)$ .

The expected results for each value of k then is given in the following table, with expected error of at most .02. Given that the value of the areas are shifted simply by a constant here, the metric, curvature, and modular information presented above still holds.

k	Area		
1	0.72		
8	0.68		
16	0.56		
24	0.34		
30	0.11		
16			

#### 5. Points for further investigation

As for areas of possible further investigation, it would be good to fully determine a way by which to read off the geodesics from  $\phi^2$ . And to prove the area formula of Zwiebach and Headrick. Further investigation of the moduli space of tori is fairly natural, as well as of multi-punctured tori. There is also the idea of a probabilistic interpretation of geodesic density as a criterion for extremality of a metric. Finally, a great result would be to find a general principle for the extremal values of combinations of curves. This seems, in general, intractable given the swiss cross being the simples combination of simple curves and having no simple answer. But that we have calculated a number of elements of winding curves of the type 2 and -1 in the thrice punctured sphere would make it very nice if there was, in general, an algorithm for expanding results of the type presented here into general results about the 2, -1 winding curves.

#### 6. Acknowledgements

This research was performed under the advisement of Professor Dylan Thurston, Didac Martinez-Granado, and Insung Park at the Indiana University Mathematics REU program, funded by NSF grant #1757857.

#### References

[1] C. Bavard, "Inégalités isosystoliques conformes," Comment. Math. Helv. 67 (1992) 146-166.

[2] E. Calabi, "Extremal isosystolic metrics for compact surfaces." pp. 167-204 in Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992), Soc. Math. France, Paris, 1996.

- [3] J. Kahn, K. Pilgrim, D. Thurston, "Conformal surface embeddings and extremal length," arXiv:1507.05294 [math.CV]
- [4] M. Gromov, "Filling Riemannian manifolds," J. Differential Geom. 18 (1983) 1-147.
- [5] M. Headrick and B. Zwiebach, "Convex programs for minimal-area problems," arXiv:1806.00449 [hep-th].

[6] M. Headrick and B. Zwiebach, "Minimal-area metrics on the Swiss cross and punctured torus," arXiv:1806.00450 [hep-th].

[7] U. Naseer and B. Zwiebach, "Extremal isosystolic metrics with multiple bands of crossing geodesics," arXiv:1903.11755 [math.DG]

[8] P. M. Pu, "Some inequalities in certain nonorientable Riemannian manifolds", Pacific. J. Math. 2 (1952) 55-71

# Mixing of Random Walks on Punctured Graphs

Chung Kyong Nguen\* Final Report Princeton University Advised by Graham White

September 18, 2019

#### Abstract

Classical probabilistic techniques for bounding mixing time of Markov chains, such as strong stationary times and couplings, typically rely on the symmetry of the underlying graphs for the given chains. We modify existing couplings to be able compare mixing times of these regular graphs with their distorted counterparts, which are created by removing a random vertex or a sparse set of vertices in the original graph. Additionally, we investigate the mixing times of random walks with multiple walkers, where they are not allowed to be in the same vertex.

# Contents

<sup>\*</sup>Funded by NSF REU at Indiana University

## **1** Preliminaries

First, we introduce the basic definitions and the concept of mixing of Markov chains.

**Definition 1.1.** A sequence of random variables  $(X_0, X_1, ...)$  is a Markov chain with state space  $\Omega$  and transition matrix P if for all  $x, y \in \Omega$ , all  $t \ge 1$ , and all events  $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$  satisfying

$$P(H_{t-1} \cap \{X_t = x\}) = P(X_{t+1} = y | X_t = x)$$
(1)

**Definition 1.2.** We call  $\pi$  a stationary distribution of the Markov chain with transition matrix *P* if it satisfies:

$$\pi = \pi P \tag{2}$$

**Definition 1.3.** *The total variation distance between two probability distributions*  $\mu$  *and*  $\nu$  *on*  $\Omega$  *is defined by* 

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \sum_{x \in \Omega, \mu(x) > \nu(x)} |\mu(x) - \nu(x)|$$
(3)

**Definition 1.4.** A coupling of two probability distributions  $\mu$  and  $\nu$  is a pair of random variables (X, Y) defined on a single probability space such that the marginal distribution of X is  $\mu$  and the marginal distribution of Y is  $\nu$ . That is, a coupling (X, Y) satisfies  $P(X = x) = \mu(x)$  and  $P(Y = y) = \nu(y)$ .

**Theorem 1.1.** Let  $\mu$  and  $\nu$  be two probability distributions on  $\Omega$ . Then

$$\|\mu - \nu\|_{TV} \le P(X \ne Y) : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu.$$
(4)

*Proof.* Note that for any coupling (X, Y) of  $\mu$  and  $\nu$  and any event  $A \subset \Omega$ 

$$\mu(A) - \nu(A) = P(X \in A) - P(Y \in A) \le P(X \in A, Y \in A) \le$$
  
$$\le P(X \neq Y)$$
(5)

**Definition 1.5.** *The mixing time is defined by* 

$$t_{mix}(\varepsilon) = \min\{t : d(t) = \max_{x \in \Omega} \left\| P^t(x, \cdot) - \pi \right\|_{TV} < \varepsilon\}$$
(6)

**Theorem 1.2.** Let  $\{(X_t, Y_t)\}$  be a coupling such that the two chains stay together at all times after their first simultaneous visit to a single state, more precisely

if 
$$X_s = Y_s$$
, then  $X_t = Y_t, t \ge s$ . (7)

Suppose  $X_0 = x, Y_0 = y$  and  $t_{couple} = \min(t : X_t = Y_t, t \ge s)$  Then,

$$\left\|P^{t}(x,\cdot) - P^{t}(y,\cdot)\right\|_{TV} \le P_{x,y}(t_{couple} > t)$$
(8)

*Proof.* Notice that  $P^t(x, z) = P_{x,y}(X_t = z)$  and  $P^t(y, z) = P_{x,y}(Y_t = z)$ . Consequently,  $(X_t, Y_t)$  is a coupling of  $P^t(x, \cdot)$  with  $Pt(y, \cdot)$  and so,

$$\left\|P^{t}(x,\cdot) - P^{t}(y,\cdot)\right\|_{TV} \le P_{x,y}(X_{t} \ne Y_{t})$$

$$\tag{9}$$

which suffices for the proof.

**Corollary 1.2.1.** Suppose that for each pair of states  $x, y \in \Omega$  there is a coupling  $(X_t, Y_t)$  with  $X_0 = x$  and  $Y_0 = y$ . For each such coupling, let  $t_{couple}$  be the coalescence time of the chains, as defined in previously. Then

$$d(t) \le \max_{x,y \in \Omega} P_{x,y}(t_{couple} > t)$$
(10)

## 2 Punctured Symmetric Group

In this section, we investigate the following random walk. Consider shuffling a deck of n cards in the following way: take a random card and put it on top. Additionally, disallow a random permutation  $\sigma$ , that is if we choose to transition to disallowed permutation, the Markov chain stays in the same position.

**Theorem 2.1.** Let  $(X_t)$  be the random walk on  $S_n \setminus \{\sigma\}$  corresponding to the random-to-top shuffle on n cards, with one blocked permutation  $\sigma$ . Then

$$t_{mix}(\varepsilon) \le n \log n + n(\log \varepsilon^{-1} + 1) \tag{11}$$

**Lemma 2.2.** Let  $(X_t)$  and  $(Y_t)$  be two copies of the given Markov chain. Then, at time  $\tau$ , which corresponds to the time when we have chosen all the cards and the card *i* for the second time, the chains have coupled.

*Proof.* Consider the following coupling of two Markov chains with the given transition probabilities  $(X_t)$  and  $(Y_t)$ : for any card *i* that we move to the top in the chain  $(X_t)$ , we move the same card *i* to the top in the chain  $(Y_t)$ , if it is possible. Otherwise, the chain  $(Y_t)$  stays in the same position. Let  $\tau_0$  be the time when each card has been chosen, but not necessarily moved once.

Indeed, first assume that we never run into the situation, where we are unable to move both chains. Then, we note that by our construction, after we choose and move cards i in both chains to top, they will remain at the same position in corresponding decks. Therefore, by time  $\tau_0$  the chains will have coupled.

Suppose, in the permutation  $\sigma$ , the card *i* is sent to the top of the deck. Now, consider the situation, where, a card *i* is moved to the top in the chain  $(X_t)$ , but not in the chain  $(Y_t)$ . Then, we claim that at time  $\tau$ , which corresponds to the time when we have chosen all the cards and then waited for the card *i* to be chosen once more the chains have coupled. Indeed, if we remove the cards *i* from both chains at time  $\tau_0$ , then both decks will be identical. Therefore, after time  $tau_0$  we also have to wait time *T* to move the card *i* to the top.

*Proof of Theorem 2.1.* It is clear that  $\tau_0$  behaves like a coupon collector time and T is a geometric random variable with mean n. Then,

$$\mathbb{P}(\tau > n\log n + n +) \tag{12}$$

## **3** Punctured Hypercube

Let graph  $H_n = (V, E)$  be such that  $V = \{0, 1\}^n$ , that is comprised of all binary strings of length n. Two vertices (v, w) are connected if they differ by one bit.

#### **3.1 Removing a single vertex**

Suppose we remove a random vertex from this graph. Since the graph is vertex transitive, we may assume without loss of generality that it is  $11 \dots 1$ . Consider the following random walk on this graph: with probability  $\frac{1}{2}$  we stay in the same position, and with probability  $\frac{1}{2n}$  we move to any of the neighboring vertices. If we choose to move to the removed vertex, we stay in the same position.

**Theorem 3.1.** For the simple random walk on  $H_n/\{11...1\}$ 

$$t_{mix}(\varepsilon) \le \frac{1}{2}n\log n + n\log(\varepsilon^{-1})$$
(13)

First, we will need to prove the following lemma:

**Lemma 3.2.** Suppose  $(X_t) = (X_t^1, X_t^2, \dots, X_t^n)$ , and let  $W_t(X_t) = \sum_i X_t^i$ . Then,  $(W_t)$  is a Markov chain and

$$\|P_x(X_t \in \cdot) - \pi\|_{TV} = \|P_{W(x)}(W_t \in \cdot) - \pi\|_{TV}$$
(14)

*Proof.* We can explicitly give the transition probabilities for  $(W_t)$ . Indeed, if  $W_t = k$ , then  $X_t$  has k ones and n - k zeros, and so for  $k \neq n - 1$ :

$$\mathbb{P}(W_{t+1} = k+1 | W_t = k) = \frac{n-k}{2n}$$

$$\mathbb{P}(W_{t+1} = k-1 | W_t = k) = \frac{k}{2n}$$

$$\mathbb{P}(W_{t+1} = k | W_t = k) = \frac{1}{2}$$
(15)

If k = n - 1, then

$$\mathbb{P}(W_{t+1} = n - 1 | W_t = n - 2) = \frac{n - 1}{2n}$$

$$\mathbb{P}(W_{t+1} = n - 1 | W_t = n - 1) = \frac{n + 1}{2n}$$
(16)

Next, we have to show that this projection chain mixes as fast as the original random walk. To see this, we note that if  $S_w = \{x : W(x) = w\}$ , then the map  $x \mapsto \mathbb{P}_1(X_t = x)$  and  $\pi$  are constant over  $S_w$ , which can be seen as the equivalence classes of the chain  $(X_t)$ . Then,

$$\sum_{x:W(x)=w} |P_x(X_t = x) - \pi(x)| = \left| \sum_{x:W(x)=w} P_x(X_t = x) - \pi(x) \right|$$

$$= |P_{W(x)}(W_t = w) - \pi_W(w)|$$
(17)

Therefore, we have reduced the study of our random walk to the study of slightly modified lazy Ehrenfest Urn on S = 0, 1, 2, ..., n - 1.
**Proof of the Theorem.** The proof is a slight modification of the argument given by the theorem 18.3 in [?]. Consider the following coupling: if both chains are not in n - 1, then one of the chain remains in the same position, and the other moves. If one of the chains is in n - 1 and the other is in 0, then . If one of the chains is in n - 1 and the other is in [n - 2], then .

Then, if  $D_t = |Z_t - Y_t|$  notice that

$$\mathbb{E}_{z,y}(D_{t+1}|Z_t = z_t, Y_t = y_t) \le \left(1 - \frac{1}{n}\right) D_t$$
(18)

Consequently, we can show

$$\mathbb{E}_{z,y}(D_t \mathbb{1}_{\tau>t}) \le \left(1 - \frac{1}{n}\right)^t n,\tag{19}$$

where  $\tau$  is the coupling time of  $(Z_t)$  and  $(Y_t)$ . Moreover, we  $D_t$  is at least as likely to move downwards as it is to move upwards, and as such we can couple it with  $(S_t)$ , which is a random walk on  $Z_n$ , such that  $D_t < S_t$ . Let  $\tau' = \min\{t \ge 0 : S_t = 0\}$ . Then,  $\tau$  is dominated by  $\tau'$ , and so

$$\mathbb{P}_k(\tau > u) \le \mathbb{P}_k(\tau' > u) \le \frac{c_1 k}{\sqrt{u}}$$
(20)

Then

$$\mathbb{P}_{z,w}(\tau > s + u | D_s) = \mathbb{1}_{\tau > s} \mathbb{P}_{D_s}(\tau > u) \le \frac{c_1 D_s \mathbb{1}_{\tau > s}}{\sqrt{u}}$$
(21)

Taking expectation on both sides we get:

$$\mathbb{P}_{z,y}(\tau > s+u) \le \frac{c_1 n e^{-s/n}}{\sqrt{u}} \tag{22}$$

Substituting  $s = \frac{1}{2}n \log n$  and  $u = \alpha(\varepsilon)$ , we get the desired result.

**Theorem 3.3.** For the random walk on  $H_n \setminus v$ , where v is a vertex chosen uniformly at random

$$t_{mix}(\varepsilon) \le \frac{1}{2}n\log n + n\log(\varepsilon^{-1})$$
(23)

*Proof.* Given that the hypercube is a vertex-transitive graph, we can assume that v = 011...1, and thus

$$\operatorname{Var}_{\pi}(W) < \frac{n}{4}$$
$$\mathbb{E}_{\pi}(W) = \frac{\frac{n}{2}2^{n} - n + 1}{2^{n} - 1} = \frac{n}{2} + \frac{n}{2^{n+1} - 2}$$
(24)

Let R(t) be the number of not refreshed coordinates by time t in the blocked hypercube and R'(t) in the original hypercube. Then for the given random walk, we can couple these random variables, such that R(t) > R'(t).

Note that

$$\mathbb{E}_{\mathbf{l}}(W(X_t)|R(t)) = \frac{1}{2}(R(t)+n) \ge \frac{1}{2}(R'(t)+n)$$

$$\mathbb{E}_{\mathbf{l}}(W(X_t)) \ge \frac{n}{2} \left[ 1 + \left(1 - \frac{1}{n}\right)^t \right]$$

$$\operatorname{Var}_{\mathbf{l}}(W(X_t)) = \frac{1}{4} \operatorname{Var}_{\mathbf{l}}(R(t)) + \frac{1}{4}(n - \mathbb{E}_{\mathbf{l}}R(t)) \le \frac{n}{4}$$
(25)

Then, since we have shown that

$$|\mathbb{E}_{\mathbf{1}}(W(X_t)) - \mathbb{E}_{\pi}(W)| \ge \frac{n}{2} \left(1 - \frac{1}{n}\right)^t$$

$$\sigma = \sqrt{\max(\operatorname{Var}_{\pi}(W), \operatorname{Var}_{\mathbf{1}}(W(X_t)))} \le \frac{\sqrt{n}}{2}$$
(26)

we can apply proposition 7.9 from [?], for f(x) = W(x) and get the desired bound.

#### 3.2 Removing a Sparse Subset

So far, we have seen that removing a single vertex does not have strong impact on the mixing time, which is to be expected, since the graphs we considered had exponentially large vertex sets. Now, let us remove some sparse set S out of  $G_n$ , where  $|S| = \frac{2^{\frac{n}{2}}}{n}$  and is a set of points chosen uniformly at random. First, we have to show that lazy random walk on this graph is well-behaved, in a sense that  $G_n \setminus S$  is connected.

Lemma 3.4. For any two blocked vertices x and y,

$$\lim_{n \to \infty} \mathbb{P}(d(x, y) \le 2) = 0 \tag{27}$$

*Proof.* For any two points in the graph  $G_n$ , the probability that the distance between them is no greater than 2 is

$$\mathbb{P}(d(x,y) \le 2) = \frac{n^2}{2^n - 1}$$
(28)

Taking union bound over the set S, we get that the probability of two blocked points being at distance 2 or closer is

$$\mathbb{P}(d(x,y) \le 2) = \frac{n^2 k(k-1)}{2^n - 1}$$
(29)

Then, sending n to infinity and plugging in  $k = \frac{2^{\frac{n}{2}}}{n}$ , we get:

$$\lim_{n \to \infty} \mathbb{P}(d(x, y) \le 2) = \frac{n^2 k(k-1)}{2^n - 1} \le \frac{n^2 k^2}{2^n - 1} = 0$$
(30)

#### **Lemma 3.5.** $G_n \setminus S$ is connected with high probability.

*Proof.* Consider any pair of disconnected points x and y. Take any path  $p = (x, s_1, s_2, \ldots, s_l, y)$  from x to y that is in  $G_n$ . Since, x and y are disconnected due to removal of the set S, there exists a vertex  $s_k$  that belongs to S. Let  $s_i$  be the first such vertex, then since the distance between any two blocked points is greater than 2,  $s_{i+1} \notin S$ . Now, we claim that x and  $s_{i+1}$  are connected in  $G_n \setminus S$ . Indeed, the distance between  $s_{i-1}$  and  $s_{i+1}$  is 2, and as such there must be a path between them. Repeating this process of replacing blocked vertices, with existing paths of length 2, we can show that any two vertices are connected.

**Lemma 3.6.** Let our chain  $(X_t)$  run up to time  $t_0$ . If S is time spent near blocked vertices, then if  $t_0 \sim n \log n$ 

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{S}{t_0} \ge \frac{1}{n}\right) \to 0 \tag{31}$$

*Proof.* To prove this we construct a Markov chain that will bound our time being next to blocked vertices. The Markov chain will have three states:

- being next to blocked vertex, d = 1
- being at distance 2 away from the closest blocked vertex, d = 2
- being at distance  $\geq 3$  away from the closest blocked vertex,  $d \geq 3$

Using Lemma ??, we can assert that within any ball of radius 1, there is at most 1 blocked vertex, and so the transition probabilities are given by:

$$\begin{bmatrix} \frac{n+1}{2n} & \frac{n-1}{2n} & 0\\ \frac{1}{2n} & \frac{1}{2} & \frac{n-1}{2n}\\ 0 & \frac{1}{2n} & \frac{2n-1}{2n} \end{bmatrix}$$

Calculating the stationary distribution  $\pi_0$  of this auxiliary chain yields that  $\pi_0(d = 1) = \frac{1}{n^2}$ . Then, we apply Chebyshev's inequality

**Theorem 3.7.** Let  $H_n$  be the *n*-dimensional hypercube and remove a set *S* consisting of  $\frac{2^{\frac{n}{2}}}{n}$  vertices chosen uniformly at random. Then, the upper bound on mixing time is given by

$$t_{mix}(\varepsilon) \le n \log n + \alpha(\varepsilon)n \tag{32}$$

where  $\alpha(\varepsilon)$  is a constant depending only on  $\varepsilon$ .

*Proof.* We construct a coupling of two chains  $(X_t)$  and  $(Y_t)$  as follows: if  $(X_t)$  has its *i*-th bit updated at time step t, then switch the *i*-th bit of the chain  $(Y_t)$  to the same bit, if that is possible; otherwise, the chain  $(Y_t)$  stays in the same position.

For the usual metric  $\rho$  on graphs and the given coupling, when both chains are not next to blocked vertices, the distance is contracted on average at least by a factor of  $(1 - \frac{1}{n})$ .

When one of the chains is next to a blocked vertex, the distance is contracted on average at least by a factor of  $(1 - \frac{1}{2n})$ . Then, since previous lemma asserts that we spend at most the fraction of  $\frac{1}{n}$  time near blocked vertices,

$$\mathbb{E}_{x,y}(\rho(X_1, Y_1)) \le \rho(x, y) \left(1 - \frac{1}{n} + \frac{1}{2n^2}\right)$$
(33)

Finally, note that

$$\begin{aligned} \left\| P^{t}(x,\cdot) - P^{t}(y,\cdot) \right\|_{TV} &\leq \mathbb{P}_{x,y}(X_{t} \neq Y_{t}) \\ &\leq \mathbb{P}_{x,y}(\rho(X_{t},Y_{t}) \geq \varepsilon) \leq \frac{\mathbb{E}_{x,y}(\rho(X_{t},Y_{t}))}{\varepsilon} \\ &\leq \frac{\operatorname{diam}(\Omega)}{\varepsilon} \left( 1 - \frac{1}{n} + \frac{1}{2n^{2}} \right)^{t} \leq \frac{n}{\varepsilon} e^{\frac{(2n-1)t}{2n^{2}}} \end{aligned}$$
(34)

and taking  $n \to \infty$  and  $t = n \log n + \alpha(\varepsilon)n$  yields us the desired bound.

### 4 Two Walkers

#### 4.1 Cycle

Our random walk consists of two walkers  $X_t^1$  and  $X_t^2$  on  $\mathbb{Z}_n$ , such that at each time step we choose one of the walker with probability  $\frac{1}{2}$  and with probability  $\frac{1}{4}$  walk step on that walker to one of its neighboring vertices or stay with probability  $\frac{1}{2}$ , and if we happen to move the walker into an occupied vertex, walkers swap the positions. Intuitively, it is somewhat similar to the previous problems in that we have some blocked vertices for the walkers, but now they are dependent on the position of the both walkers. To tackle this issue, we note that the state space of this random walk is  $Z_n^2 \setminus S$ , where  $S = \{(x, y) \in \mathbb{Z}_n^2 : x = y\}$ . Therefore, we can reformulate this problem as a random walk on a torus with set S removed.

**Theorem 4.1.** For the given random walk of two walkers on  $\mathbb{Z}_n$ 

$$n^2/32 \le t_{mix}(\varepsilon) \le n^2/4 \tag{35}$$

*Proof.* So, as was previously noted, we can consider the dynamics of a single walker on  $\mathbb{Z}_n^2 \setminus S$  instead. Note that the coordinates (x, y) can be parametrized as (x, d(x, y)), where d(x, y) is the signed distance between the two points. The properties of our walk guarantee that d(x, y) is never 0 or n. Moreover, we claim that  $(X_t, X_t - Y_t)$ , where  $X_t$  and  $Y_t$  are the positions of two walkers, is a Markov chain and we give explicit transition probabilities.

Indeed, first note that  $(X_t)$  behaves like a simple lazy random walk on  $\mathbb{Z}_n$  with laziness coefficient  $\frac{3}{4}$ . Additionally,  $D_t = X_t - Y_t$  behaves like a simple lazy random walk on the cycle  $\{1, 2, \ldots, n-1\}$ . To see this, assume that  $X_t = x$  and  $Y_t = y$ , where x > y, then for  $d \notin \{1, n-1\}$ 

$$\mathbb{P}(D_{t+1} = d+1|D_t = d) = \mathbb{P}(X_t = x+1|X_t = x) + \mathbb{P}(Y_t = y-1|Y_t = y) = \frac{1}{2}$$
  

$$\mathbb{P}(D_{t+1} = d-1|D_t = d) = \mathbb{P}(X_t = x-1|X_t = x) + \mathbb{P}(Y_t = y+1|Y_t = y) = \frac{1}{2}$$
  

$$\mathbb{P}(D_{t+1} = 1|D_t = 1) = \mathbb{P}(D_{t+1} = n-1|D_t = n-1) = \frac{1}{2}$$
(36)

We construct the following coupling of two Markov chains  $M_t^1$  and  $M_t^2$ , as follows:

#### 4.2 Hypercube

Here we consider a random walk, which consists of two walkers  $X_t$  and  $Y_t$  on  $H_n$ , where  $H_n$  is *n*-dimensional hypercube, such that at each time step we flip a coin and choose a walker, which we move to one of the neighboring vertices with uniform probability, and if we happen to move into occupied vertex, walkers swap their positions.

**Theorem 4.2.** For the given random walk of two walkers on  $H_n$ 

$$2n\log 2n + \alpha_2(\varepsilon) \le t_{mix}(\varepsilon) \le 2n\log 2n + \alpha_1(\varepsilon)$$
(37)

*Proof.* To prove this statement, we construct a projection chain of this random walk. First, note that analogously to the previous problem, we can consider single particle dynamics on  $H_n \times H_n = H_{2n}$  with the set  $S = \{aa \in H_{2n} : a \in H_n\}$  removed. The edge set is defined by the following relation:  $u = (u_1, u_2, \dots, u_{2n})$  and  $v = (v_1, v_2, \dots, v_{2n})$  are connected, if first

half of u differs from the first half of v by one bit and second halves are identical, or second half of u differs from the second half of v by one bit while first halves are identical, or first half of u is the second half of v and vice versa and the Hamming distance between first half of u and second is 1.

Thus we claim that if  $W(X_t) = \sum_{i=1}^{2n} X_t^i$  and  $d(X_t) = d_H(A(X_t), B(X_t))$ , where A(x) and B(x) are first and second halves of the string x, and  $d_H$  is Hamming distance function, then  $(W(X_t), d(X_t))$  is a projection chain. Indeed,

To obtain the lower bound, one can consider the Hamming weight as a statistic, and since it behaves like an Ehrenfest urn, the lower bound is immediate.

To obtain the upper bound, we can notice that

## **5** Convergence of Conditional Distribution

In this section we attempt to understand mixing time of the joint distribution of two arbitrary cards in the random transposition shuffle through understanding the speed of convergence of the conditional distributions. Intuitively, if  $X_t$  and  $Y_t$  are the positions of the first and second card respectively, one has:

$$\mathbb{P}(X_t = x, Y_t = y) = \mathbb{P}(X_t = x) \times \mathbb{P}(Y_t = y | X_t = x)$$
(39)

Our aim is to investigate convergence of both of the distributions on the RHS.

**Theorem 5.1.** Let  $\mu$  be the uniform distribution and P be the conditional distribution of the particle Y at time t conditioned on the position of the particle X at time t, then

$$\|P^{t}(x,\cdot) - \mu\|_{TV} \le Ce^{t/(n-2)}$$
 (40)

# References

- [LP] Levin, D.A, Peres, Y. Markov Chains and Mixing Times
- [AD] Aldous, D. and P. Diaconis. 1986. *Shuffling cards and stopping times*, Amer. Math. Monthly 93, no. 5, 333â348.

# Cauchy-Fueter Formulas for Universal Clifford Algebras of Indefinite Signature

### Ely Sandine<sup>\*</sup>

Advised by Prof. Matvei Libine

#### Abstract

We extend constructions of standard Clifford Analysis to the case of indefinite, nondegenerate signature. We define (p, q)-monogenic functions by means of Dirac operators that factor a (p, q)-wave operator. We prove two different Cauchy-Fueter integral formulas for these functions. The two formulas arise from dealing with singularities in separate ways, and are inspired by the methods of [L]. These theorems indicate the merit of these methods for dealing with singularities.

We also include a brief discussion of conformal mappings on  $\mathbb{R}^{p+1,q}$  acting on sets of (p,q)-monogenic functions. We provide a group action of O(p,q) on monogenic functions which map into the associated Clifford Algebra, as well as a computation for the pullback of the form  $D_{p,q}x$  under a conformal inversion.

### 1 Introduction

Given the ubiquity of Complex Analysis, considering possible extensions and generalizations leads to several theories. A modern introduction to Quaternionic Analysis is presented in [Su]. As the quaternions are non-commutative, there are two families of quaternionic valued functions that are analogous to the holomorphic functions of Complex Analysis, termed *left-regular* and *right-regular*. These satisfy the Cauchy-Riemann-Fueter equations

$$0 = \frac{\partial f}{\partial x_0} + i\frac{\partial f}{\partial x_1} + j\frac{\partial f}{\partial x_2} + k\frac{\partial f}{\partial x_3} \text{ and } 0 = \frac{\partial f}{\partial x_0} + \frac{\partial f}{\partial x_1}i + \frac{\partial f}{\partial x_2}j + \frac{\partial f}{\partial x_3}k$$

respectively. The left-regular functions satisfy the following identity, with C being a three dimensional contour in  $\mathbb{H}$  that wraps around  $q_0$  once, and Dq a fixed quaternionic valued three-form:

$$\frac{1}{2\pi^2} \int_C \frac{(q-q_0)^{-1}}{|q-q_0|^2} Dqf(q) = f(q_0).$$

<sup>\*</sup>This research was made possible by the Indiana University Bloomington Math REU Program, funded by NSF Award #1757857. We would like to thank Prof. Chris Connell for organizing and facilitating the wonderful program. We would also like to thank Ms. Mandie McCarthy for her administrative work, the various professors who gave talks, and the other REU students for their company.

This is termed the Cauchy-Fueter Integral Formula, and is analogous to the complex Cauchy Integral Formula, with Dq corresponding to idz. There is a similar formula for right regular functions. The proof of the Cauchy-Fueter Formula relies primarily on Stokes' Theorem, and that the function  $(x_0^2 + x_1^2 + x_2^2 + x_3^2)^{-1}$  is harmonic at all points except the origin.

A natural extension beyond the quaternionic case is analysis on real Universal Clifford Algebras. When constructed from a quadratic space with non-degenerate form, the algebras are of dimension  $2^{p+q}$ , where (p,q) is the signature of a quadratic form, with  $\mathbb{R}$ ,  $\mathbb{C}$ , and  $\mathbb{H}$ (the algebra of quaternions) corresponding to the cases of signature (0,0), (1,0), and (2,0). Clifford Analytic functions, or monogenic functions, are the proper analogue of holomorphic functions and regular functions. An introduction can be found in [G], and a proof of the Cauchy-Fueter Formula for the case of definite signature (p,0) is exhibited, for instance, in [GM]. In this case, the main player is the harmonic function  $(x_1^2 + \cdots + x_p^2)^{-\frac{p}{2}}$ . In [R], functions mapping to complex Universal Clifford Algebras of non-degenerate form over  $\mathbb{C}$ are studied, leading to a theory of Complexified Clifford Analysis and corresponding integral formulas.

Another direction is taken in [L], in which the real split quaternions are studied. In this case, solutions of the modified wave equation  $\frac{\partial^2 f}{\partial x_0^2} - \frac{\partial^2 f}{\partial x_1^2} - \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_2^2}$  are related to regular functions, and  $(x_0^2 - x_1^2 - x_2^2 + x_3^2)^{-1}$  is used in the proof. This leads to a situation in which the set of singularities, the null cone, is no longer a point, and intersects open sets containing the origin. This issue is resolved in two distinct ways, leading to separate integral formulas. In the first method, the author considers the *complexified* Split Quaternions, also known as biquaternions, in which coefficients on the four basis vectors, 1, *i*, *j*, and *k*, are elements of  $\mathbb{C}$ , not just  $\mathbb{R}$ . This algebra contains both the split quaternions and the classical quaternions as subspaces. By considering the holomorphic extension of a regular function into a complexified vector space, and deforming the contour of integration into the classical quaternions, it is shown that for all  $\epsilon$  sufficiently small, a suitable integral about an  $\epsilon$ -deformed contour is equal to the value of that function evaluated at a given point. In the second method, the author considers a single parameter family of integrals, in which a purely imaginary term is added to the denominator to prevent singularities, and shows that the limit as this term approaches zero converges to the desired value.

In this work, we begin by recalling the construction of Universal Clifford Algebras from non-degenerate quadratic forms. These constructions are done in both the real and complex cases. Following this, we define Dirac operators, which are used to define left-monogenic and right-monogenic functions. We then describe the monogenic functions that appear in the formulas, and prove lemmas regarding the key differential form  $D_{p,q}x$ . We prove many of these results in the complex case, and then restrict attention to the real Clifford Algebras in order to smooth our transition from real to complex vector spaces in the proof of Theorem 20. We first present a proof of Cauchy's Integral Formula in the case of a definite quadratic form, Theorem 16, to establish notation and for reference when proving the later formulas. We adapt the proofs of the main theorems of [L] to prove the existence of two integral formulas, Theorems 20 and 23. In the first case, we use Complexified Clifford Analysis, as described by [R], along with several homotopies, which allow us to deform the manifolds of integration into a real Clifford algebra of definite signature, and apply Theorem 16. The second formula uses a direct limiting argument to establish the result up to a constant, which is then determined by application of the first formula. In a larger context, Theorems 20 and 23 suggest that the methods of handling singularities presented in [L] are quite general.

In the final section, we consider certain spaces of monogenic functions, and how they are preserved under conformal maps. The group of conformal mappings on  $\mathbb{R}^{p+1,q}$  with respect to the quadratic form (p+1,q), for  $p+q \ge 1$ , is generated by translations, scalings, orthogonal transformations, and an inversion (for introduction, see, for example [Sc]). In the cases of Complex and Quaternionic Analysis, the conformal transformations on  $\mathbb C$  and  $\mathbb H$  lead to group actions on the spaces of holomorphic and regular functions. In the complex case, the action is through the fractional linear (Möbius) transformations, and the quaternionic case is discussed, for example, in [Su]. The actions of conformal transformations on  $\mathbb{R}^{p,q}$  lift to actions on solutions to associated wave equation ( $[K\emptyset]$ ), which are related to minimal representations of the group. This motivates the study of conformal mappings acting on monogenic functions. We observe that (p,q)-monogenic functions are preserved under translations and scalings, and provide a group action of O(p,q) acting on the space of monogenic functions that map into the associated Clifford Algebra. We finally include a calculation of the pullback of  $D_{p,q}x$  under the inversion action. Analogies with previous cases, along with this pullback computation, suggest the possibility of a similar group action to (21), although such an action is not proved.

### 2 Clifford Algebra Construction and Conventions

In this section we establish the algebraic environment. We begin by constructing Universal Clifford Algebras from real quadratic vector spaces with non-degenerate form. We employ a similar construction in the complex case, and observe that there is a natural algebra homomorphism which allows us to view real Universal Clifford Algebras as subalgebras of complex Clifford Algebras. We then define two conjugations on a subspace of the complex Universal Clifford Algebra, and describe their real analogues that arise via the homomorphism.

Let V be an n-dimensional real vector space, and Q a quadratic form on V. By the process of diagonalization, we have that there exist integers p, q and orthogonal basis  $e_1, \dots, e_n$  of (V, Q), such that

$$Q(e_j) = \begin{cases} 1 & 1 \le j \le p \\ -1 & p+1 \le j \le p+q \\ 0 & p+q < j \end{cases}$$

By Sylvester's Law of Inertia, we have that p and q are independent of basis chosen. The ordered pair (p,q) is the signature of (V,Q). If p + q = n, we say Q(x) is a non-degenerate quadratic form. From this point on we restrict our attention to non-degenerate forms. If Q(x) is such a form with q = 0 or p = 0, we call it a positive or negative definite quadratic form respectively. We consider a quotient of the tensor algebra over V, where (S) is the ideal generated by the elements of S,

$$\bigotimes V = \mathbb{R} \oplus V \oplus (V \otimes V) \oplus \cdots, \qquad S = \{v \otimes v + Q(v) : v \in V\}, \qquad A_Q = \bigotimes V/(S).$$
(1)

 $A_Q$ , constructed in this manner, is the Universal Clifford Algebra associated to (V, Q). We note the sign convention chosen for elements of the ideal is not standard in the literature,

resulting in a possible interchange of p and q. We note that for  $v_1$  and  $v_2$  orthogonal elements of (V, Q),

$$v_1 \otimes v_2 + v_2 \otimes v_1 = (v_1 + v_2) \otimes (v_1 + v_2) - v_1 \otimes v_1 - v_2 \otimes v_2 = -Q(v_1 + v_2) + Q(v_1) + Q(v_2) = 0.$$
(2)

We thus have that  $A_Q$  is a finite-dimensional algebra over  $\mathbb{R}$  generated by  $e_1, \dots, e_n$ , and let  $A_{p,q}$  be correspond to the  $A_Q$  constructed from (V,Q) where Q has signature (p,q). We will let  $\{e_1, e_2, \dots, e_p, \tilde{e}_{p+1}, \dots, \tilde{e}_{p+m}\}$  be an orthogonal basis of V such that  $Q(e_j) = 1$ ,  $Q(\tilde{e}_j) = -1$  for all applicable j. In the positive definite case we let  $A_n = A_{n,0}$ , and have that the basis of V is given by  $\{e_1, e_2, \dots, e_n\}$ . We consider finite sets of the type B = $\{i_1, i_2, \dots, i_n\} \subseteq \{1, \dots, n\}$  with  $i_j < i_{j+1}$ , and define

$$e_B = e_{i_1 i_2 \cdots i_j} = e_{i_1} \otimes e_{i_2} \cdots \otimes \tilde{e}_{i_n}.$$

$$\tag{3}$$

We let  $e_{\emptyset} = e_0$  be the multiplicative identity for the algebra. We have that if  $\Omega = \{1, 2, \dots, n\}$ ,  $\{e_B : B \subseteq \Omega\}$  forms a basis of  $A_{p,q}$  over  $\mathbb{R}$ .  $A_{p,q}$  is a Universal Clifford Algebra, with the property that, for every linear transformation  $M : V \to V$ , there exists a unique algebra homorphism  $\tilde{M}: A_{p,q} \to A_{p,q}$  such that  $\tilde{M}(e_j) = M(e_j)$  for  $1 \leq j \leq p+q$ , and  $\tilde{M}(e_0) = e_0$ . We will be especially concerned with the subspace  $\mathbb{R} \oplus V$ , which we identify with  $\mathbb{R}^{p+q+1} \subset A_{p,q}$ . We thus obtain the following relations for  $1 \leq i, j \leq n$  with  $i \neq j$ ,

$$e_0^2 = e_0, \qquad e_0 e_i = e_i e_0 = e_i, \qquad e_i e_j = -e_j e_i, \qquad e_i^2 = -e_0, \qquad \tilde{e}_j^2 = e_0.$$
 (4)

We thus have that we can write any  $x \in \mathbb{R}^{p+q+1}$  as  $x = \sum_{j=0}^{p} x_j e_j + \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j$ . Specifically, in the positive definite case, for  $x \in A_n$ , we can express  $x = \sum_{j=0}^{n} x_j e_j$ .

We can also construct a complex Universal Clifford Algebra from a quadratic space over  $\mathbb{C}$ . We let  $V^{\mathbb{C}}$  be an *n*-dimensional complex vector space, with quadratic form Q. In this case, we have that that by diagonalization, there exists an orthogonal basis  $e_1, \ldots, e_n$ , and integer p such that  $Q(e_j) = 1$  if  $1 \leq j \leq p$ , and  $Q(e_j) = 0$  otherwise. We say that  $V^{\mathbb{C}}$  is non-degenerate if p = n, and only consider this case. We perform a similar construction to (1), with  $V^{\mathbb{C}}$  now being a vector space over  $\mathbb{C}$ ,

$$\bigotimes V^{\mathbb{C}} = \mathbb{C} \oplus V^{\mathbb{C}} \oplus (V^{\mathbb{C}} \otimes V^{\mathbb{C}}) \oplus \cdots, \qquad S = \{ v \otimes v + Q(v) : v \in V^{\mathbb{C}} \}, \qquad A_Q^{\mathbb{C}} = \bigotimes V^{\mathbb{C}} / (S).$$

We denote the resultant algebra by  $A_n^{\mathbb{C}}$ . Analogues of (2), (3) and (4) hold by the same proofs.

We can also consider  $A_n^{\mathbb{C}}$  as an  $\mathbb{R}$  algebra generated by  $e_1, e_2, \ldots, e_n, ie_1, ie_2, \ldots, ie_n$ . We note that the natural inclusion  $\iota : A_{p,q} \to A_{p+q}^{\mathbb{C}}$  defined on the generators by  $\iota(e_j) = e_j$ ,  $\iota(\tilde{e}_j) = ie_j$  for  $0 \leq j \leq n$  extends to an injective  $\mathbb{R}$ -algebra homomorphism, so we can consider  $A_{p+q}$  as a unital real subalgebra of  $A_{p+q}^{\mathbb{C}}$ . We identify the linear span of  $e_0, e_1, \cdots, e_n$ over  $\mathbb{C}$  with  $\mathbb{C}^{n+1} \subset A_n(\mathbb{C})$ . We define the *Clifford conjugation* on  $\mathbb{C}^{n+1}$  by mapping

$$z = z_0 e_0 + \sum_{j=1}^n z_j e_j \to z^+ = z_0 e_0 - \sum_{j=1}^n z_j e_j.$$

We also have *complex conjugation*, defined component-wise

$$z = z_0 e_0 + \sum_{j=1}^n z_j e_j \to \bar{z} = \bar{z_0} e_0 + \sum_{j=1}^n \bar{z_j} e_j.$$

The fixed points of these actions are the span of  $e_0$ , identified with  $\mathbb{C}$ , and the span of  $e_0, \ldots, e_p$ , identified with  $\mathbb{R}^{p+1}$ , and the actions can be viewed as reflecting over the respective spaces. These two operations commute, and lead to two useful quadratic forms. The first such form is

$$N(z) = zz^{+} = z^{+}z = \sum_{j=0}^{n} z_{j}^{2}.$$

We note that N(z) is complex valued, and multiplicative. Via polarization, we get the bilinear form

$$\langle z, w \rangle = \frac{z^+ w + z w^+}{2} = \sum_{j=0}^n z_j w_j.$$

We let  $\mathcal{N}_{\mathbb{C}} = \{z \in \mathbb{C}^{n+1} : N(z) = 0\}$ , and have for all  $z \in \mathbb{C}^{n+1} \setminus \mathcal{N}_{\mathbb{C}}$ , z is invertible with inverse given by  $N(z)^{-1}z^+$ . We also consider the following form, which is real valued, and usually not multiplicative:

$$||z||^2 = z(\bar{z})^+ = \bar{z}z^+ = \sum_{j=0}^n |z_j|^2.$$

We consider  $A_{p,q}$  as a subalgebra of  $A_{p+q}^{\mathbb{C}}$ , and describe the restrictions of these conjugations and quadratic forms to  $\mathbb{R}^{p+q+1} \subset A_{p+q}$ . We have

$$\begin{aligned} x &= x_0 e_0 + \sum_{j=1}^p x_j e_j + \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j \to x^+ = x_0 e_0 - \sum_{j=1}^p x_j e_j - \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j, \\ x &= x_0 e_0 + \sum_{j=1}^p x_j e_j + \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j \to \bar{x} = x_0 e_0 + \sum_{j=1}^p x_j e_j - \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j, \\ N(x) &= x x^+ = x^+ x = \sum_{j=0}^p x_j^2 - \sum_{j=p+1}^{p+q} \tilde{x}_j^2, \text{ and} \\ \|x\|^2 &= x(\bar{x})^+ = \bar{x} x^+ = \sum_{j=0}^p x_j^2 + \sum_{j=p+1}^q \tilde{x}_j^2. \end{aligned}$$

From the quadratic form N(x), we obtain the bilinear form

$$\langle x, y \rangle = \frac{1}{2}(x^+y + xy^+) = \sum_{j=0}^p x_j y_j - \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{y}_j.$$

In line with the complex case, we consider with  $\mathcal{N}_{p,q} = \{x \in \mathbb{R}^{p+q+1} : N(x) = 0\}$ . For all  $x \in \mathbb{R}^{p+q+1} \setminus \mathcal{N}_{p,q}$ , we have that x is invertible with inverse given by  $N(x)^{-1}x^+$ . We note that  $\mathcal{N}_{n,0} = \{0\}$ .

# 3 Dirac Operators, Monogenic Functions, and Green's Function

We begin this section with the definition of complex Clifford analogues of the classical Dirac Operators. We next define monogenic functions over complex Clifford Algebras and a Green's Function, as dealt with by [R] in the case of definite signature. As was done with the conjugations and forms of the previous section we, using the natural inclusion, restrict these differential operators to define their analogues for real Clifford Algebras. These constructions are used in the statements and proofs of the Cauchy-Fueter formulas. We apply these operators to solutions of the (p,q) wave equation to obtain special monogenic functions, termed Green's Functions, which play a role in our formulas, analogous to that of  $z^{-1}$  in the classical Cauchy Integral Formula.

**Definition 1.** We let  $U \subseteq \mathbb{C}^{n+1}$  be an open set, and  $M_n^{\mathbb{C}}$  a left  $A_n^{\mathbb{C}}$  module. We let  $f: U \to M_n^{\mathbb{C}}$  be a holomorphic function, and define

$$\nabla_{\mathbb{C}}^{+}f = e_{0}\frac{\partial f}{\partial z_{0}} + \sum_{j=1}^{n} e_{j}\frac{\partial f}{\partial z_{j}}, \text{ and}$$
$$\nabla_{\mathbb{C}}f = e_{0}\frac{\partial f}{\partial z_{0}} - \sum_{j=1}^{n} e_{j}\frac{\partial f}{\partial z_{j}}.$$

We similarly can let  $\tilde{M}_n^{\mathbb{C}}$  be a right  $A_n^{\mathbb{C}}$  module,  $g: U \to \tilde{M}_n^{\mathbb{C}}$ , and define

$$g \nabla_{\mathbb{C}}^{+} = \frac{\partial g}{\partial z_{0}} e_{0} + \sum_{j=1}^{n} \frac{\partial g}{\partial z_{j}} e_{j}, \text{ and}$$
  
 $g \nabla_{\mathbb{C}} = \frac{\partial g}{\partial z_{0}} e_{0} - \sum_{j=1}^{n} \frac{\partial g}{\partial z_{j}} e_{j}.$ 

We note that if we let  $\square_{\mathbb{C}}$  be the complex Laplacian, we obtain the factorizations

$$\nabla_{\mathbb{C}}\nabla_{\mathbb{C}}^{+}f = \nabla_{\mathbb{C}}^{+}\nabla_{\mathbb{C}}f = \sum_{j=0}^{n} \frac{\partial^{2}f}{\partial z_{j}^{2}} = \Box_{\mathbb{C}}f \qquad \text{and} \qquad g\nabla_{\mathbb{C}}\nabla_{\mathbb{C}}^{+} = g\nabla_{\mathbb{C}}^{+}\nabla_{\mathbb{C}} = \sum_{j=0}^{n} \frac{\partial^{2}g}{\partial z_{j}^{2}} = \Box_{\mathbb{C}}g.$$

**Definition 2.** A holomorphic function  $f : U \subseteq \mathbb{C}^{n+1} \to M_n^{\mathbb{C}}$ , with U open and  $M_n^{\mathbb{C}}$  a left  $A_{p,q}$  module, is complex-left-monogenic if  $\nabla^+ f = 0$ . Alternatively, a holomorphic function  $g : U \subseteq \mathbb{C}^{n+1} \to \tilde{M}_n^{\mathbb{C}}$ , with U open and  $\tilde{M}_{p,q}$  a right  $A_{p,q}$  module, is complex-right-monogenic if  $g\nabla^+ = 0$ .

This factorization leads to a natural method of constructing complex left and rightmonogenic functions. If  $\phi: U \to M_n^{\mathbb{C}}$  is complex-harmonic, with U and  $M_n^{\mathbb{C}}$  as above, then  $\nabla_{\mathbb{C}} f$  is complex-left-monogenic. Similarly, if  $\tilde{\phi}: U \to \tilde{M}_n^{\mathbb{C}}$  is complex-harmonic,  $\nabla_{\mathbb{C}} \tilde{\phi}$  is complex-right-monogenic.

We restrict these definitions to the subalgebra  $A_{p,q} \subset A_{p+q}^{\mathbb{C}}$ . We let  $U \subset \mathbb{R}^{p+q+1}$  be open, and have, if f is a  $C^1$  function  $U \to M_{p+q}$  (g is a  $C^1$  function  $U \to \tilde{M}_{p,q}$ ), which maps to a left (right)  $A_{p,q}$  module, labeled  $M_{p,q}$  ( $\tilde{M}_{p,q}$ ): Definition 3.

$$\begin{aligned} \nabla_{p,q}^{+}f &= e_{0}\frac{\partial f}{\partial x_{0}} + \sum_{j=1}^{p}e_{j}\frac{\partial f}{\partial x_{j}} - \sum_{j=p+1}^{p+q}\tilde{e}_{j}\frac{\partial f}{\partial \tilde{x}_{j}},\\ \nabla_{p,q}f &= e_{0}\frac{\partial f}{\partial x_{0}} - \sum_{j=1}^{p}e_{j}\frac{\partial f}{\partial x_{j}} + \sum_{j=p+1}^{p+q}\tilde{e}_{j}\frac{\partial f}{\partial \tilde{x}_{j}},\\ g\nabla_{p,q}^{+} &= \frac{\partial g}{\partial x_{0}}e_{0} + \sum_{j=1}^{p}\frac{\partial g}{\partial x_{j}}e_{j} - \sum_{j=p+1}^{p+q}\frac{\partial g}{\partial \tilde{x}_{j}}\tilde{e}_{j}, \text{ and}\\ g\nabla_{p,q} &= \frac{\partial g}{\partial x_{0}}e_{0} - \sum_{j=1}^{p}\frac{\partial g}{\partial x_{j}}e_{j} + \sum_{j=p+1}^{p+q}\frac{\partial g}{\partial \tilde{x}_{j}}\tilde{e}_{j}.\end{aligned}$$

We let  $\square_{p,q}$  be a wave operator on  $\mathbb{R}^{p+q+1}$ ,

$$\Box_{p,q}f(x) = \sum_{j=0}^{p} \frac{\partial^2 f}{\partial x_j^2} - \sum_{j=p+1}^{p+q} \frac{\partial^2 f}{\partial \tilde{x}_j^2} f.$$

Thus,

$$abla_{p,q} 
abla_{p,q}^+ f = 
abla_{p,q}^+ 
abla_{p,q} 
abla_{p,q} f = \Box_{p,q} f, ext{ and } g 
abla_{p,q} 
abla_{p,q}^+ = g 
abla_{p,q}^+ 
abla_{p,q} = \Box_{p,q} g.$$

**Definition 4.** A smooth function  $f : U \subset \mathbb{R}^{p+q+1} \to M_{p,q}$ , with U open and  $M_{p,q}$  a left  $A_{p,q}$  module, is (p,q)-left-monogenic if  $\nabla^+ f = 0$ . Alternatively, a smooth function  $g : U \subset \mathbb{R}^{p+q+1} \to \tilde{M}_{p,q}$ , with U open and  $\tilde{M}_{p,q}$  a right  $A_{p,q}$  module, is (p,q)-right-monogenic if  $g\nabla^+ = 0$ .

Using this definition, and the factorization of the wave equation, we can construct (p, q)monogenic functions by considering solutions to the (p, q)-wave operator. Specifically, if we have  $\phi : U \to M_{p,q}$ , with U and M as above, such that  $\Box_{p,q}\phi = 0$ , this implies  $\nabla_{p,q}f$  is (p,q)-left-monogenic. Likewise, if we have  $\tilde{\phi} : U \to \tilde{M}_{p,q}, \Box_{p,q}\tilde{\phi} = 0$ , then  $\tilde{\phi}\nabla_{p,q}$  is (p,q)right-monogenic. The following case is an important example of such a construction, in which we consider a Green's function on  $\mathbb{C}^{p+q+1}$ , and a suitable restriction to the Universal Clifford Algebra  $A_{p,q}$ .

**Definition 5.** For  $n \geq 2$ , we define the following function  $H_n(z)$ :  $\mathbb{C}^{n+1} \setminus \mathcal{N}_{\mathbb{C}} \to \mathbb{C}$  by

$$H_n(z) = \frac{1}{(\sum_{j=0}^n z_j^2)^{(n-1)/2}} = \frac{1}{(N(z))^{(n-1)/2}}.$$

We note that if n is even, this may not be well defined. In this case, we let  $\mathbb{C}_{G}^{n+1} = \mathbb{C}^{n+1} \setminus \{z \in \mathbb{C}^{n+1} : N(z) \in \mathbb{R}, N(z) \leq 0\}$ , and define  $z^{\frac{1}{2}} : \mathbb{C}_{G} \to \mathbb{C}^{n+1}$  to map z to the unique w such that  $\operatorname{Re} w > 0$ ,  $w^{2} = z$ . As a consequence,  $H_{n}(z)$  mapping  $\mathbb{C}_{G}^{n+1} \to \mathbb{C}$  is well defined. From this point on we restrict the domain of  $H_{n}(z)$  to  $\mathbb{C}_{G}^{n+1}$  and do not distinguish the parity of n, as the same proofs hold in either case. We note that on this domain, we have that  $H_{n}(z)$  is complex harmonic, and thus  $\nabla_{\mathbb{C}}H_{n}(z)$  and  $H_{n}(z)\nabla_{\mathbb{C}}$  are left and right-monogenic respectively. Thus, we obtain the function on  $\mathbb{C}_{G}^{n+1}$  that is left and right-monogenic.

Lemma 6.

$$G_n(z) = \nabla_{\mathbb{C}} H_n(z) = H_n(z) \nabla_{\mathbb{C}} = (1-n) \frac{z_0 - \sum_{j=1}^n z_j e_j}{(\sum_{j=0}^n z_j^2)^{(n+1)/2}} = (1-n) \frac{z^+}{N(z)^{(n+1)/2}},$$
$$G_n(z) \nabla_{\mathbb{C}}^+ = \nabla_{\mathbb{C}}^+ G_n(z) = \Box_{\mathbb{C}} H_n(z) = 0.$$

By restricting this function to  $\mathbb{R}^{p+q+1}$  as done with the operators before, we observe the following solution to the (p,q)-wave operator, and the corresponding (p,q)-left and (p,q)-right-monogenic function:

**Definition 7.** For all  $p + q \ge 2$ , we let  $\mathbb{R}^{p+q+1}_G = \mathbb{R}^{p+q+1} \setminus \{x \in \mathbb{R}^{p+q+1} : N(x) \le 0\}$ , and define the functions  $H_{p,q}(x) : \mathbb{R}^{p+q+1}_G \to \mathbb{R}$ ,  $G_{p,q}(x) : \mathbb{R}^{p+q+1}_F$  by

$$H_{p,q}(x) = \frac{1}{\left(\sum_{j=0}^{p} x_j^2 - \sum_{j=p}^{p+q} \tilde{x}_j^2\right)^{(p+q-1)/2}} = \frac{1}{N(x)^{(p+q-1)/2}}, \text{ and}$$

$$G_{p,q}(x) = \nabla_{p,q} H_{p,q}(z) = H_{p,q}(x) \nabla_{p,q}$$
  
=  $(1 - p - q) \frac{x_0 - \sum_{j=1}^p x_j e_j - \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j}{(\sum_{j=0}^p x_j^2 - \sum_{j=p+1}^{p+q} \tilde{x}_j^2)^{(p+q+1)/2}} = (1 - p - q) \frac{x^+}{N(x)^{(p+q+1)/2}}.$ 

Consequently,

$$G_{p,q}(x)\nabla_{p,q}^+ = \nabla_{p,q}^+ G_{p,q}(x) = \Box_{p,q} H_{p,q}(x) = 0.$$

### 4 Differential Forms

We construct the differential forms,  $D_n z$  and  $D_{p,q} x$ , that will allow us to state and prove the Cauchy-Fueter Formulas. We prove several lemmas regarding these forms.

**Definition 8.** We let  $dV_{\mathbb{C}}$  be the n+1 complex form on  $\mathbb{C}^{n+1}$  normalized so that

$$dV_{\mathbb{C}} = dz_0 \wedge dz_1 \wedge \dots \wedge dz_n, \qquad dV_{\mathbb{C}}(e_0, e_1, e_2, \dots, e_n) = 1.$$
(5)

We recall that in our restrictions to  $\mathbb{R}^{p+q+1}$ , we have, as  $\tilde{e}_j = ie_j$ , the following two cases for  $j \leq p$  and j > p

$$dz_j(x_j e_j) = x_j = dx_j(x_j e_j) \implies dz_j\Big|_{\mathbb{R}^{p+1,q}} = dx_j, \text{ and}$$
$$dz_j(\tilde{x}\tilde{e}_j) = \tilde{x}i = id\tilde{x}_j(\tilde{x}\tilde{e}_j) \implies dz_j\Big|_{\mathbb{R}^{p+1,q}} = id\tilde{x}_j.$$

Substituting these into (5), we have

$$dV_{p,q} = i^q (dx_0 \wedge \dots \wedge dx_p \wedge d\tilde{x}_{p+1} \wedge \dots \wedge d\tilde{x}_{p+q}), \quad dV_{p,q}(e_0, e_1, \dots, e_p, \tilde{e}_{p+1}, \dots, \tilde{e}_{p+q}) = i^q.$$
(6)

**Definition 9.** We let  $D_n z$  be the unique  $\mathbb{C}^{n+1}$  valued complex n-form on  $\mathbb{C}^{n+1}$  such that, for all  $(z_0, z_1, \cdots, z_n) \in \mathbb{C}^{n+1}$ , we have

$$\langle z_0, D_{p+q} z(z_1, z_2, \cdots, z_n) \rangle = dV_{\mathbb{C}}(z_0, z_1, z_2, \cdots z_n).$$
 (7)

We can express  $D_n z$  as a sum of n forms by substituting basis vectors into (7) yielding

$$D_n z = \sum_{j=0}^n (-1)^j e_j d\hat{z}_j,$$
(8)

where

$$d\hat{z}_j = dz_0 \wedge dz_1 \wedge \dots \wedge \widehat{dz_j} \wedge \dots dz_n.$$

**Definition 10.** We let  $D_{p,q}x$  be the restriction of  $D_nz$  to  $\mathbb{R}^{p+q+1}$ , so that we have for all  $(x_0, x_1, \cdots, x_p, \tilde{x}_{p+1}, \cdots, \tilde{x}_{p+q})$ 

$$\langle x_0, D_{p,q} x(x_1, x_2, \cdots, x_n) \rangle = dV_{p,q}(x_0, x_1, x_2, \cdots, x_n),$$

$$D_{p,q}x = i^{q} \left(\sum_{j=0}^{p} (-1)^{j} e_{j} d\hat{x}_{j} - \sum_{j=p+1}^{p+q} (-1)^{j} \tilde{e}_{j} d\hat{x}_{j}\right),$$

where

$$d\hat{x}_j = dx_0 \wedge \dots \wedge dx_p \wedge d\hat{x}_{p+1} \wedge \dots \wedge d\hat{x}_{j-1} \wedge d\hat{x}_{j+1} \wedge \dots \wedge d\hat{x}_{p+q}.$$

**Proposition 11.** We let g be a holomorphic function from the open set  $U \subseteq \mathbb{C}^n$  to the right  $A_n^{\mathbb{C}}$  module  $\tilde{M}_n^{\mathbb{C}}$  and f be a holomorphic map from U to a left  $A_n^{\mathbb{C}}$  module  $M_n^{\mathbb{C}}$  such that there exists a product  $\cdot$  defined on  $\tilde{M}_n^{\mathbb{C}} \times M_n^{\mathbb{C}}$  (for instance one of the modules is  $A_n^{\mathbb{C}}$ ). Then,

$$d(gD_nzf) = (g\nabla_{\mathbb{C}}^+)fdV_{\mathbb{C}} + g(\nabla_{\mathbb{C}}^+f)dV_{\mathbb{C}}.$$

Proof. As  $dz_j \wedge d\hat{z}_j = (-1)^j dV$ ,

$$d(fD_nzg) = d(\sum_{j=0}^n (-1)^j ge_j d\hat{z}_j f)$$
  
=  $\sum_{j=0}^n (-1)^j (\frac{\partial g}{\partial z_j} e_j dz_j \wedge d\hat{z}_j f + ge_j dz_j \wedge d\hat{z}_j \frac{\partial f}{\partial z_j})$   
=  $\sum_{j=0}^n (-1)^{j+j} (\frac{\partial g}{\partial z_j} e_j f dV_{\mathbb{C}} + ge_j \frac{\partial f}{\partial z_j} dV_{\mathbb{C}})$   
=  $(g\nabla_{\mathbb{C}}^+) f dV_{\mathbb{C}} + g(\nabla_{\mathbb{C}}^+ f) dV_{\mathbb{C}}.$ 

By substituting in the constant 1 for g and f respectively

$$d(gD_nz) = (g\nabla_{\mathbb{C}}^+)dV_{\mathbb{C}}, \qquad d(D_nzf) = (\nabla_{\mathbb{C}}^+f)dV_{\mathbb{C}}.$$
(9)

**Corollary 12.** f is complex-left-monogenic on  $U \iff D_n z f$  is closed on U. g is complex-right-monogenic on  $U \iff gD_n z$  is closed on U.

We restrict these forms to real subspaces  $A_{p,q}$  to obtain an analog for  $C^1$  functions on  $\mathbb{R}^{p+q+1}$  mapping to appropriate modules.

$$d(fD_{p,q}xg) = (f\nabla_{p,q}^{+})gdV_{p,q} + f(\nabla_{p,q}^{+}g)dV_{p,q},$$

$$d(fD_{p,q}x) = (f\nabla_{p,q}^{+})dV_{p,q}, \qquad d(D_{p,q}xf) = (\nabla_{p,q}^{+}f)dV_{p,q}.$$
(10)

**Corollary 13.** f is (p,q)-left-monogenic on  $U \iff D_{p,q}xf$  is closed on U. g is (p,q)-right-monogenic on  $U \iff gD_{p,q}x$  is closed on U.

We consider  $\mathbb{R}^{p+q+1} \subset A_{p,q}$ , and an open set U contained in  $\mathbb{R}^{p+q+1}$ , with oriented boundary  $\partial U$ . At a point  $x \in \partial U$ , we let  $n_x = n_0 e_0 + n_1 e_1 + \cdots + n_p e_p + \tilde{n}_{p+1} \tilde{e}_{p+1} + \cdots + \tilde{n}_{p+q} \tilde{e}_{p+q}$ be a outward pointing normal unit vector to U at x. We let  $dS_{p,q}$  be the contraction of  $n_x$ with  $dV_{p,q}$ , so that we have for  $v_1, \ldots, v_n$  in the tangent space of  $\partial U$  at x,

$$dS_{p,q}(v_1, \dots, v_n) = dV_{p,q}(n_x, v_1, \dots, v_{p+q}),$$
  
$$dS_{p,q} = i^q \left(\sum_{j=0}^p (-1)^j n_j d\hat{x}_j + \sum_{j=p+1}^{p+q} (-1)^j n_j \hat{\tilde{x}}_j\right).$$

**Lemma 14.** With  $\bar{n}_x$  denoting the complex conjugate of the vector defined above,

$$D_{p,q}x\bigg|_{\partial U} = \bar{n}_x dS_{p,q}.$$

*Proof.* By definition  $n_x$  is orthogonal to the tangent space, so for every vector  $x = x_0e_0 + x_1e_1 + \cdots + x_pe_p + \tilde{x}_{p+1}\tilde{e}_{p+1} + \cdots + \tilde{x}_{p+q}\tilde{e}_{p+q}$  in the tangent space, we have that from our orthogonality relation that it satisfies

$$\sum_{j=0}^{p} n_j x_j + \sum_{j=p+1}^{p+q} n_j \tilde{x}_j = 0 \implies n_0 dx_0 + \dots + n_p dx_p + n_{p+1} d\tilde{x}_{p+1} + \dots + n_{p+q} d\tilde{x}_{p+q} = 0.$$

Without loss of generality, we suppose  $n_0 \neq 0$ , and so we can isolate  $dx_0$ , and substitute it into  $dV_{p,q}$ , and so we obtain on this tangent space, using the symbol  $d\tilde{x}$  to consider wedge products of all but  $dx_0$ ,  $dx_j$ ,

$$d\tilde{x}_{j} = dx_{1} \wedge dx_{2} \wedge \dots \wedge dx_{j-1} \wedge dx_{j+1} \wedge \dots \wedge d\tilde{x}_{p+q} \implies dx_{j} \wedge d\tilde{x}_{j} = (-1)^{j-1} d\hat{x}_{0},$$
  
$$dx_{0} = -\frac{1}{n_{0}} (n_{1} dx_{1} + n_{2} dx_{2} + \dots + n_{p} dx_{p} + n_{p+1} d\tilde{x}_{p+1} + \dots + n_{p+q} d\tilde{x}_{p+q}), \qquad (11)$$

$$dS_{p,q} = i^q (n_0 d\hat{x}_0 + \sum_{j=1}^p (-1)^j (-\frac{n_j^2}{n_0}) dx_j \wedge d\check{x}_j) + \sum_{j=p+1}^{p+q} (-1)^j (-\frac{n_j^2}{n_0}) d\check{x}_j \wedge d\check{x}_j)$$
$$= \frac{i^q}{n_0} (\sum_{j=0}^{p+q} n_j^2) d\hat{x}_0 = \frac{i^q}{n_0} d\hat{x}_0$$

We perform a very similar computation for  $D_{p,q}x$ , substituting in (11) to get

$$\begin{aligned} D_{p,q}x \bigg|_{\partial U} &= i^{q} (\sum_{j=0}^{p} (-1)^{j} e_{j} d\hat{x}_{j} - \sum_{j=p+1}^{p+q} (-1)^{j} \tilde{e}_{j} d\hat{\tilde{x}}_{j}) \\ &= i^{q} (\sum_{j=0}^{p} (-1)^{j} \frac{1}{n_{0}} n_{j} e_{j} dx_{j} \wedge d\check{x}_{j} - \sum_{j=p+1}^{p+q} (-1)^{j} \frac{1}{n_{0}} n_{j} \tilde{e}_{j} d\tilde{x}_{j} \wedge d\check{\tilde{x}}_{j}) = \frac{i^{q} \bar{n}_{x}}{n_{0}} d\hat{x}_{0} = \bar{n}_{x} dS_{p,q}. \end{aligned}$$

We let  $K_r$  and  $S_r$  be the boundaries of the sets  $\{x \in \mathbb{R}^{p+q+1} : N(x) \leq r\}$  and  $\{x \in \mathbb{R}^{p+q+1} : ||x||^2 \leq r\}$ , and note that the outward pointing normal vectors at x will be  $\frac{\bar{x}}{||x||}$  and  $\frac{x}{||x||}$  respectively, so we get the following corollary, analogous to Lemma 3 of [L].

Corollary 15.

$$D_{p,q}x\Big|_{K_r} = \frac{x}{\|x\|}dS, \qquad D_{p,q}x\Big|_{S_r} = \frac{\bar{x}}{\|x\|}dS = \bar{x}\frac{dS}{r}.$$

We note that these expressions are the same when q = 0.

### 5 Standard Cauchy-Fueter Formula

We present the Cauchy Integral Formula for Clifford Algebras of definite signature. A modern introduction can be found in [GM], in which monogenic functions are defined on  $\mathbb{R}^{p+q}$ , not  $\mathbb{R}^{p+q+1}$ , and different notation is used. We present this theorem and proof for clarity, and also for comparison with later formulas and proofs.

**Theorem 16.** Cauchy-Fueter Formula for Universal Clifford Algebras of Signature (n, 0): Let  $U \subset \mathbb{R}^{n+1} \subset A_n$  be an open bounded set with smooth boundary  $\partial U$  and f be a (n, 0)-leftmonogenic function defined on a neighborhood of  $\overline{U}$  for  $n \geq 2$ . We have, where  $\omega_n$  is the surface area of the unit n-sphere:

$$\int_{\partial U} G_{n,0}(x-x_0) D_{n,0} x f(x) = \begin{cases} (1-n)\omega_n f(x_0) & x_0 \in U \\ 0 & x_0 \notin \bar{U} \end{cases}.$$

If g is (n, 0)-right-monogenic:

$$\int_{\partial U} g(x) D_{n,0} x G_{n,0}(x - x_0) = \begin{cases} (1 - n) \omega_n g(x_0) & x_0 \in U \\ 0 & x_0 \notin \bar{U} \end{cases}.$$

*Proof.* We prove only the left-monogenic case, with the other case proceeding symmetrically. By translation, we can consider  $x_0$  to be fixed at zero. We note that as the form is definite,  $\mathbb{R}^{n+1}_G = \mathbb{R}^{n+1} \setminus \{0\}$ , and for  $x \in \mathbb{R}^{n+1} \setminus \{0\} \cap U$ 

$$d(G_{n,0}D_{n,0}xf) = (G_{n,0}\nabla_{n,0}^+)fdV_{n,0} + G_{n,0}(\nabla_{n,0}^+f)dV_{n,0} = 0.$$

Suppose  $0 \notin U$ . In this case,  $G_{n,0}D_{n,0}xf$  is a closed form on U, and so by Stokes' Theorem, the integral of it about  $\partial U$  will be zero. Now, if  $0 \in U$ , as U is open, we consider the ball  $B_r = \{x \in G_{n,0} : ||x|| \leq r\} \subset U$ . We have by Stokes' Theorem, applied to the manifold  $U \setminus B_{r,0}$ ,

$$\int_{\partial U} G_{n,0} D_{n,0} x f = \int_{\partial B_r} G_{n,0}(x) D_{n,0} x f.$$

We apply Corollary 15 and substitute the definition of  $G_{n,0}$  to continue,

$$\int_{\partial U} G_{n,0} D_{n,0} x f = \int_{\partial B_r} G_{n,0}(\frac{xf}{\|r\|} dS_{n,0})$$
$$= \int_{\partial B_r} (1-n) \frac{xx^+ f}{(N(x))^{(n+1)/2}} \frac{dS_{n,0}}{r} = \int_{\partial B_r} (1-n) \frac{f\|x\|^2 dS_{n,0}}{(N(x))^{(n+1)/2} r}.$$

We note that  $N(x) = ||x||^2 = r^2$ , split the integral into two, and recognize the second of the resultant integrals as that of a constant over a sphere.

$$= \frac{1-n}{r^n} \int_{\partial B_r} f(x) dS_{n,0} = \frac{1-n}{r^n} \left( \int_{\partial B_r} (f(x) - f(0)) dS_{n,0} + \int_{\partial B_r} f(0) dS_{n,0} \right)$$
$$= \frac{1-n}{r^n} \int_{\partial B_r} (f(x) - f(0)) dS_{n,0} + (1-n)\omega_n f(0). \quad (12)$$

To deal with the first integral, we use Cauchy-Swartz inequality, and have that if M is the sup of |f(x) - f(0)| on  $\partial B_r$ ,

$$\left|\frac{1-n}{r^n}\int_{\partial B_r} (f(x) - f(0))dS_{n,0}\right| \le \frac{n-1}{r^n}\int_{\partial B_{r,x_0}} |f(x) - f(x_0)||dS_{n,0}| \le \omega_n M(n-1).$$

Thus, as we let  $r \to 0$ , M goes to zero as f continuous, and so the first integral of (12) vanishes, yielding the theorem.

We next prove two analogues of this result for (p, q)-monogenic functions, theorems 20 and 23. These statements (and subsequent proofs) follow closely from Theorems 13 and 16 of [L].

### 6 First Cauchy-Fueter Formula Formulation

In this section we present an integral formula for (p, q)-monogenic functions. We begin by defining the one-parameter map  $h_{p,q,\epsilon}$  which will be used to state the integral formula. This map, along with  $l_{p,q,\epsilon}$ , will serve to change the contour of integration through  $\mathbb{C}^{p+q+1}$  from one contained in the real span of  $e_0, e_1, \ldots, e_p, \tilde{e}_{p+1}, \ldots, \tilde{e}_{p+q}$  to one contained in the real span of  $e_0, \ldots, e_p, e_{p+1}, \ldots, e_{p+q}$ , from whence the standard integral formula may be applied. **Definition 17.** We let  $h_{p,q,\epsilon} : \mathbb{C}^{p+q+1} \to \mathbb{C}^{p+q+1}$  for  $0 \le \epsilon \le 1$  be defined by

$$z = \sum_{j=0}^{p+q} z_j e_j \to z_{h,\epsilon} = \sum_{j=0}^{p} (1+i\epsilon) z_j e_j + \sum_{j=p+1}^{q+1} (1-i\epsilon) z_j e_j.$$

Corollary 18. If  $x = \sum_{j=0}^{p} x_j e_j + \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j \in \mathbb{R}^{p+q+1} \subset A_{p,q} \subset A_{p+q}^{\mathbb{C}}$ ,  $N(x_{h,\epsilon}) = (1 - \epsilon^2) N(x) + 2i\epsilon \|x\|^2,$ 

$$x_{h,1} = \sum_{j=0}^{p} (1+i)x_j e_j + \sum_{j=p+1}^{p+q} (1-i)x_j (ie_j) = \sum_{j=0}^{p} (1+i)x_j e_j + \sum_{j=p+1}^{p+q} (1+i)\tilde{x}_j e_j, \text{ and}$$
$$N(x_{h,1}) = 2i||x||^2.$$

**Definition 19.** We let  $h_{p,q,\epsilon,x_0}: x \to x_0 + h_{p,q,\epsilon}(x-x_0)$ .

**Theorem 20.** Cauchy-Fueter Formula for Universal Clifford Algebras of Signature (p,q): Let  $U \subset \mathbb{R}^{p+q+1} \subset A_{p,q}$  be an open bounded set with smooth boundary  $\partial U$ , and let f be a (p,q)left-monogenic function defined on a neighborhood of  $\overline{U}$ , mapping to a left  $A_{p,q}^{\mathbb{C}}$  module  $M_{p+q}^{\mathbb{C}}$ , with  $p+q \geq 2$ . Suppose f extends into a complex-left-monogenic function  $f^{\mathbb{C}}: W^{\mathbb{C}} \to M_{p+q}^{\mathbb{C}}$ , with  $W^{\mathbb{C}} \subset \mathbb{C}^{p+q+1} \subset A_{p+q}^{\mathbb{C}}$  an open subset containing  $\overline{U}$ . We have

$$\int_{(h_{p,q,\epsilon,x_0})_*(\partial U)} G_{p,q}(z-x_0) D_{p+q} z f^{\mathbb{C}}(z) = \begin{cases} (1-p-q)\omega_{p+q} f(x_0) & x_0 \in U \\ 0 & x_0 \notin \bar{U} \end{cases}$$

for all  $\epsilon > 0$  sufficiently close to 0. If g is right-monogenic mapping to a right  $A_{p+q}^{\mathbb{C}}$  module  $\tilde{M}_{p+q}^{\mathbb{C}}$ , which extends to a complex-right-holomorphic function  $g^{\mathbb{C}}: W^{\mathbb{C}} \to \tilde{M}_{p+q}^{\mathbb{C}}$ , we have

$$\int_{(h_{p,q,\epsilon,x_0})_*(\partial U)} g^{\mathbb{C}}(z) D_{p+q} z G_{p,q}(z-x_0) = \begin{cases} (1-p-q)\omega_{p+q} g(x_0) & x_0 \in U\\ 0 & x_0 \notin \bar{U} \end{cases}$$

We begin by giving an outline of the proof. As with the proof of the definite case, we will set  $x_0 = 0$  by translation, so that  $h_{p,q,\epsilon,x_0} = h_{p,q,\epsilon}$ , and consider only the left-monogenic case. We first consider  $(h_{p,q,\epsilon,x_0})_*(\partial U)$ . We note that if we consider the image of  $\mathbb{R}^{p+q+1}$  under  $h_{p,q,\epsilon}$ , denoted by  $(h_{p,q,\epsilon})_*(\mathbb{R}^{p+q+1})$ , we have for all  $x \neq 0$ ,  $N(x) = (1 - \epsilon^2)N(x) + 2i\epsilon ||x||$ . This is not a negative real, so we have  $(h_{p,q,\epsilon})_*(\mathbb{R}^{p+q+1} \setminus \{0\}) \in \mathbb{C}_G^{p+q+1}$ . Thus, we have that if  $\partial U$  is a contour in  $\mathbb{R}^{p+q+1}$  of a set as stated in the theorem, its image under  $h_{p,q,\epsilon}$ , denoted by  $(h_{p,q,\epsilon})_*(\partial U)$ , will be contained in region of  $\mathbb{C}_G^{p+q+1}$ , so our Green's function is well defined. We note that this implies if f is a complex-left-monogenic function on  $(h_{p,q,\epsilon})_*(U)$ , we have the following result by Corollary 12 applied to  $(h_{p,q,\epsilon})_*(U) \setminus (h_{p,q,\epsilon})_*(B_{p,q,r})$ , where  $(h_{p,q,\epsilon})_*(B_{p,q,r})$  is the image of  $B_{p,q,r} \subset U = \{x \in \mathbb{R}^{p+q+1} : ||x|| \leq r\}$  with boundary  $S_{p,q,r}$ ,

$$\int_{(h_{p,q,\epsilon})_*(\partial U)} G_{p+q}(z) D_{p+q} z f^{\mathbb{C}}(z) - \int_{(h_{p,q,\epsilon})_* S_r} G_{p+q}(z) D_{p+q} z f^{\mathbb{C}}(z)$$
$$= \int_{(h_{p,q,\epsilon})_*(U) \setminus (h_{p,q,\epsilon})_*(B_{p,q,r})} d(G_{p+q}(z) D_{p+q} z f^{\mathbb{C}}(z)) = \int_{(h_{p,q,\epsilon})_*(U) \setminus (h_{p,q,\epsilon})} 0 dV_{\mathbb{C}} = 0.$$
(13)

By varying the parameter from  $\epsilon$  to 1, we can continuously deform  $(h_{\epsilon})_*(S_{p,q,r})$  into  $(h_{p,q,1})_*(S_{p,q,r})$ in  $\mathbb{C}_G^{p+q+1}$  for any  $\epsilon \leq 1$ , and so we have

$$\int_{(h_{p,q,\epsilon})_*(\partial U)} G_{p+q}(z) D_{p+q} z f^{\mathbb{C}}(z) = \int_{(h_{p,q,1})_* S_r} G_{p+q}(z) D_{p+q} z f^{\mathbb{C}}(z)$$
(14)

We will move the deformed sphere into  $\mathbb{R}^{p+q+1} \subset A_{p+q,0} \subset A_{p+q}^{\mathbb{C}}$  by means of the following map.

**Definition 21.** Let  $l_{p,q,\epsilon} : \mathbb{C}^{p+q+1} \to \mathbb{C}^{p+q+1}$  for  $0 \le \epsilon \le 1$  be the rotation

$$z \to z_{l,\epsilon} = \frac{1 - i\epsilon}{\sqrt{1 + \epsilon^2}} z.$$

Corollary 22.

$$N(z_{l,\epsilon}) = \frac{1 - 2i\epsilon - \epsilon^2}{1 + \epsilon^2} N(z), \qquad ||z_{l,\epsilon}|| = ||z||.$$

We consider the composition of  $h_{p,q,\epsilon}$  and  $l_{p,q,\epsilon}$  applied to sphere  $S_{p,q,r}$ . If  $z = h_{p,q,1}(x)$  for some  $x \in S_{p,q,r}$ , we have.  $N(z) = 2i||x||^2$ . Therefore,  $N(z_{l,\epsilon})$  will not be a negative real for all  $0 \le \epsilon \le 1$ , which implies that the image of  $(h_{p,q,1})_*(B_{p,q,r})$  under  $l_{p,q,\epsilon}$  as we vary  $\epsilon$  from 0 to 1 will be contained in  $\mathbb{C}_G^{p+q+1}$ . We also note that the composition of both of these is orientation preserving, as seen by applying them to the basis vectors. We can now state the full argument.

*Proof.* Let  $M = \sup_{x \in \partial U} ||x||$ . We restrict  $W^{\mathbb{C}}$  to be the  $\delta$  neighborhood of  $\overline{U}$  for some  $\delta > 0$ . We consider  $0 < \epsilon < \frac{\delta}{M}$ . Thus, we have that  $(h_{\epsilon})_*(\partial U) \subseteq \mathbb{C}_G^{p+q+1}$  lies inside  $W^{\mathbb{C}}$ , and in this region, we have that the integrand is closed by Corollary 12, so the integral will be constant for all  $0 < \epsilon < \frac{\delta}{M}$ . If  $0 \notin \overline{U}$ , we have that  $(h_{\epsilon})_*(U) \subseteq \mathbb{C}_G^{p+q+1}$ , the form is closed, and so the integral about the boundary is zero, and we are done.

If not, we choose an r sufficiently small such that  $r < \frac{\delta}{2}$ , and  $S_r = \{x \in \mathbb{R}^{n+1} : ||x|| = r\}$  is contained in U, with the orientation given by it being the boundary of  $\{x \in \mathbb{R}^{n+1} : ||x|| \le r\}$ . We consider  $(h_{\epsilon})_*S_r$ , and we have by (13)

$$\int_{(h_{p,q,\epsilon})_*(\partial U)} G_{p+q} D_{p+q} z f^{\mathbb{C}} = \int_{(h_{p,q,\epsilon})_* S_r} G_{p+q} D_{p+q} z f^{\mathbb{C}}.$$

As  $r < \frac{\delta}{2}$ ,  $(h_{p,q,1})_* S_r$  is contained in  $W^{\mathbb{C}}$ , so we apply (14). By our mapping,  $l_{p,q,1}$ , we have that, where  $\tilde{S}_{r\sqrt{2}}$  is the sphere of radius  $r\sqrt{2}$  contained in  $\mathbb{R}^{p+q+1} \subset A_{p+m,0}$ ,

$$= \int_{(h_{p,q,1})_* S_r} G_{p+q}(z) D_{p+q} z f(z) = \int_{\tilde{S}_{r\sqrt{2}}} G_{p+q,0}(x) D_{p+q} x f^{\mathbb{C}}(x),$$

Thus, we have by Stokes' Theorem in  $\mathbb{C}_G^{n+1}$ , and the result for the definite case:

$$\int_{(h_{p,q,\epsilon})_*(\partial U)} G_{p,q} D_{p+q} z f^{\mathbb{C}} = \int_{\tilde{S}_{r\sqrt{2}}} G_{p+q} D_{p+q} z f^{\mathbb{C}}$$
$$= \int_{\tilde{S}_{r\sqrt{2}}} G_{p+q,0} D_{p+q,0} x f^{\mathbb{C}} = (1-p-q)\omega_{p+q} f(0).$$

### 7 Second Cauchy-Fueter Formula Formulation

We state and prove a second integral formula, one that does not change the contour of integration. In this case, a direct limiting argument is used, resulting in a longer proof. We first introduce a hybrid spherical coordinates system in which our integral simplifies. In this system, we are able to adapt the methods of [L] to prove the theorem up to a constant,  $C_{p,q}$ . We present two separate proofs in which we apply Theorem 20 to a constant function in order to establish the value of  $C_{p,q}$ .

**Theorem 23.** Let  $U \subset \mathbb{R}^{p+q+1}$  be a bounded open region with smooth boundary  $\partial U$ . Let  $f: U \to M_{p,q}$ , an  $A_{p,q}$  module, be a left-monogenic function with  $p + q \geq 2$ . Suppose  $\partial U$  intersects the cone  $\{x \in \mathbb{R}^{p+q+1} : N(x - x_0) = 0\}$  transversally. We have

$$\lim_{\epsilon \to 0} \int_{\partial U} G_{p,q,\epsilon}(x-x_0) \cdot D_{p,q} x f(x) = \begin{cases} (1-p-q)\omega_{p+q} f(x_0) & x_0 \in U\\ 0 & x_0 \notin \bar{U} \end{cases}$$

where  $G_{p,q,\epsilon}$  is the modified Green's function defined by

$$G_{p,q,\epsilon} = \frac{(1-p-q)(x_0 - \sum_{j=1}^p x_j e_j - \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j)}{(\sum_{j=0}^p x_j^2 - \sum_{j=p+1}^{p+q} \tilde{x}_j^2 + i\epsilon(\sum_{j=0}^p x_j^2 + \sum_{j=p+1}^{p+q} \tilde{x}_j))^{((p+q)+1)/2}} = \frac{(1-p-q)x^+}{(N(x) + i\epsilon \|x\|^2)^{\frac{(p+q)+1}{2}}}$$

Similarly, if g right-monogenic satisfying the same restraints and mapping to  $\tilde{M}_{p,q}$ , a right  $A_{p,q}$  module, we have:

$$\lim_{\epsilon \to 0} \int_{\partial U} g(x) \cdot D_{p,q} x \cdot G_{p,q,\epsilon}(x-x_0) = \begin{cases} (1-p-q)\omega_{p+q}g(x_0) & x_0 \in U\\ 0 & x_0 \notin \bar{U} \end{cases}$$

*Proof.* We consider the change of coordinates to hybrid spherical coordinates given by

$$\Phi(\theta, \phi_1, \dots, \phi_p, \rho, \psi_1, \dots, \psi_{q-1}) = (x_0, x_1, \dots, x_p, \tilde{x}_{p+1}, \dots, \tilde{x}_{p+q})$$

$$x_0 = \rho \cos \theta \cos \phi_1$$

$$x_1 = \rho \cos \theta \sin \phi_1 \cos \phi_2$$

$$x_2 = \rho \cos \theta \sin \phi_1 \sin \phi_2 \cos \phi_3$$

$$\vdots$$

$$x_{p-1} = \rho \cos \theta \sin \phi_1 \cdots \cos \phi_p$$

$$x_p = \rho \cos \theta \sin \phi_1 \cdots \sin \phi_p$$

$$\tilde{x}_{p+1} = \rho \sin \theta \cos \psi_1$$

$$\tilde{x}_{p+2} = \rho \sin \theta \sin \psi_1 \cos \psi_2$$

$$\vdots$$

$$\tilde{x}_{p+q-1} = \rho \sin \theta \sin \psi_1 \cdots \sin \psi_{q-1}$$

$$0 \le \phi_1, \phi_2, \dots, \phi_{p-1}, \psi_1, \psi_2, \dots, \psi_{q-2} \le \pi$$

$$0 \le \phi_p, \psi_{q-1} \le 2\pi$$

$$0 \le \theta \le \frac{\pi}{2}.$$
(15)

We will use these coordinates for proving the theorem, and will establish several results that motivate the choice of coordinates. In these coordinates, we have

$$N(x) = \sum_{j=0}^{p} x_j^2 - \sum_{j=p+1}^{p+q} \tilde{x}_j^2 = \rho^2 \cos^2 \theta - \rho^2 \sin^2 \theta = \rho^2 \cos 2\theta, \text{ and}$$
$$\|x\|^2 = \sum_{j=0}^{p} x_j^2 + \sum_{j=p+1}^{p+q} \tilde{x}_j^2 = \rho^2.$$

We thus have that the null cone  $\mathcal{N}_{p,q}$  is the set such that  $\theta = \frac{\pi}{4}$ , and we structure our argument in the vein of [L]. We use the symmetry of the change of basis matrix with respect to p and q-1 to calculate the determinant by means of block matrices.

**Lemma 24.** We let  $S_{n,\alpha}$  be the Jacobian matrix corresponding to the transformation into standard n dimensional spherical coordinates,  $(\rho, \alpha_1, \ldots, \alpha_n) \to (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$ ,

$$\det(S_{n,\alpha}) = \rho^n \sin^{n-1} \alpha_1 \sin^{n-2} \alpha_2 \cdots \sin \alpha_{n-1}$$

The determinant of the Jacobian matrix of the change of basis (15) is given by

$$\det(D\Phi) = -\rho \cos^p \theta \sin^{q-1} \theta \det(S_{p,\phi}) \det(S_{q-1,\psi}).$$

Proof. We let A be the  $(p+1) \times (p+1)$  matrix that is obtained from  $S_{p,\phi}$  by multiplying the first column by  $-\rho \sin \theta$  and multiplying the rest of the columns are by  $\cos \theta$ . We let the first column of A be  $\vec{a}$ . We let B be the  $(p+1) \times q$  matrix whose first column is  $\vec{b}$ , described below, and all other columns are zero. Similarly, we let  $\Gamma$  be the  $q \times (p+1)$  matrix whose first column is  $\vec{c}$  and the rest of the columns are zero. Finally, we let  $\Delta$  be  $S_{q-1,\psi}$ , where each entry is multiplied by  $\sin \theta$ , and we label its first column by  $\vec{d}$ . Expressing these definitions in terms of matrix multiplication, and computing the Jacobian matrix, we have the following

$$A = S_{p,\phi} \cdot \begin{pmatrix} -\rho \sin \theta & 0 & 0 & \dots \\ 0 & \cos \theta & 0 & \dots \\ 0 & 0 & \cos \theta & \dots \\ 0 & 0 & 0 & \ddots \end{pmatrix}, \qquad \Delta = S_{q-1,\psi} \cdot (\sin \theta)$$
$$\vec{a} = (-\rho \sin \theta) \cdot \begin{pmatrix} \cos \phi_1 \\ \sin \phi_1 \cos \phi_2 \\ \vdots \\ \sin \phi_1 \sin \phi_2 \cdots \cos \phi_p \\ \sin \phi_1 \sin \phi_2 \cdots \sin \phi_p \end{pmatrix}, \qquad \vec{b} = (\frac{-\cos \theta}{\rho \sin \theta}) \cdot \vec{a},$$
$$\vec{c} = (\rho \cos \theta) \begin{pmatrix} \cos \psi_1 \\ \sin \psi_1 \cos \psi_2 \\ \vdots \\ \sin \psi_1 \sin \psi_2 \cdots \cos \psi_{q-1} \\ \sin \psi_1 \sin \psi_2 \cdots \sin \psi_{q-1} \end{pmatrix}, \qquad \vec{d} = (\frac{\sin \theta}{\rho \cos \theta}) \vec{c}, \text{ and}$$

$$D\Phi = \begin{pmatrix} \frac{\partial x_0}{\partial \theta} & \frac{\partial x_0}{\partial \phi_1} & \frac{\partial x_0}{\partial \phi_2} & \cdots & \frac{\partial x_0}{\partial \phi_p} & \frac{\partial x_0}{\partial \rho} & \frac{\partial x_0}{\partial \phi_1} & \cdots & \frac{\partial x_0}{\partial \phi_{q-1}} \\ \frac{\partial x_1}{\partial \theta} & \frac{\partial x_1}{\partial \phi_1} & \frac{\partial x_1}{\partial \phi_2} & \cdots & \frac{\partial x_1}{\partial \phi_p} & \frac{\partial x_1}{\partial \rho} & \frac{\partial x_1}{\partial \phi_1} & \cdots & \frac{\partial x_1}{\partial \phi_{q-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_p}{\partial \theta} & \frac{\partial x_p}{\partial \phi_1} & \frac{\partial x_p}{\partial \phi_2} & \cdots & \frac{\partial x_{p+1}}{\partial \phi_p} & \frac{\partial x_p}{\partial \rho} & \frac{\partial x_{p+1}}{\partial \phi_1} & \cdots & \frac{\partial x_{p+1}}{\partial \phi_{q-1}} \\ \frac{\partial \tilde{x}_{p+1}}{\partial \theta} & \frac{\partial \tilde{x}_{p+1}}{\partial \phi_1} & \frac{\partial \tilde{x}_{p+1}}{\partial \phi_2} & \cdots & \frac{\partial \tilde{x}_{p+1}}{\partial \phi_p} & \frac{\partial \tilde{x}_{p+1}}{\partial \rho} & \frac{\partial \tilde{x}_{p+1}}{\partial \phi_1} & \cdots & \frac{\partial \tilde{x}_{p+1}}{\partial \phi_{q-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \tilde{x}_{p+q}}{\partial \theta} & \frac{\partial \tilde{x}_{p+q}}{\partial \phi_1} & \frac{\partial \tilde{x}_{p+q}}{\partial \phi_2} & \cdots & \frac{\partial \tilde{x}_{p+q}}{\partial \phi_p} & \frac{\partial \tilde{x}_{p+q}}{\partial \rho} & \frac{\partial \tilde{x}_{p+q}}{\partial \phi_1} & \cdots & \frac{\partial \tilde{x}_{p+q}}{\partial \phi_{q-1}} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} A & B \\ \Gamma & \Delta \end{pmatrix}.$$

**Lemma 25.** If A is an  $m \times m$  matrix, B is an  $m \times n$  matrix,  $\Gamma$  is an  $n \times m$  matrix, and  $\Delta$  is an invertible  $n \times n$  matrix,  $\det \begin{pmatrix} A & B \\ \Gamma & \Delta \end{pmatrix} = \det(A - B\Delta^{-1}\Gamma)\det(\Delta).$ 

*Proof.* This standard result on block matrices is proved by factoring into triangular matrices:

$$\begin{pmatrix} A & B \\ \Gamma & \Delta \end{pmatrix} = \begin{pmatrix} A - B\Delta^{-1}\Gamma & B\Delta^{-1} \\ 0_{n,m} & I_{n,n} \end{pmatrix} \begin{pmatrix} I_{m,m} & 0_{m,n} \\ \Gamma & \Delta \end{pmatrix}.$$

Using this,

$$\det(D\Phi) = \det\begin{pmatrix} A & B\\ \Gamma & \Delta \end{pmatrix} = \det(A - B\Delta^{-1}\Gamma)\det(\Delta).$$

We next calculate  $A - B\Delta^{-1}\Gamma$ . We first compute  $B^{-1}\Delta\Gamma$ , labelling the *j*th row of the matrix  $\Delta^{-1}$  by  $\vec{r_i}$  (horizontal vector),

$$B(\Delta^{-1}\Gamma) = B\left(\begin{pmatrix}\vec{r_1}\\\vec{r_2}\\\vdots\\\vec{r_q}\end{pmatrix}\begin{pmatrix}\vec{c} & \vec{0} & \dots\end{pmatrix}\right) = \begin{pmatrix}\vec{b} & \vec{0} & \dots\end{pmatrix}\begin{pmatrix}\vec{c} \cdot \vec{r_1} & 0 & \dots\\\vec{c} \cdot \vec{r_2} & 0 & \dots\\\vdots\\\vec{c} \cdot \vec{r_q} & 0 & \dots\end{pmatrix} = \left((\vec{c} \cdot \vec{r_1})\vec{b} & \vec{0} & \dots\right).$$

We note that  $\vec{c}$  is the first column of  $\Delta$  scaled by  $\frac{\rho \cos \theta}{\sin \theta}$ , and the so the scalar product of it and the first row of the inverse of  $\Delta$  will be  $\frac{\rho \cos \theta}{\sin \theta} \cdot 1$ , as 1 is the upper left entry of  $\Delta^{-1}\Delta$ . Thus, we have that  $B\Delta^{-1}\Gamma$  is a matrix with first column  $\frac{\rho\cos\theta}{\sin\theta}\vec{b}$ , and the rest of columns zero. Thus, as  $\vec{b} = \frac{-\cos\theta}{\rho\sin\theta}\vec{a}$ , we have that  $A - B\Delta^{-1}\Gamma$  is A with the first column multiplied by  $1 + \frac{\cos^2 \theta}{\sin^2 \theta} = \frac{1}{\sin^2 \theta}$ . Using the multilinearity of the determinant, and factoring out the  $\cos \theta$ and  $\sin\theta$  from the columns of A and  $\Delta$ ,

$$\det(D\Phi) = \det(A - B\Delta^{-1}\Gamma) \det(\Delta) = \frac{1}{\sin^2 \theta} \det(A) \det(\Delta)$$
$$= \frac{1}{\sin^2 \theta} (-\rho \sin \theta \cos^p \theta \sin^q \theta) \det(S_{p,\phi}) \det(S_{q-1,\psi})$$
$$= -\rho \cos^p \theta \sin^{q-1} \theta \det(S_{p,\phi}) \det(S_{q-1,\psi}).$$

We note that this determinant is always negative, and so we will negate it in computations in order that our change of variables preserve orientation.

Now that we have established this coordinate transformation, we can prove the theorem in a convenient setting. We first compute

#### Lemma 26.

$$\nabla_{p,q}^{+}G_{p,q,\epsilon} = G_{p,q,\epsilon}\nabla_{p,q}^{+} = \frac{i\epsilon(1-p-q)(p+q+1)(\|x\|^{2}-x^{+}\bar{x})}{(N(x)+i\epsilon\|x\|^{2})^{(p+q+3)/2}}.$$

We use (10), and Stokes' theorem to have, where  $B_r$  is a ball of radius r about the origin, and  $S_r$  its boundary,

#### Corollary 27.

$$d(G_{p,q,\epsilon} \cdot D_{p,q}z \cdot f) = (G_{p,q,\epsilon}\nabla_{p,q}^+)fdV_{p,q} + G_{p,q,\epsilon}(\nabla_{p,q}^+f)dV_{p,q} = (G_{p,q,\epsilon}\nabla_{p,q}^+)fdV_{p,q}$$

$$\int_{\partial U} G_{p,q,\epsilon} \cdot D_{p,q} x \cdot f = \int_{U \setminus B_r} \frac{i\epsilon(p+q+1)(1-p-q)(\|x\|^2 - x^+ \bar{x})}{(N(x) + i\epsilon \|x\|^2)^{(p+q+3)/2}} f dV_{p,q} + \int_{S_r} G_{p,q,\epsilon} \cdot D_{p,q} x \cdot f.$$

We will establish (in order) analogues of Lemma 17, Lemma 18, and Lemma 20 of [L], so that we may analyze the above integrals, and prove the theorem.

**Lemma 28.** If we fix a  $\theta_0 \in (0, \frac{\pi}{4})$  and let p, q be non-negative integers (with  $z^{1/2}$  being defined as after Definition 5) we have two distributions which send a test function  $g(\theta)$  into the limits

$$\lim_{\epsilon \to 0^+} \int_{\frac{\pi}{4} - \theta_0}^{\frac{\pi}{4} + \theta_0} \frac{g(\theta)d\theta}{\left(\cos(2\theta) + i\epsilon\right)^{\frac{p+q+3}{2}}} \qquad and \qquad \lim_{\epsilon \to 0^-} \int_{\frac{\pi}{4} - \theta_0}^{\frac{\pi}{4} + \theta_0} \frac{g(\theta)d\theta}{\left(\cos(2\theta) + i\epsilon\right)^{\frac{p+q+3}{2}}}$$

*Proof.* In the case of  $p + q = 1 \mod 2$ , this is a consequence of Lemma 17 of [L]. If not, we modify the proof to fit the fractional case. We have that  $\frac{p+q+3}{2} = n + \frac{1}{2}$  for some non-negative integer n, and we induct on n. For the base case, n = 0, we integrate by parts,

$$\int_{\frac{\pi}{4}-\theta_{0}}^{\frac{\pi}{4}+\theta_{0}} \frac{g(\theta)d\theta}{(\cos(2\theta)+i\epsilon)^{1/2}} = \int_{\frac{\pi}{4}-\theta_{0}}^{\frac{\pi}{4}+\theta_{0}} \frac{2\sin 2\theta}{(\cos(2\theta)+i\epsilon)^{1/2}} \frac{g(\theta)d\theta}{2\sin 2\theta} \\ = -\frac{2(\cos(2\theta)+i\epsilon)^{1/2}g(\theta)}{(1/2-n)2\sin 2\theta} \Big|_{\frac{\pi}{4}-\theta_{0}}^{\frac{\pi}{4}+\theta_{0}} + \int_{\frac{\pi}{4}-\theta_{0}}^{\frac{\pi}{4}+\theta_{0}} 2(\cos(2\theta)+i\epsilon)^{\frac{1}{2}} \frac{d}{d\theta} \left(\frac{g(\theta)}{2\sin 2\theta}\right) d\theta.$$

As  $(\cos 2\theta + i\epsilon)^{1/2}$  is integrable for all values of  $\epsilon$ , including  $\epsilon = 0$ , the limits as  $\epsilon \to 0^{\pm}$  exist and depend continuously on  $g(\theta)$ . Now we consider the case of n > 0, in which case we can integrate by parts, and use the inductive hypothesis,

$$\int_{\frac{\pi}{4}-\theta_{0}}^{\frac{\pi}{4}+\theta_{0}} \frac{g(\theta)d\theta}{(\cos(2\theta)+i\epsilon)^{n+1/2}} = \int_{\frac{\pi}{4}-\theta_{0}}^{\frac{\pi}{4}+\theta_{0}} \frac{2\sin 2\theta}{(\cos(2\theta)+i\epsilon)^{n+1/2}} \frac{g(\theta)d\theta}{2\sin 2\theta}$$
$$= \frac{(\cos 2\theta+i\epsilon)^{(1/2-n)}}{1/2-n} \frac{g(\theta)}{2\sin 2\theta} \Big|_{\pi/4-\theta_{0}}^{\pi/4+\theta_{0}} + \int_{\frac{\pi}{4}-\theta_{0}}^{\frac{\pi}{4}+\theta_{0}} \frac{1}{(n-\frac{1}{2})(\cos 2\theta+i\epsilon)^{n-\frac{1}{2}}} \frac{\mathrm{d}}{\mathrm{d}\theta} \left(\frac{g(\theta)}{2\sin 2\theta}\right) d\theta.$$

The first limit is well defined, and the second converges by the inductive hypothesis.  $\Box$  Lemma 29.

$$\lim_{\epsilon \to 0} \int_{U \setminus B_r} \frac{i\epsilon(p+q+1)(1-p-q)(\|x\|^2 - x^+ \bar{x})}{(N(x) + i\epsilon \|x\|^2)^{(p+q+3)/2}} f dV_{p,q} = 0.$$

*Proof.* We write the integral in the hybrid spherical coordinates (15), and integrate out the variables  $r, \phi_1, \ldots, \phi_n, \psi_p, \ldots, \psi_{q-1}$ . When we do this, we retain an integral of the form

$$\epsilon \int_0^{\frac{\pi}{2}} \frac{g(\theta)d\theta}{\left(\cos 2\theta + i\epsilon\right)^{\frac{p+q+3}{2}}}.$$
(16)

Due to the transversality of the manifold with respect to the null cone,  $g(\theta)$  is smooth for  $\theta$  lying in the interval  $[\frac{\pi}{4} - \theta_0, \frac{\pi}{4} + \theta_0]$  with  $\theta_0 \in (0, \frac{\pi}{4})$ , so we can apply our previous lemma for  $\int_{\frac{\pi}{4}-\theta_0}^{\frac{\pi}{4}+\theta_0}$ , and the limit on the remainder of the interval is defined, so the limit of the entire integral will exist, meaning that when we take  $\epsilon \to 0$ , the entire expression vanishes.

We define a constant, and evaluate two limits that will come up in our proof.

$$C_{p,q} = \lim_{\epsilon \to 0} \int_0^{\frac{\pi}{2}} \frac{\cos^p \theta' \sin^{q-1} \theta'}{(\cos 2\theta' + i\epsilon)^{(p+q+1)/2}} d\theta'.$$
 (17)

Proposition 30.

$$\lim_{r \to 0^+} (\lim_{\epsilon \to 0} \int_{S_r} \frac{\|x\|^2 f(x)}{(N(x) + i\epsilon r^2)^{\frac{p+q+1}{2}}} \frac{dS_{p,q}}{r}) = \omega_p \omega_{q-1} i^q C_{p,q} f(0).$$

Proof. We can split  $dS_{p,q} = (r^{p+q} \cos^p \theta \sin^{q-1} \theta d\theta d\Omega_{p,\phi} d\Omega_{q-1,\psi}) i^q$ , where  $\Omega_{n,\alpha} = \frac{\det(S_{n,\alpha})}{\rho^n} d\alpha_1 d\alpha_2 \cdots d\alpha_n$  represents the angular integral components. The factor of  $i^q$  comes from the normalization of the Euclidean volume element on this space, (6). We note that the factors of r in the numerator and denominator cancel. We let  $F_{\epsilon}(\theta)$  be the antiderivative of  $\frac{\cos^p \theta \sin^{q-1} \theta}{(\cos 2\theta + i\epsilon)^{(p+q+1)/2}}$ . When we do this, and integrate by parts we get

$$F_{\epsilon}(\theta) = \int_{0}^{\theta} \frac{\cos^{p} \theta' \sin^{q-1} \theta'}{(\cos 2\theta' + i\epsilon)^{(p+q+1)/2}} d\theta',$$

$$\int_{S_r} \frac{f(x)}{(\cos 2\theta + i\epsilon)^{\frac{p+q+1}{2}}} dS_{p,q}$$

$$= i^q \int_{S_p} \int_{S_{q-1}} \int_{\theta=0}^{\theta=\pi/2} \frac{f(x)}{(\cos 2\theta + i\epsilon)^{\frac{p+q+1}{2}}} (\cos^p \theta \sin^{q-1} \theta d\theta d\Omega_{p,\phi} d\Omega_{q-1,\psi})$$

$$= i^q \int_{S_p} \int_{S_{q-1}} fF_{\epsilon}(\theta) \Big|_{\theta=0}^{\theta=\pi/2} d\Omega_{p,\phi} d\Omega_{q-1,\psi} - i^q \int_{S_p} \int_{S_{q-1}} F_{\epsilon}(\theta) \frac{\partial f}{\partial \theta} d\theta d\Omega_{\phi,p} d\Omega_{\psi,q-1}.$$

By the chain rule, we have  $\frac{\partial f}{\partial \theta} = \sum_i \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \theta} = r \cdot g(x)$ , where g(x) is a smooth function. Thus, the second term will be a constant multiple of r (by Lemma 28), and so in the limit of  $r \to 0$  it will vanish. In the first term, as r goes to zero, f will approach f(0) (we implicitly apply the argument presented in the proof of Theorem 16), and so we can carry it out, and integrate  $\Omega_{p,\phi}$  and  $\Omega_{q-1,\psi}$  to be the respective surface areas to get

$$i^{q} \int_{S_{p}} \int_{S_{q-1}} \left. fF_{\epsilon}(\theta) \right|_{\theta=0}^{\theta=\pi/2} \to f(0)\omega_{p}\omega_{q-1} \lim_{\epsilon \to 0} \left. F_{\epsilon}(\theta) \right|_{\theta=0}^{\theta=\pi/2} = i^{q}\omega_{p}\omega_{q-1}C_{p,q}f(0).$$

The following integral appears as a cross term.

#### Proposition 31.

$$\lim_{r \to 0^+} (\lim_{\epsilon \to 0} \int_{S_r} \frac{\sum_{i=1}^q \tilde{x}_{j+p} \tilde{e}_{j+p} (\sum_{j=0}^p x_j e_j) f(x)}{(N(x) + i\epsilon r^2)^{\frac{p+q+1}{2}}} \frac{dS_{p,q}}{r} = 0).$$

Proof. We note that as in the previous case, once we convert to spherical coordinates, the factors of r will cancel. Additionally, when we convert to spherical coordinates, we have that the *i*th term of the outer sum will be of the form  $\cos \psi_i \cdot h_i$ , where  $h_i$  is a function does not depend on  $\psi_i$ , except for i = q, which will have instead  $\sin \psi_{i-1}$ . Additionally, we use Lemma 24, to see that the Jacobian as a function of  $\psi_j$  will be proportional to  $\sin^{q-1-i}\psi_j$  for some integer k > 0, with  $\psi_{q-1}$ . By absorbing the remaining factors of  $dS_{p,q}(\cos 2\theta + i\epsilon)^{-\frac{p+q+1}{2}}$  (including  $i^q$  and possible negative signs) into the  $h_i$ , so we still have  $\frac{\partial h_i}{\partial \psi_i} = 0$ , then for all  $\epsilon$ , we can split the integral into

$$\lim_{r \to 0^+} (\lim_{\epsilon \to 0} \sum_{i=1}^{q-2} \int_{S_r} \cos \psi_i \sin^{q-1-i} \psi_i h_i f(x) d\theta d\phi_1 d\phi_2 \cdots d\phi_p d\psi_1 d\psi_2 \cdots d\psi_{q-1} + \int_{S_r} \cos \psi_{q-1} h_{q-1} f(x) d\theta d\phi_1 d\phi_2 \cdots d\phi_p d\psi_1 d\psi_2 \cdots d\psi_{q-1} + \int_{S_r} \sin \psi_{q-1} h_q f(x) d\theta d\phi_1 d\phi_2 \cdots d\phi_p d\psi_1 d\psi_2 \cdots d\psi_{q-1}).$$

We take the  $\epsilon$  limit of each  $h_i$ , which will exist by Lemma 28, and integrate the *i*th term with respect to  $\psi_i$  for  $1 \le i \le q-2$ . When we integrate with respect to  $\psi_i$ , we have

$$\int_{\psi_i=0}^{\psi_i=\pi} \cos\psi_i \sin^{q-1-i}\psi_i h_i f(x) d\psi_i = \frac{\sin^{q-i}\psi_i}{q-i} h_i f(x) \Big|_{\psi_i=0}^{\pi} - \int_0^{\pi} \frac{\sin^{q-i}\psi_i}{q-i} h_i \frac{\partial f}{\partial \psi_i} d\psi_i.$$

The boundary term is zero, and the second integral yields a factor of r by the chain rule as demonstrated in the proof of Proposition 30, and so as we take the limit as r goes to zero, it too will vanish. For i = q - 1, i = q, we have that the bounds of integration will instead be 0 and  $2\pi$ , and the same argument works in this case as well, showing that the the integral of Proposition 31 as a whole will be on the order of r, and vanish in the limit.

Now that we have established these two equalities, we can prove an analog of Lemma 20 of [L].

**Lemma 32.** Let  $C_{p,q}$  be as in (17), then

$$\lim_{r \to 0^+} (\lim_{\epsilon \to 0} \int_{S_r} (1 - p - q) \frac{x^+ \cdot D_{p,q} x \cdot f(x)}{(N(x) + i\epsilon r^2)^{\frac{p+q+1}{2}}}) = i^q (1 - p - q) \omega_p \omega_{q-1} C_{p,q} f(0)$$

*Proof.* We begin by factoring out the (1 - p - q) from both sides. We use Corollary 15, and have

$$\int_{S_r} \frac{x^+ \cdot D_{p,q} x \cdot f}{\left(N(x) + i\epsilon r^2\right)^{\frac{p+q+1}{2}}} = \int_{S_r} \frac{x^+ \bar{x}f}{\left(N(x) + i\epsilon r^2\right)^{\frac{p+q+1}{2}}} \frac{dS_{p,q}}{r}.$$
(18)

We calculate, if  $x = x_0 e_0 + \sum_{j=1}^p x_j e_j + \sum_{j=p+1}^{p+q} \tilde{x}_j \tilde{e}_j$ ,

$$x^{+}\bar{x} = \|x\|^{2} - 2x_{0}\left(\sum_{j=p+1}^{p+q} \tilde{x}_{j}\tilde{e}_{j}\right) + 2\left(\sum_{j=1}^{p} x_{j}e_{j}\right)\left(\sum_{j=p+1}^{p+q} \tilde{x}_{j}\tilde{e}_{j}\right) = \|x\|^{2} - 2\sum_{j=p+1}^{q} \tilde{x}_{j}\tilde{e}_{j}\left(\sum_{j=0}^{p} x_{p}e_{i}\right).$$

We split the integral of (18) into two parts to be analyzed separately,

$$= \int_{S_r} \frac{\|x\|^2 f}{(N(x) + i\epsilon r^2)^{\frac{p+q+1}{2}}} \frac{dS_{p,q}}{r} - 2 \int_{S_r} \frac{\sum_{i=1}^q \tilde{x}_{j+p} \tilde{e}_{j+p} (\sum_{j=0}^p x_j e_j) f(x)}{(N(x) + i\epsilon r^2)^{\frac{p+q+1}{2}}} \frac{dS_{p,q}}{r}.$$

Applying Propositions 30 and 31 yields the lemma.

We can prove Theorem 23 up to a constant. If 0 is not contained within our manifold U, we have that the integral about the boundary will be the integral of Lemma 29 on the interior, which goes to zero. If not, we can apply Corollary 27, and we have that the first term goes to zero by Lemma 29, and the second integral goes to  $i^q \omega_p \omega_{q-1} C_{p,q}$  by the Lemma 32.

#### Lemma 33.

$$C_{p,q} = (-i)^q \frac{\omega_{p+q}}{\omega_p \omega_{q-1}}.$$

*Proof.* We prove this by applying both integral formulas to the constant function  $f(x) = \frac{1}{1-p-q}$  on the sphere  $S_{p,q} = \{x \in \mathbb{R}^{p+q+1} : ||x|| = 1\}$ . When we apply the first integral formula, we obtain for all  $\epsilon$  sufficiently close to zero:

$$\int_{(h_{p,q,\epsilon})*(S_{p,q})} \frac{z^+ D z}{N(z)^{\frac{p+q+1}{2}}} = \omega_{p+q}.$$

We note that  $h_{p,q,\epsilon}$  is a linear transformation, and we can write this as an integral over the sphere  $S_{p,q}$  once we calculate the pullback of  $D_z$ . We expand the form in the standard basis, and note that for  $w_j \in S_{p,q}$ , with  $h_{p,q,\epsilon}(w_j) = z_j$ , if  $j \leq p$ ,  $dz_j = (1 + i\epsilon)dw_j$ , and if j > p, we have  $dz_j = (1 - i\epsilon)dw_j$ . We expand Dz using (8), and have where  $\tilde{D}w$  is a constant p + q form that depends polynomially on  $\epsilon$ :

$$Dz = \sum_{j=0}^{p} (-1)^{j} e_{j} (1+i\epsilon)^{p} (1-i\epsilon)^{q} d\hat{w}_{j} - \sum_{j=p+1}^{p+q} (-1)^{j} e_{j} (1+i\epsilon)^{p+1} (1-i\epsilon)^{q-1} d\hat{w}_{j}$$
$$= Dw + \epsilon \tilde{D}w.$$

We let  $\tilde{z}$  be defined as

$$z = \sum_{j=0}^{p+q} z_j e_j \to \tilde{z} = \sum_{j=0}^{p} z_j e_j - \sum_{j=p+1}^{p+q} z_j e_j$$

Using this operator, and Corollary 18 we obtain expressions for the final two terms in our integral,

$$z^{+} = (w + i\epsilon \tilde{w})^{+} = w^{+} + i\epsilon \tilde{w}^{+},$$
  
$$N(z) = (1 - \epsilon^{2})N(w) + i\epsilon ||w||^{2} = (1 - \epsilon^{2})N(w) + i\epsilon$$

$$\int_{(h_{p,q,\epsilon})*(S_{p,q})} \frac{z^{+}Dz}{N(z)^{\frac{p+q+1}{2}}} = \int_{S_{p,q}} \frac{(w^{+} + i\epsilon\tilde{w}^{+})(Dw + \epsilon\tilde{D}w)}{((1-\epsilon^{2})N(w) + i\epsilon)^{\frac{p+q+1}{2}}} \\
= \int_{S_{p,q}} \frac{w^{+}Dw}{((1-\epsilon^{2})N(w) + i\epsilon)^{\frac{p+q+1}{2}}} + \int_{S_{p,q}} \frac{\epsilon\tilde{w}^{+}\tilde{D}w + i\epsilon\tilde{w}^{+}(Dw + \epsilon\tilde{D}w)}{((1-\epsilon^{2})N(w) + i\epsilon)^{\frac{p+q+1}{2}}}.$$
(19)

With the first integral of (19), we do the change of variables,  $\epsilon' = \frac{\epsilon}{1-\epsilon^2}$ ,  $\epsilon = \frac{-1+\sqrt{1+4(\epsilon')^2}}{2\epsilon'}$ , with  $\epsilon(0) = 0$ . We note that  $\epsilon'$  is a smooth function of  $\epsilon$  for  $\epsilon$  sufficiently small.

$$\int_{S_{p,q}} \frac{w^+ Dw}{((1-\epsilon^2)N(w)+i\epsilon)^{\frac{p+q+1}{2}}} = \left(\frac{\epsilon'}{\epsilon}\right)^{\frac{p+q+1}{2}} \int_{S_{p,q}} \frac{w^+ Dw}{(N(w)+i\epsilon')^{\frac{p+q+1}{2}}}$$

We take the limit as  $\epsilon$  goes to zero, and have that the outer term will approach 1, and the integral will approach  $i^q \omega_p \omega_{q-1}$  by the second formulation of Cauchy Fueter formula, as  $\epsilon \to 0 \iff \epsilon' \to 0$ .

For the second integral of (19), we note that we can factor out  $\epsilon$ , and then change the inner coordinates from  $\epsilon$  to  $\epsilon'$ , and have where  $D_{\epsilon'}$  is a constant differential form that depends smoothly on  $\epsilon'$ :

$$\int_{S_{p,q}} \frac{\epsilon \tilde{w}^+ \tilde{D}w + i\epsilon \tilde{w}^+ (Dw + \epsilon \tilde{D}w)}{((1-\epsilon^2)N(w) + i\epsilon)^{\frac{p+q+1}{2}}} = \epsilon \left(\frac{\epsilon'}{\epsilon}\right)^{\frac{p+q+1}{2}} \int_{S_{p,q}} \frac{D_{\epsilon'}}{(N(w) + i\epsilon')^{\frac{p+q+1}{2}}}$$

We can convert to the spherical coordinates of (15), and apply Lemma 28 in the same manner as is done in the proof of the second integral formula. When we do this, we have that the integral approaches a constant C as  $\epsilon'$  goes to zero, so as  $\epsilon$  goes to zero, the second term approaches zero. We prove Lemma 33 a second way.

*Proof.* We apply both integral formulas to the constant function  $f(x) = \frac{1}{1-p-q}$  to show this equivalence. We let  $S_{p,q,1}$  be the unit sphere in  $\mathbb{R}^{p+q+1}$ . We have by the first integral formula that there exists an  $\epsilon_1 > 0$  such that for all  $0 < \epsilon \leq \epsilon_1$ ,

$$\int_{(h_{p,q,\epsilon})*(S_{p,q,1})} \frac{z^+ D_{p+q} z}{(N(z))^{\frac{p+q+1}{2}}} = \omega_{p+q}.$$

We note that this is a strict equality, and not a limiting argument, so we can fix this  $\epsilon_1$  throughout the proof. We will consider the following integral:

$$\int_{S_{p,q,1}} \frac{z^+ D_{p,q} z}{(N(z) + i\epsilon_2)^{\frac{p+q+1}{2}}}$$

We have by the first integral formula

$$\lim_{\epsilon_2 \to 0} \int_{S_{p,q,1}} \frac{z^+ D_{p,q} z}{\left(N(z) + i\epsilon_2\right)^{\frac{p+q+1}{2}}} = \lim_{\epsilon_2 \to 0} \int_{S_{p,q,1}} \frac{x^+ D_{p+q} x}{\left(N(x) + i\epsilon_2 \|x\|\right)^{\frac{p+q+1}{2}}} = i^q C_{p,q} \omega_p \omega_{q-1}.$$

We let  $\Gamma_{\epsilon_2}(z) = \left(\frac{1}{N(z)^{\frac{p+q+1}{2}}} - \frac{1}{(N(z)+i\epsilon_2)^{\frac{p+q+1}{2}}}\right)$ , and have by Stokes' Theorem, if U is the manifold bounded by  $(h_{p,q,\epsilon_1})_*(S_{p,q,1})$  and  $S_{p,q,1}$ :

$$\begin{split} i^{q}C_{p,q}\omega_{p}\omega_{q-1} &= \int_{S_{p,q,1}} \frac{z^{+}D_{p,q}z}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}} \\ &= \int_{(h_{p,q,\epsilon_{1}})_{*}(S_{p,q,1})} \frac{z^{+}D_{p,q}z}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}} - \int_{U} d(\frac{z^{+}D_{p,q}z}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}}) \\ &= \int_{(h_{p,q,\epsilon_{1}})_{*}(S_{p,q,1})} \frac{z^{+}D_{p,q}z}{(N(z))^{\frac{p+q+1}{2}}} - \int_{(h_{p,q,\epsilon_{1}})_{*}(S_{p,q,1})} z^{+}\Gamma_{\epsilon_{2}}(z)D_{p+q}z - \int_{U} d(\frac{z^{+}D_{p,q}z}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}}) \\ &= \omega_{p+q} - \int_{(h_{p,q,\epsilon_{1}})_{*}(S_{p,q,1})} z^{+}\Gamma_{\epsilon_{2}}(z)D_{p+q}z - \int_{U} d(\frac{z^{+}D_{p,q}z}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}}). \end{split}$$

We apply the following lemmas, and the constant is computed.

#### Lemma 34.

$$\lim_{\epsilon_2 \to 0} \int_U d(\frac{z^+ D_{p,q} z}{(N(z) + i\epsilon_2)^{\frac{p+q+1}{2}}}) = 0.$$
(20)

*Proof.* By (9) we have

$$\int_{U} d\left(\frac{z^{+}D_{p,q}z}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}}\right) = \int_{U} \left(\frac{z^{+}}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}}\nabla^{+}\right) dV_{\mathbb{C}}.$$

We do a similar computation to Lemma 26:

$$\nabla^{+}\left(\frac{z^{+}}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}}\right) = \left(\frac{z^{+}}{(N(z)+i\epsilon_{2})^{\frac{p+q+1}{2}}}\right)\nabla^{+} = \frac{i\epsilon_{2}(p+q+1)}{(N(z)+i\epsilon_{2})^{\frac{p+q+3}{2}}}.$$

We can parametrize the sphere  $S_{p,q,1}$  by (15), with  $\rho = 1$ , and extend this to a parametrization of U by considering the parametrization under  $h_{p,q,\epsilon}$  as  $\epsilon$  varies 0 to  $\epsilon_1$ . We have that  $N(z) = \cos 2\theta + i\epsilon$  by Corollary 18. We can integrate out the variables  $\phi_1, \phi_2, \ldots, \phi_p, \psi_1, \ldots, \psi_{q-1}$ , get the integral to be of the form

$$= \int_{\theta=0}^{\theta=\frac{\pi}{2}} \int_{\epsilon=0}^{\epsilon=\epsilon_1} \frac{i\epsilon_2 g(\theta,\epsilon)}{\left(\cos 2\theta + i\epsilon + i\epsilon_2\right)^{\frac{p+q+3}{2}}} d\epsilon d\theta.$$

In this case,  $g(\theta, \epsilon)$  a smooth function that arises from the Jacobian of this transformation (it will be a polynomial in  $\epsilon$ ,  $\cos \theta$ ,  $\sin \theta$  that does not depend on  $\epsilon_2$ ). We prove a very similar result to Lemma 29 that will prove (20):

**Proposition 35.** Let  $g(\theta, \epsilon)$  be a smooth function, and  $n \ge 1$  an integer.

$$\lim_{\epsilon_2 \to 0} \int_{\theta=0}^{\theta=\frac{\pi}{2}} \int_{\epsilon=0}^{\epsilon=\epsilon_1} \frac{i\epsilon_2 g(\theta,\epsilon)}{(\cos 2\theta + i\epsilon + i\epsilon_2)^{\frac{n}{2}}} d\epsilon d\theta = 0.$$

We show this inductively, with two base cases corresponding to the parity of n, n = 1 and n = 2. In each of these cases we integrate by parts with respect to  $\epsilon$ :

$$\int_{0}^{\frac{\pi}{2}} \int_{0}^{\epsilon_{1}} \frac{i\epsilon_{2}g(\theta,\epsilon)}{(\cos 2\theta + i\epsilon + i\epsilon_{2})^{\frac{1}{2}}} = 2i\epsilon_{2} \int_{0}^{\frac{\pi}{2}} g(\theta,\epsilon)(\cos 2\theta + i\epsilon_{2} + i\epsilon_{1})^{\frac{1}{2}} \Big|_{\epsilon=0}^{\epsilon=\epsilon_{1}} d\theta - 2i\epsilon_{2} \int_{0}^{\frac{\pi}{2}} \int_{0}^{\epsilon_{1}} (\cos 2\theta + i\epsilon + i\epsilon_{2})^{\frac{1}{2}} \frac{\partial g(\theta,\epsilon)}{\partial \epsilon} d\epsilon d\theta,$$

$$\int_{0}^{\frac{\pi}{2}} \int_{0}^{\epsilon_{1}} \frac{i\epsilon_{2}g(\theta,\epsilon)}{(\cos 2\theta + i\epsilon + i\epsilon_{2})^{1}} = i\epsilon_{2} \int_{0}^{\frac{\pi}{2}} g(\theta,\epsilon) \log(\cos 2\theta + i\epsilon_{2} + i\epsilon_{1}) \Big|_{\epsilon=0}^{\epsilon=\epsilon_{1}} d\theta - i\epsilon_{2} \int_{0}^{\frac{\pi}{2}} \int_{0}^{\epsilon_{1}} \log(\cos 2\theta + i\epsilon + i\epsilon_{2}) \frac{\partial g(\theta,\epsilon)}{\partial \epsilon} d\epsilon d\theta.$$

Here, as with the proof of Lemma 18 of [L] and Lemma 29 above, the complex logarithm and  $z^{1/2}$  are defined everywhere except the negative real axis. We thus have that as  $\epsilon_2$  goes to zero, the limits of the integrals exist, and so when we take into account the factor of  $\epsilon_2$ multiplying the integrals, the terms go to zero. We consider the case of general n,

$$\int_{0}^{\frac{\pi}{2}} \int_{0}^{\epsilon_{1}} \frac{i\epsilon_{2}g(\theta,\epsilon)}{(\cos 2\theta + i\epsilon + i\epsilon_{2})^{\frac{n}{2}}} = \frac{2i\epsilon_{2}}{n-1} \int_{0}^{\frac{\pi}{2}} \frac{g(\theta,\epsilon)}{(\cos 2\theta + i\epsilon + i\epsilon_{2})^{\frac{n-1}{2}}} \bigg|_{\epsilon=0}^{\epsilon=\epsilon_{1}} d\theta$$
$$- \frac{2i\epsilon_{2}}{n-1} \int_{0}^{\frac{\pi}{2}} \int_{0}^{\epsilon_{1}} \frac{\frac{\partial g(\theta,\epsilon)}{\partial\epsilon}}{(\cos 2\theta + i\epsilon + i\epsilon_{2})^{\frac{n-1}{2}}} d\epsilon d\theta.$$

The second integral goes to zero by the inductive hypothesis, so we consider the first,

$$\frac{2i\epsilon_2}{n-1}\int_0^{\frac{\pi}{2}} \frac{g(\theta,\epsilon_1)}{(\cos 2\theta + i\epsilon_1 + i\epsilon_2)^{\frac{n-1}{2}}} d\theta - \frac{2i\epsilon_2}{n-1}\int_0^{\frac{\pi}{2}} \frac{g(\theta,0)}{(\cos 2\theta + i\epsilon_2)^{\frac{n-1}{2}}} d\theta.$$

The magnitude of the integrand in the first of these integral is bounded by  $G\epsilon_1^{-\frac{n-1}{2}}$ , where G is the sup of  $g(\theta, \epsilon_1)$  on the interval  $[0, \frac{\pi}{2}]$ , and so the integral is bounded, and when we multiply it by  $i\epsilon_2$ , and take the limit, it will go to zero. The second integral is similar to (16) and vanishes by the same reasoning.

Lemma 36. With 
$$\Gamma_{\epsilon_2}(z) = \left(\frac{1}{N(z)^{\frac{p+q+1}{2}}} - \frac{1}{(N(z)+i\epsilon_2)^{\frac{p+q+1}{2}}}\right)$$
, we have  
$$\lim_{\epsilon_2 \to 0} \int_{(h_{p,q,\epsilon_1})_*(S_{p,q,1})} z^+ \Gamma_{\epsilon_2}(z) D_{p+q} z = 0.$$

*Proof.* We use the Cauchy Schwartz inequality, and let  $M_1 = \int_{(h_{p,q,\epsilon_1})_*(S_{p,q,1})} |z^+ Dz|$ , and have

$$\left| \int_{(h_{p,q,\epsilon_1})_*(S_{p,q,1})} z^+ D z \Gamma_{\epsilon_2}(z) \right| \le M_1 \sup_{z \in (h_{p,q,\epsilon_1})_*(S_{p,q,1})} |\Gamma_{\epsilon_2}(z)|$$

We note that the sup on  $(h_{p,q,\epsilon_1})_*(S_{p,q,1})$  of  $\left|\left(\frac{1}{N(z)^{\frac{p+q+1}{2}}} - \frac{1}{(N(z)+i\epsilon_2)^{\frac{p+q+1}{2}}}\right)\right|$  goes to zero. We note that for all z, both terms are well defined as  $\epsilon_2$  goes to zero, as N(z) has positive imaginary part of at least  $\epsilon_1$  by Corollary 18. Utilizing this, one can show that this expression goes to zero uniformly via binomial expansion.

### 8 Conformal Mappings acting on Monogenic Functions

In this section, we consider mappings from  $\mathbb{R}^{p+q+1}$  to itself that are conformal with respect to the quadratic form of signature (p + 1, q), and how these mappings relate to left and right-monogenic functions. We write  $\mathbb{R}^{p+q+1}$  with this quadratic form as  $\mathbb{R}^{p+1,q}$ , and write the quadratic form as Q(x), to be more consistent with the literature. We observe that the set of monogenic functions is invariant under translations and scalings of  $\mathbb{R}^{p+1,q}$ , as seen from the definitions. We then provide a group action of O(p,q) on the functions that map into A(p,q). We then compute the pullback of  $D_{p,q}x$  under the inversion  $x \to Q(x)^{-1}x$ , and compare it to the quaternionic case presented in [Su].

We first consider the group O(p,q) acting on the (p,q)-monogenic functions. For these arguments, we consider (p,q)-left-monogenic functions, with the (p,q)-right-monogenic functions being acted on in the same way. We consider O(p,q) acting on  $\mathbb{R}^{p+1,q}$  by having each matrix M act linearly on the last p+q coordinates. We consider  $\mathcal{F} = \{f : \mathbb{R}^{p+q+1} \to A_{p,q} :$  $\nabla^+ f(x) = 0\}$ . As  $A_{p,q}$  is a universal Clifford Algebra, we have that there exists a unique extension of  $M : \mathbb{R}^{p+q} \to \mathbb{R}^{p+q}$  to an algebra homomorphism  $\tilde{M} : A_{p,q} \to A_{p,q}$  which satisfies  $\tilde{M}(e_j) = M(e_j)$  for  $1 \leq j \leq p+q$  and  $\tilde{M}(e_0) = e_0$ . We consider the mapping on  $\mathcal{F}$  given by

$$M \in O(p,q) : f \to \tilde{M}(f(M^{-1}(x))).$$

$$\tag{21}$$

**Lemma 37.** This is a group action of O(p,q) on  $\mathcal{F}$ . Similarly, if we define  $\mathcal{G} = \{g : \mathbb{R}^{p+q+1} \to A_{p,q} : g(x)\nabla^+ = 0\}$ , we have the following group action of O(p,q) on  $\mathcal{G}$ :

$$M \in O(p,q) : g \to \tilde{M}(g(M^{-1}(x)))$$

*Proof.* We prove the result only in the left-monogenic case. We first show that this is indeed a group action.

$$M_1 \circ (M_2 \circ f) = M_1 \circ \tilde{M}_2 f(M_2^{-1}(x)) = \tilde{M}_1 \tilde{M}_2 f(M_2^{-1} M_1^{-1}(x)) = \widetilde{M}_1 M_2 f((M_1 M_2)^{-1} x).$$

To show that the image of a left-monogenic function is left-monogenic, we follow an analogous method to the proof of Theorem 20 of [Su], and use Corollary 13. We let  $\mu$  be the action that maps x to the image of x under  $M^{-1} \in O(p,q)$ . We calculate the pullback of  $D_{p,q}x$  to be

$$\mu^* D_{p,q} x(h_1, h_2, \dots, h_{p+q}) = D_{p,q} x(M^{-1}(h_1), M^{-1}(h_2), \dots, M^{-1}(h_{p+q})).$$

By Definition 10, with  $det(M) = det(M^{-1}) = \pm 1, \forall h_0, h_1, \dots, h_{p+q} \in \mathbb{R}^{p+q+1}$ :

 $\langle M^{-1}(h_0), \mu^* D_{p,q} x(h_1, \dots, h_{p+q}) \rangle = dV(M^{-1}h_0, \dots, M^{-1}h_{p+q}) = \det(M)dV(h_0, h_1, \dots, h_{p+q}).$ 

Moreover, as  $M^{-1}$  preserves the quadratic form of signature (p + 1, q) it also preserves the associated bilinear form,

$$\langle M^{-1}(h_0), M^{-1}(D_{p,q}x(h_1,\ldots,h_{p+q})) \rangle = \langle h_0, (D_{p,q}x(h_1,\ldots,h_{p+q})) \rangle = dV(h_0,\ldots,h_{p+q}).$$

By noting the similarity of the previous two equations, we have, as  $M(D_{p,q}x) = \tilde{M}(D_{p,q}x)$ 

$$\mu^* D_{p,q} x(h_1, \dots, h_{p+q}) = \det(M) M^{-1}(D_{p,q} x(h_1, \dots, h_{p+q})) \Longrightarrow D_{p,q} x = \det(M) \tilde{M}(\mu^* D_{p,q} x).$$
(22)

We have by Corollary 13:

$$d(D_{p,q}xf) = D_{p,q}x \wedge df = 0$$

We have that by the chain rule, as  $\tilde{M}$  is linear:

$$d(\tilde{M}(f(M^{-1}(x)))) = \tilde{M}df(M^{-1}(x)) = \tilde{M}(\mu^* df.)$$

To test if  $f' = \tilde{M}^{-1}(f(M(x)))$  is also left-monogenic we compute:

$$d(D_{p,q}x \cdot \tilde{M}(f(M^{-1}(x)))) = D_{p,q}x \wedge d(\tilde{M}(f(M^{-1}(x))))$$
  
= det(M) $\tilde{M}(\mu^* D_{p,q}x) \wedge d(\tilde{M}(f(M^{-1}(x)))) = det(M)\tilde{M}(\mu^* D_{p,q}x) \wedge \tilde{M}(\mu^* df)$   
= det(M) $\tilde{M}(\mu^*(D_{p,q}x \wedge df)) = det(M)\tilde{M}(\mu^*(0)) = 0.$ 

We consider a similar method for a modified inversion:  $x \to \text{Inv}(x) = \frac{x^+}{N(x)}$ . We note that this is a conformal map with respect to the quadratic form, Q, of signature (p+1,q) with a conformal factor of  $\pm \frac{1}{Q(x)}$ . We write the Jacobian derivative matrix as [D Inv].

**Lemma 38.** The pullback of the form  $D_{p,q}x$  under the mapping  $\mu : x \to \text{Inv}(x) = \frac{x^+}{Q(x)}$  is given by:

$$\mu^* D_{p,q} x = \frac{(-1)^{p+q+1}}{Q(x)^{p+q-1}} [D \operatorname{Inv}(x)] D_{p,q} x$$
(23)

where  $[D \operatorname{Inv}(x)]D_{p,q}x$  denotes applying the linear map corresponding to the derivative of  $\operatorname{Inv}(x)$  to  $D_{p,q}x$ .

*Proof.* As Inv(x) is conformal with conformal factor  $\Omega = \frac{1}{Q(x)}$ , by expressing Q(x) as the diagonal matrix  $I^{p+1,q}$ , and using the definition of a conformal mapping ([Sc]),

$$[D \operatorname{Inv}(x)]^T I^{p+1,q} [D \operatorname{Inv}(x)] = \Omega^2 I^{p+1,q} = \frac{1}{Q(x)^2} I^{p+1,q}.$$

We can take the determinant of both sides.  $det\left(\frac{1}{Q(x)^2}I^{p+1,q}\right) = Q(x)^{-2(p+q+1)}det(I^{p+1,q}),$ and so  $det[D \operatorname{Inv}] = \pm Q(x)^{-(p+q+1)}$ . This determinant will be a rational function, and so to determine the sign we need to evaluate it at a single point. By substituting  $x = e_0$ ,

$$[D \operatorname{Inv}] = \begin{pmatrix} \frac{Q(x) - 2x_0^2}{Q(x)^2} & \frac{-2x_1x_0}{Q(x)^2} & \cdots & \frac{2x_0x_{p+1}}{Q(x)^2} & \cdots & \frac{2x_0x_{p+q}}{Q(x)^2} \\ \frac{2x_0x_1}{Q(x)^2} & \frac{-Q(x) + 2x_1^2}{Q(x)^2} & \cdots & \frac{-2x_1x_{p+1}}{Q(x)^2} & \cdots & \frac{-2x_1x_{p+q}}{Q(x)^2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{2x_0x_{p+1}}{Q(x)^2} & \frac{2x_1x_{p+1}}{Q(x)^2} & \cdots & \frac{-Q(x)^2 - 2x_{p+1}^2}{Q(x)^2} & \cdots & \frac{-2x_{p+1}x_{p+q}}{Q(x)^2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{2x_0x_{p+q}}{Q(x)^2} & \frac{2x_1x_{p+q}}{Q(x)^2} & \cdots & \frac{-2x_{p+1}x_{p+q}}{Q(x)^2} & \cdots & \frac{-Q(x) - 2x_{p+q}^2}{Q(x)^2} \end{pmatrix}, \\ \left[ D \operatorname{Inv} \right] \Big|_{(1,0,\dots)} = -I,$$

we get this signature to be  $(-1)^{p+q+1}$ . This informs us how the volume element  $dV_{p,q}$  will transform under the inversion mapping, which we use to determine how  $D_{p,q}x$  transforms. We can calculate the pullback of  $D_{p,q}x$  under the inversion (denoted by  $\mu : x \to \text{Inv}(x)$ ) as was done in the orthogonal case.

$$\langle [D \operatorname{Inv}(x)]h_0, D_{p,q}x([D \operatorname{Inv}(x)]h_1), \dots, [D \operatorname{Inv}(x)]h_{p+q}) \rangle = dV([D \operatorname{Inv}(x)]h_0, [D \operatorname{Inv}(x)]h_1, \dots) = \frac{(-1)^{p+q+1}dV_{p,q}(h_0, h_1, \dots)}{Q(x)^{p+q+1}},$$

 $\langle [D \operatorname{Inv}(x)]h_0, [D \operatorname{Inv}(x)]D_{p,q}x(h_1, \dots, h_{p+q}) \rangle = \frac{1}{Q(x)^2} \langle h_0, D_{p,q}x(h_1, \dots, h_{p+q}) \rangle = \frac{dV_{p,q}(h_0, h_1, \dots)}{Q(x)^2}$ 

By comparing these two lines, we see

$$D_{p,q}x([D\operatorname{Inv}(x)]h_1,\dots,[D\operatorname{Inv}(x)]h_{p+q}) = \frac{(-1)^{p+q+1}}{Q(x)^{p+q-1}}[D\operatorname{Inv}(x)]D_{p,q}x(h_1,\dots,h_{p+q})$$
$$\mu^*D_{p,q}x = \frac{(-1)^{p+q+1}}{Q(x)^{p+q-1}}[D\operatorname{Inv}(x)]D_{p,q}x.$$

This is analogous to the pullback of O(p,q), (22), insofar as the differential form is acted upon by a linear map, although in this case the map is not constant. This feature prevents a statement similar to Lemma 37, as when we take the exterior derivative of  $\mu^*Dx$  we get an additional term due to Liebniz's rule.

In the quaternionic case, presented in [Su], the pullback under quaternionic inversion is found to be:

$$\mu^* Dq = -\frac{q^{-1}}{|q|^4} \cdot Dq \cdot q^{-1}.$$
(24)

While this appears distinct from (23), we show that they coincide. In quaternionic analysis, multiplication by a quaternion is a linear map,  $\mathbb{H} \to \mathbb{H}$ , which can each be identified with  $\mathbb{R}^4$ . Expressing these maps as real matrices, we have that multiplication by  $q^{-1}$  on the left and right correspond to the matrices:

$$M_{left} = \frac{1}{|q|^2} \begin{pmatrix} q_0 & q_1 & q_2 & q_3 \\ -q_1 & q_0 & q_3 & -q_2 \\ -q_2 & -q_3 & q_0 & q_1 \\ -q_3 & q_2 & -q_1 & q_0 \end{pmatrix}, \qquad M_{right} = \frac{1}{|q|^2} \begin{pmatrix} q_0 & q_1 & q_2 & q_3 \\ -q_1 & q_0 & -q_3 & q_2 \\ -q_2 & q_3 & q_0 & -q_1 \\ -q_3 & -q_2 & q_1 & q_0 \end{pmatrix}.$$

We can calculate the matrix,  $[D \operatorname{Inv}(q)]$  for p = 3, q = 0, and note that it factors into the above matrices:

$$[D \operatorname{Inv}(q)] = \frac{1}{|q|^4} \begin{pmatrix} |q|^2 - 2q_0^2 & -2q_0q_1 & -2q_0q_2 & -2q_0q_3 \\ 2q_0q_1 & |q|^2 - 2q_1^2 & 2q_1q_2 & 2q_1q_3 \\ 2q_0q_2 & 2q_1q_2 & |q|^2 - 2q_2^2 & 2q_2q_3 \\ 2q_0q_3 & 2q_1q_3 & 2q_2q_3 & |q|^2 - 2q_3^2 \end{pmatrix}$$
$$= -M_{left}M_{right} = -M_{right}M_{left}. \quad (25)$$

As  $|q|^2$  serves the role of Q(x), we can write (23) as:

$$\mu^* D_{3,0} x = \frac{1}{Q(x)^2} [D \operatorname{Inv}(x)] D_{3,0} x,$$
$$\mu^* Dq = \frac{1}{|q|^4} [D \operatorname{Inv}(q)] Dq = \frac{1}{|q|^4} (-M_{left} M_{right}) Dq = -\frac{1}{|q|^4} q^{-1} \cdot Dq \cdot q^{-1}.$$

Thus, our pullback computation is consistent with the quaternionic case. A more elegant formula similar to (24) seems to be unlikely in the general case as  $\mathbb{R}^{p+q+1}$  is not closed under multiplication, and so multiplication by an element similar to  $q^{-1}$  on both sides of Dqwould carry the image out of  $\mathbb{R}^{p+q+1}$  and into the full Clifford Algebra  $A_{p,q}$ . Moreover, the factorization exhibited in (25) appears not extend to general Universal Clifford Algebras. Despite this, if such an action were to exist, by analogy with similar cases, it would likely be of the following form:

$$f(x) \in \mathcal{F} \to \frac{x^+}{N(x)^{(p+q+1)/2}} \cdot f(\operatorname{Inv}(x)),$$
$$g(x) \in \mathcal{G} \to g(\operatorname{Inv}(x)) \cdot \frac{x^+}{N(x)^{(p+q+1)/2}}.$$

### References

- [G] D. J. H. Garling, Clifford Algebras: An Introduction, Cambridge Univ. Press, Cambridge, UK, 2011.
- [GM] J. Gilbert, M. Murray, Clifford Algebras and Dirac Operators in Harmonic Analysis, Cambridge Univ. Press, Cambridge, UK, 1991.

- [KØ] T. Kobayashi, B. Ørsted, Analysis on the minimal representation of O(p, q) I, II, III, Adv. Math. 180 (2003), no. 2, 486-512, 513-550, 551-595.
- [L] M. Libine, An invitation to split quaternionic analysis in I. Sabadini, F. Sommen (Eds.), Hypercomplex Analysis and its Applications, Birkhäuser, 2011, pp. 161-180; also arXiv:1009.2540.
- [R] J. Ryan, Complexified Clifford Analysis, Complex Variables Theory Appl. 1 (1982/83), no. 1, 119-149.
- [Sc] M. Schottenloher, A Mathematical Introduction to Conformal Field Theory. SpringerVerlag, Lecture Notes in Physics 759, 2nd ed, 2008.
- [Su] A. Sudbery, Quaternionic analysis, Math. Proc. Camb. Phil. Soc. 85 (1979), 199-225.
# On the Large Time Diffusion of Isotropic and Anisotropic Particles

Mikhail Sweeney

July 26th, 2019

### 1 Abstract

We first look at Ebbens et. al's 2010 paper and, using abstract diffusion theory methods, replicate their results concerning the mean square displacement and effective diffusivity of a spherical particle with both rotational and translational self propulsion.

We then look at Fan, Pak, and Sandoval's 2017 paper, in which they used an approximation to calculate the effective diffusivity of an elliptic particle in a rotating magnetic field with translational self propulsion. We then apply abstract diffusion theory methods to rigorously verify the claims made in the paper and show that the effective diffusivity of the anisotropic particle is very similar to the effective diffusivity of the above self-propelled runner-and-tumbler.

### 2 Definitions and Lemmas

Before we begin the derivation of the results claimed in the abstract, we first introduce a handful of definitions from Markov Theory and prove two brief lemmas that will be useful in the calculations for both papers.

#### 2.1 Definitions

We first recall the elementary definition of the

**Definition 2.1** (Markov Property). A stochastic process  $X_t : \Omega \to S$  satisfies the Markov Property if  $\forall f : S \to \mathbb{R}$  bounded and measurable,

$$\mathbb{E}_x[f(X_{t+s})|\mathcal{F}_t] = \mathbb{E}_{X_t}[f(X_s)] = P_s f(X_t)$$

where  $\mathcal{F}_t = \sigma\{X_r : r \leq t\}.$ 

And thus

**Definition 2.2** (Markov Process). A stochastic process  $X_t$  is a Markov Process if and only if it satisfies the Markov Property.

We also have the basic fact that every Markov process  $X_t$  admits a semigroup of operators  $\{P_t\}$ , where  $P_t f(x) = \mathbb{E}_x[f(X_t)]$  and  $P_s P_t = P_{s+t}$ .

A Markov Process may also under certain conditions admit a stationary distribution, which is defined as follows.

**Definition 2.3** (Stationary Distribution).  $\rho$  is a stationary distribution for the Markov Process  $X_t$  with semigroup  $\{P_t\}$  if  $\rho(P_t f) = \rho(f)$  for all f bounded measurable and  $t \ge 0$ , with  $\rho(g) := \int g(x) \rho(dx)$ , where g is bounded measurable.

Markov processes also inherently have a transition probability funciton P that governs the probabilities of moving between states.

**Definition 2.4** (Transition Probability). The transition probability of a Markov Process  $X_t$  is a function P(s, x, t, B) such that  $P(s, x, t, B) = P(X_t \in B | X_s = x)$ , satisfying the following properties:

- 1.  $P(s, x, t, \cdot)$  is a probability measure on  $\mathcal{B}(\mathbb{R})$  for  $s \leq t$  fixed and  $x \in \mathbb{R}$ .
- 2.  $P(s, \cdot, t, B)$  is  $\mathcal{B}(\mathbb{R})$ -measurable for  $s \leq t$  fixed and  $B \in \mathcal{B}(\mathbb{R})$ .

And further, if it exists,

**Definition 2.5** (Transition Density). The transition density of a Markov Process  $X_t$  is the function p(s, x, t, y) such that P(s, x, t, dy) = p(s, x, t, y)dy.

**Remark 2.1.** Recall as well that for homogeneous Markov processes, i.e. those with stationary and independent increments, we can write their transition probabilities as P(s, x, t, B) = P(t - s, x, B).

#### 2.2 Two Lemmas

We now prove two lemmas concerning Markov Processes and their expectations.

**Lemma 2.1.** Let  $f_t(x) = f(t, x) : \mathbb{R}^+ \times S \to \mathbb{R}$  be a bounded integrable function. Then for any time-homogeneous Markov processes  $X_t$  on a state space S, which admits a stationary measure  $\rho$  and a transition density p(t, x, y) with an eigenfunction expansion

$$p(t, x, y)\rho(dy) = \left(1 + \sum_{k=2}^{\infty} e^{-\lambda_k t} \phi_k(x)\phi_k(y)\right)\rho(dy),$$

Where  $\{\phi_i\}$  is an orthonormal basis for  $L^2(S, \rho)$ . Then the following holds:

$$\lim_{t \to \infty} \frac{1}{2t} \mathbb{E}_{X_0} \int_0^t f_s(X_s) ds = \lim_{t \to \infty} \frac{1}{2t} \int_0^t \mathbb{E}_{\rho}[f_s(X_s)] ds$$

*Proof.* Assuming  $X_0 = x_0$ ,

$$\begin{split} \frac{1}{t} \mathbb{E}_{X_0} \Big[ \int_0^t f_s(X_s) ds \Big] &= \frac{1}{t} \int_S \int_0^t f_s(y) ds \ \mathbb{P}^{x_0}(X_s \in dy) = \frac{1}{t} \int_S \int_0^t f_s(y) ds \ p(s, x_0, y) dy \\ &= \frac{1}{t} \int_S \int_0^t f_s(y) \Big( 1 + \sum_{k=2}^\infty e^{-\lambda_k s} \phi_k(x_0) \phi_k(y) \Big) \rho(dy) ds \\ &= \frac{1}{t} \int_S \int_0^t f_s(y) \rho(dy) ds + \frac{1}{t} \int_S \int_0^t f_s(y) \sum_{k=2}^\infty e^{-\lambda_k s} \phi_k(x_0) \phi_k(y) \rho(dy) ds \end{split}$$

Then as  $f_t$  is bounded, let C be the pointwise bound of  $f_s(y)$ , i.e.  $|f_s(y)| < C$ . Further, as  $\{\phi_k\}$  is an orthonormal family,  $||\phi_k(z)||_{L^2(\rho)} \leq 1 \forall z \in S$ . Thus all of the  $\phi_k$  are at least pointwise bounded on S, so  $\forall k \in \mathbb{N}, |\phi_k(x)| \leq B_k \in \mathbb{R}$ . Thus

$$\begin{split} \frac{1}{t} \int_{S} \int_{0}^{t} f_{s}(y) \sum_{k=2}^{\infty} e^{-\lambda_{k}s} \phi_{k}(x_{0}) \phi_{k}(y) \rho(dy) ds &\leq \frac{C}{t} \sum_{k=2}^{\infty} B_{k}^{2} \int_{0}^{t} e^{-\lambda_{k}s} ds \\ &\leq \frac{C}{t} \sum_{k=2}^{\infty} B_{k}^{2} (\frac{-e^{-\lambda_{k}t}}{\lambda_{k}} + \frac{1}{\lambda_{k}}) \end{split}$$

Thus,

$$\lim_{t \to \infty} \frac{1}{t} \int_S \int_0^t f_s(y) \sum_{k=2}^\infty e^{-\lambda_k s} \phi_k(x_0) \phi_k(y) \rho(dy) ds \le \lim_{t \to \infty} \frac{C}{t} \sum_{k=2}^\infty B_k^2 \left(\frac{-e^{-\lambda_k t}}{\lambda_k} + \frac{1}{\lambda_k}\right) \le 0$$

Concluding, by Fubini's theorem and the previous steps,

$$\lim_{t \to \infty} \frac{1}{2t} \mathbb{E}_{X_0} \Big[ \int_0^t f_s(X_s) ds \Big] = \lim_{t \to \infty} \frac{1}{2t} \int_S \int_0^t f_s(y) \rho(dy) ds = \lim_{t \to \infty} \frac{1}{2t} \int_0^t \mathbb{E}_{\rho}[f(X_s)] ds$$
As desired.

As desired.

Corollary 2.1. Similarly, it follows from 2.1 that by just adding another time integral, under the same assumptions as 2.1

$$\lim_{t \to \infty} \frac{1}{2t} \mathbb{E}_{X_0} \int_0^t \int_0^t f_s(X_s) ds_2 \, ds_1 = \lim_{t \to \infty} \frac{1}{2t} \int_0^t \int_0^t \mathbb{E}_{\rho}[f_s(X_s)] ds_2 \, ds_1$$

**Lemma 2.2.** Let  $f_t(x) = f(t, x) : \mathbb{R}^+ \times S \to \mathbb{R}$  be a bounded integrable function. Then assuming  $s_1 < s_2$ , for time homogeneous Markov processes  $X_t$  on a state space S and  $\forall k \in \mathbb{R}^+$ ,

$$\mathbb{E}_{X_k}[f_{s_1}(X_{s_1})f_{s_2}(X_{s_2})] = \mathbb{E}_{X_k}[a(X_{s_1})]$$

Where  $a(x) := \mathbb{E}_{X_k}[f_{s_1}(x)P_{s_2-s_1}f_{s_2}(x)].$ 

*Proof.* Observe that

$$\mathbb{E}_{X_k}[f_{s_1}(X_{s_1})f_{s_2}(X_{s_2})] = \mathbb{E}_{X_k}[\mathbb{E}_{X_k}[f_{s_1}(X_{s_1})f_{s_2}(X_{s_2})|\mathcal{F}_{s_1}]]$$
  
=  $\mathbb{E}_{X_k}[f_{s_1}(X_{s_1})\mathbb{E}_{X_k}[f_{s_2}(X_{s_2})|\mathcal{F}_{s_1}]]$   
=  $\mathbb{E}_{X_k}[f_{s_1}(X_{s_1})\mathbb{E}_{X_{s_1}}[f_{s_2}(X_{s_2-s_1})]]$   
=  $\mathbb{E}_{X_k}[f_{s_1}(X_{s_1})P_{s_2-s_1}f_{s_2}(X_{s_1})]$ 

Where the third equality is justified by  $X_t$  being a Markov process and thus satisfying the Markov property. 

## 3 Effective Diffusivity of the Runner-Tumbler Model

We look to calculate the effective diffusivity of the runner tumbler model given by the following system of SDE's:

$$dX_t = v\cos(\theta_t) + \sqrt{2D}dB_t^x \tag{1}$$

$$dY_t = v\sin(\theta_t) + \sqrt{2D}dB_t^y \tag{2}$$

$$d\theta_t = w + \sqrt{2D_r} dB_t \tag{3}$$

Where v is a constant governing the translational velocity and w is a constant governing the rotational velocity. Further  $B_t$ ,  $B_t^x$ , and  $B_t^y$  are completely uncorrelated Brownian motions.

#### 3.1 Goals

We are looking to first calculate the large-time Mean-Squared Displacement (MSD) for the position of our particle, as well as the Mean-Squared Angular Displacement (MSAD). These expressions are given by:

$$MSD = \mathbb{E}_{\theta_0}[X_t^2] + \mathbb{E}_{\theta_0}[Y_t^2]$$
(4)

$$MSAD = \mathbb{E}_{\theta_0}[\theta_t^2] \tag{5}$$

Where  $\mathbb{E}_{\theta_0}$  is expectation with respect to the law of  $\theta_t$  at t = 0. The MSD and MSAD are measures of the deviation of the position and angle, respectively, of a particle with respect to the initial position  $(X_0, Y_0, \theta_0)$  over time.

We further look for the effective diffusion coefficient,  $D_{eff}$ , of the particle, which describes the rate at which the particle diffuses. The effective diffusion coefficient is given by the expression

$$D_{eff} = \lim_{t \to \infty} \frac{1}{4t} [\text{MSD}] \tag{6}$$

**Theorem 3.1** (MSAD). The MSAD for the Runner-Tumbler model is given by the expression

$$MSAD = w^2 t^2 + 2D_r t \tag{7}$$

*Proof.* We first solve for  $\theta_t$ . By integration and the fact that for all constants  $c, \int_0^t c \, dB_s = c \, B_t$ ,

$$d\theta_t = wdt + \sqrt{2D_r}dB_t$$
$$\theta_t = wt + \sqrt{2D_r}B_t$$

And so we calculate the MSAD:

$$MSAD = \mathbb{E}_{\theta_0}[(wt + \sqrt{2D_r B_t})^2]$$
$$= \mathbb{E}_{\theta_0}[w^2 t^2 + wt \sqrt{2D_r B_t} + 2D_r B_t^2]$$
$$= w^2 t^2 + 2D_r t$$

Where the third equality follows from the fact that  $B_t \sim \mathcal{N}(0, t)$  and we can treat t as a constant when taking expectations. Note that this agrees with (4) in [1], verifying the MSAD.

**Theorem 3.2** (Effective Diffusivity of the Runner-Tumbler Model). The effective diffusivity of the Runner-Tumbler model specified by equations (1), (2), and (3) is given by the expression

$$D_{eff} = \frac{v^2 D_r}{2(D_r^2 + w^2)} + D$$

*Proof.* We begin by performing a substitution to make  $\theta_t$  time homogeneous and thus into an Ito Diffusion, a class of SDE's which exhibit very useful properties. Note that  $d\theta_t = wdt + \sqrt{2D_r}dB_t \implies \theta_t = wt + \sqrt{2D_r}B_t$  is time in-homogeneous. So, we define  $\alpha_t = \theta_t - wt = T_t(\theta_t)$ , where  $T_t(x) = x - wt$ . The, by the Ito Formula,

$$d\alpha_t = -wdt + 1d\theta_t = -wdt + wdt + \sqrt{2D_r}dB_t \tag{8}$$

$$=\sqrt{2D_r}dB_t\tag{9}$$

And so,  $\alpha_t = \sqrt{2D_r}B_t$ , i.e. scaled Brownian motion. We will consider  $\alpha_t$  as Brownian motion on the circle  $S^1 := [0, 2\pi]/\sim$ , as in our calculations, all appearances of  $\alpha_t$  are periodic. Thus  $\alpha(0) = \alpha(2\pi n) \forall n \in \mathbb{N}$ . Note further, that at t = 0,  $\alpha_t = \theta_t$  and thus  $\mathbb{E}_{\alpha_0}[\cdot] = \mathbb{E}_{\theta_0}[\cdot]$ .

Before we continue with the calculations, we must first determine the stationary distribution, if it exists, for the process  $\alpha_t$ . We can obtain the density r(x) for the stationary distribution  $\rho(dy) = r(y)dy$  as the solution to the stationary Fokker-Planck, or Forwards Kolmogorov equation:

$$-\partial_x[b(x)r(x)] + \partial_{xx}\left[\frac{(\sqrt{2D_r})^2}{2}r(x)\right] = 0$$

**Lemma 3.1.** The stationary distribution for  $\alpha_t$  is given by

$$\rho(dy) = \frac{1}{2\pi} dy$$

*Proof.* For  $\alpha_t$ , b(x) as defined above is 0, so we need to solve

$$\partial_{xx}[D_r r(x)] = 0 \tag{10}$$

This is simple however, and

$$D_r^2 r''(x) = 0$$
  

$$r''(x) = 0$$
  

$$r(x) = c_1 x + c_2$$

To determine the constants  $c_1$  and  $c_2$ , consider first that as  $r(x) : S^1 \to \mathbb{R}^+$ and thus  $r(0) = r(2\pi)$ , and so  $r(0) = c_2 = r(2\pi) = c_1 2\pi + c_2$ , and thus  $c_1 = 0$ . Further, r(x) is a density, and so

$$\int_0^{2\pi} r(x)dx = 1 \implies 2c_2\pi = 1 \implies c_2 \equiv \frac{1}{2\pi}$$

Thus

$$\rho(dy) = \frac{1}{2\pi} dy,\tag{11}$$

Proving the lemma.

We next find the density p(t, x, y) such that  $\mathbb{P}^x(\alpha_t \in dy) = p(t, x, y)\rho(dy)$  for the transition probabilities associated to the Markov Process  $\alpha_t$ . We can obtain the transition density from solving the Kolmogorov Forwards equation with dirac delta initial condition, but we rather obtain it from Proposition 2.1 in [3]:

$$p(t, x, y)dy = \left(\frac{1}{\pi} \sum_{n \in \mathbb{N}} \left[ e^{(-D_r n^2 t)} \cos(n(x-y)) \right] + \frac{1}{2\pi} \right) dy$$
(12)

$$p(t, x, y)dy = \left(2\sum_{n \in \mathbb{N}} \left[e^{(-D_r n^2 t)} \cos(n(x-y))\right] + 1\right)\rho(dy)$$
(13)

This transition density is a reformulation of the natural notion of periodic Brownian motion density:  $p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \sum_{n \in \mathbb{Z}} \exp[\frac{-\sqrt{2Dr}}{t} (x - y - 2\pi n)^2]$ 

**Remark 3.1.** Note that as  $t \to \infty$ , p(t, x, y)dy converges absolutely and exponentially fast, with rate  $D_r$  to the stationary distribution  $\rho(dy)$ . Thus for large t, we can safely approximate

$$p(t,x,y)dy \sim \left(\frac{1}{\pi}e^{-D_r t}\cos(x-y) + \frac{1}{2\pi}\right)dy \tag{14}$$

We define functions

$$g_t(x) = \cos(x + wt)$$
$$h_t(x) = \sin(x + wt)$$

So that  $g_t(\alpha_t) = \cos(\theta_t)$  and  $h_t(\alpha_t) = \sin(\theta_t)$ . Thus,

$$\begin{split} \mathbb{E}_{\theta_0}[X_t^2] &= \mathbb{E}_{\alpha_0}[X_t^2] \\ &= \mathbb{E}_{\alpha_0} \left[ (v \int_0^t g_s(\alpha_s) ds + \sqrt{2D} B_t^1)^2 \right] \\ &= v^2 \mathbb{E}_{\alpha_0} \left[ (\int_0^t g_s(\alpha_s) ds)^2 \right] + 2Dt \\ &= v^2 \mathbb{E}_{\alpha_0} \left[ \int_0^t \int_0^t g_{s_1}(\alpha_{s_1}) g_{s_2}(\alpha_{s_2}) ds_2 ds_1 \right] + 2Dt \\ &= v^2 \int_0^t \int_0^t \mathbb{E}_{\alpha_0} \left[ g_{s_1}(\alpha_{s_1}) g_{s_2}(\alpha_{s_2}) \right] ds_2 ds_1 + 2Dt \\ &= 2v^2 \int_0^t \int_{s_1}^t \mathbb{E}_{\alpha_0} \left[ g_{s_1}(\alpha_{s_1}) g_{s_2}(\alpha_{s_2}) \right] ds_2 ds_1 + 2Dt \end{split}$$

Assuming  $s_1 < s_2$ , as the integrand is symmetric over the diagonal of the square  $[0,t] \times [0,t]$ . Similarly for  $Y_t$ ,

$$\begin{split} \mathbb{E}_{\theta_0}[Y_t^2] &= \mathbb{E}_{\alpha_0}[Y_t^2] \\ &= \mathbb{E}_{\alpha_0} \left[ (v \int_0^t h_s(\alpha_s) ds + \sqrt{2D} B_t^1)^2 \right] \\ &= v^2 \mathbb{E}_{\alpha_0} \left[ (\int_0^t h_s(\alpha_s) ds)^2 \right] + 2Dt \\ &= v^2 \mathbb{E}_{\alpha_0} \left[ \int_0^t \int_0^t h_{s_1}(\alpha_{s_1}) h_{s_2}(\alpha_{s_2}) ds_2 ds_1 \right] + 2Dt \\ &= v^2 \int_0^t \int_0^t \mathbb{E}_{\alpha_0} \left[ h_{s_1}(\alpha_{s_1}) h_{s_2}(\alpha_{s_2}) \right] ds_2 ds_1 + 2Dt \\ &= 2v^2 \int_0^t \int_{s_1}^t \mathbb{E}_{\alpha_0} \left[ h_{s_1}(\alpha_{s_1}) h_{s_2}(\alpha_{s_2}) \right] ds_2 ds_1 + 2Dt \end{split}$$

By Lemma 2.2, we can further reduce these expressions to

$$\mathbb{E}_{\alpha_0}[X_t^2] = 2v^2 \int_0^t \int_{s_1}^t \mathbb{E}_{\alpha_0} \bigg[ g_{s_1}(\alpha_{s_1}) P_{s_2 - s_1} g_{s_2}(\alpha_{s_1}) \bigg] ds_2 ds_1 + 2Dt \tag{15}$$

$$\mathbb{E}_{\alpha_0}[Y_t^2] = 2v^2 \int_0^t \int_{s_1}^t \mathbb{E}_{\alpha_0} \bigg[ h_{s_1}(\alpha_{s_1}) P_{s_2-s_1} h_{s_2}(\alpha_{s_1}) \bigg] ds_2 ds_1 + 2Dt$$
(16)

And by Lemma 2.1,

$$\lim_{t \to \infty} \mathbb{E}_{\alpha_0}[X_t^2] = 2v^2 \int_0^t \int_{s_1}^t \mathbb{E}_{\rho} \bigg[ g_{s_1}(\alpha_{s_1}) P_{s_2 - s_1} g_{s_2}(\alpha_{s_1}) \bigg] ds_2 ds_1 + 2Dt \qquad (17)$$

$$\lim_{t \to \infty} \mathbb{E}_{\alpha_0}[Y_t^2] = 2v^2 \int_0^t \int_{s_1}^t \mathbb{E}_{\rho} \bigg[ h_{s_1}(\alpha_{s_1}) P_{s_2 - s_1} h_{s_2}(\alpha_{s_1}) \bigg] ds_2 ds_1 + 2Dt \quad (18)$$

Recalling that

$$P_t f(x) = \mathbb{E}_x[f(X_t)]$$

We can expand

$$P_{s_2-s_1}g_{s_2}(\alpha_{s_1}) = \int_0^{2\pi} g_{s_2}(\alpha_{s_2-s_1}) \mathbb{P}^{\alpha_{s_1}}(\alpha_{s_2-s_1} \in dy)$$
  
= 
$$\int_0^{2\pi} g_{s_2}(y)p(s_2-s_1,x,y)dy$$
  
= 
$$\int_0^{2\pi} g_{s_2}(y) \left(\frac{1}{\pi}e^{-D_r(s_2-s_1)}\cos(x-y) + \frac{1}{2\pi}\right)dy$$

By using the n = 1 rate of convergence determining term in the sum of (12). Thus

$$\lim_{t \to \infty} \mathbb{E}_{\alpha_0} [X_t^2] = 2v^2 \int_0^t \int_{s_1}^t \int_0^{2\pi} \int_0^{2\pi} g_{s_1}(x) g_{s_2}(y) \left(\frac{1}{\pi} e^{-D_r(s_2-s_1)} \cos(x-y) + \frac{1}{2\pi}\right) \frac{1}{2\pi} dy dx ds_2 ds_1 + 2Dt$$
(19)
$$\lim_{t \to \infty} \mathbb{E}_{\alpha_0} [Y_t^2] = 2v^2 \int_0^t \int_{s_1}^t \int_0^{2\pi} \int_0^{2\pi} h_{s_1}(x) h_{s_2}(y) \left(\frac{1}{\pi} e^{-D_r(s_2-s_1)} \cos(x-y) + \frac{1}{2\pi}\right) \frac{1}{2\pi} dy dx ds_2 ds_1 + 2Dt$$
(20)

But we can simplify this integral, as

$$\int_0^{2\pi} \int_0^{2\pi} g_{s_1}(x) g_{s_2}(y) dy dx = 0 = \int_0^{2\pi} \int_0^{2\pi} h_{s_1}(x) h_{s_2}(y) dy dx$$

 $\operatorname{So}$ 

$$\lim_{t \to \infty} \mathbb{E}_{\alpha_0}[X_t^2] = \frac{v^2}{\pi^2} \int_0^t \int_{s_1}^t \int_0^{2\pi} \int_0^{2\pi} g_{s_1}(x) g_{s_2}(y) e^{-D_r(s_2 - s_1)} \cos(x - y) dy dx ds_2 ds_1 + 2Dt$$
(21)

$$\lim_{t \to \infty} \mathbb{E}_{\alpha_0}[Y_t^2] = \frac{v^2}{\pi^2} \int_0^t \int_{s_1}^t \int_0^{2\pi} \int_0^{2\pi} h_{s_1}(x) h_{s_2}(y) e^{-D_r(s_2 - s_1)} \cos(x - y) dy dx ds_2 ds_1 + 2Dt$$
(22)

And the overall expression for the effective diffusion:

$$\lim_{t \to \infty} \frac{v^2}{4\pi^2 t} \int_0^t \int_{s_1}^t \int_0^{2\pi} \int_0^{2\pi} \left( g_{s_1}(x) g_{s_2}(y) + h_{s_1}(x) h_{s_2}(y) \right) e^{-D_r(s_2 - s_1)} \cos(x - y) dy dx ds_2 ds_1 + D \quad (23)$$

Using trig identities, we can rewrite  $g_{s_1}(x)g_{s_2}(y)+h_{s_1}(x)h_{s_2}(y)$  as  $\cos(x-y+w(s_1-s_2))$  and then as

$$\cos(x-y)\cos(w(s_1-s_2)) + \sin(x-y)\sin(w(s_1-s_2))$$

Thus the integrand of (23) is given by the expression

$$e^{-D_r(s_2-s_1)}(\cos^2(x-y)\cos(w(s_1-s_2)) + \sin(x-y)\cos(x-y)\sin(w(s_1-s_2)))$$

To simply this expression, we observe that

$$\int_{0}^{2\pi} \int_{0}^{2\pi} \sin(x-y) \cos(x-y) dy dx = 0$$

And that

$$\int_{0}^{2\pi} \int_{0}^{2\pi} \cos^2(x-y) dy dx = 2\pi^2$$

So the effective diffusion expression (23) reduces to

$$\lim_{t \to \infty} \frac{v^2}{2t} \int_0^t \int_{s_1}^t \cos(w(s_1 - s_2)) e^{\frac{-2D_r(s_2 - s_1)}{2}} ds_2 ds_1 + D$$
(24)

We can explicitly compute the double integral in (24) and obtain the expression

$$D_{eff} = \lim_{t \to \infty} \left( \frac{v^2 (w^2 - D_r^2)}{2t (D_r^2 + w^2)^2} + \frac{v^2 D_r}{2(D_r^2 + w^2)} + \frac{v^2 e^{-D_r t}}{2t (D_r^2 + w^2)^2} [(D_r^2 - w^2) \cos wt - 2w D_r \sin wt] + D \right)$$
(25)

Which finally reduces to

$$D_{eff} = \frac{v^2 D_r}{2(D_r^2 + w^2)} + D$$
(26)

Proving the theorem.

**Remark 3.2.** Note that (26) agrees exactly with the effective diffusion coefficient (6) in [1] when we write  $D_r = 1/\tau_r$ , verifying the results of the paper.

#### Effective Diffusivity of the Anisotropic Model 4

We now look at the effective diffusion coefficients in the x and y directions,  $D_{xx}$ and  $D_{yy}$  respectively, of the model specified in [2] and given by the following system of differential equations:

$$dX_t = U\cos(\theta_t)dt + \sqrt{2k_BT\Gamma_{11}(\theta_t)}\,dB_t^{(x)}$$
(27)

$$dY_t = U\sin(\theta_t)dt + \sqrt{2}k_B T \Gamma_{22}(\theta_t) dB_t^{(y)}$$
(28)

$$d\theta_t = w_c \sin[2(w_h t - \theta_t)]dt + \sqrt{2D_r} dB_t$$
<sup>(29)</sup>

Where  $\Gamma_{ij}(\theta_t) = \bar{D}\delta_{ij} + \frac{1}{2}\Delta D \begin{pmatrix} \cos(2\theta_t) & \sin(2\theta_t) \\ \sin(2\theta_t) & -\cos(2\theta_t) \end{pmatrix}$ , U is the constant translational velocity of the particle and both  $B_t^{(x)}$  and  $B_t^{(y)}$  are uncorrelated with  $\bar{D}_t^{(x)}$ 

 $B_t$ .

#### 4.1 Goals

We look to calculate the effective diffusion coefficients of the model by rigorous stochastic analytic methods. We hope that our results will agree with the results obtained in [2], which were obtained with approximations.

**Remark 4.1** (Making the process  $\theta_t$  time homogeneous). Recall that

$$d\theta_t = w_c \sin[2(w_h t - \theta_t)]dt + \sqrt{2D_r}dB_t \tag{30}$$

And thus  $\theta_t$  is not a time-homogeneous stochastic process. We first make  $\theta_t$  time homogeneous. Define  $\theta'_t := \theta_t - w_h t$ . Then

$$d\theta'_t = -(w_h + w_c \sin(2\theta'_t))dt + \sqrt{2D_r dB_t}$$
(31)

And thus  $\theta'_t$  is a time homogeneous stochastic process. In fact,  $\theta'_t$  is of the form

$$d\theta_t' = b(\theta_t')dt + \sigma dB_t \tag{32}$$

Where  $b(x) = -w_h - w_c \sin(2x)$  and  $\sigma = \sqrt{2D_r}$ . Note that both b(x) and  $\sigma$  are Lipschitz continuous, and thus the process  $\theta'_t$  can be viewed as an Ito diffusion on  $\mathbb{R}$ . As an Ito Diffusion, equation (11) admits a unique solution  $\theta'_t$ .

Further, because b(x) is  $\pi$  periodic, we can write define a stochastic process  $\alpha_t$  on the circle of radius 1/2,  $S := [0, \pi]/\sim$ , such that  $\alpha_t := \theta'_t \pmod{\pi}$ .

**Lemma 4.1** (Stationary Distribution of  $\alpha_t$ ). The stationary distribution  $\mu(dy)$  for  $\alpha_t$  is given by the expression

$$\mu(dy) = Ke^{h(y)/D_r}dy$$

Where  $\forall h(y) = b(y)$  with b(y) defined in remark 4.1:

$$h(y) = -w_h y + \frac{w_c}{2}\cos(2y)$$

And  $K=(\int_0^\pi e^{h(y)/D_r}dy)^{-1}$ 

*Proof.* We first obtaining the density of the stationary distribution as the solution to the stationary Forwards Kolmogorov equation. Ie, we solve the differential equation:

$$\left\{-\partial_x[b(x)r(x)] + \partial_{xx}[D_rr(x)] = 0\right\}$$
(33)

Where  $d\alpha_t = b(\alpha_t)dt + \sqrt{2D_r}DB_t$ ,  $b(x) = -\nabla h(x)$ ,

$$b(x) = -w_h - w_c \sin(2x)$$
$$h(x) = -w_h x + \frac{w_c}{2} \cos(2x)$$

We claim that

$$r(x) = e^{h(x)/D_r} \tag{34}$$

solves the above equation (12) and is thus the stationary density. Observe that

$$-\partial_x[b(x)r(x)] + \partial_{xx}[D_rr(x)] = -b'(x)r(x) - b(x)r'(x) + D_rr''(x)$$

And substituting (13) yields

$$-b'(x)e^{h(x)/D_r} - \frac{b^2(x)}{D_r}e^{h(x)/D_r} + b'(x)e^{h(x)/D_r} + \frac{b^2(x)}{D_r}e^{h(x)/D_r} = 0$$

Thus  $r(x) = e^{h(x)/D_r}$  is indeed the stationary density and so our stationary distribution  $\mu$  can be written as  $\mu(dy) = Kr(y)dy$ , where  $K = (\int_0^{\pi} r(y)dy)^{-1}$ . The normalizing constant, K, is necessary so that  $\mu(S) = \int_0^{\pi} Kr(y)dy = 1$ . Thus the stationary distribution of our process is given by

$$\mu(E) = K \int_E e^{h(y)/D_r} dy \tag{35}$$

Or equivalently,

$$\mu(dy) = Ke^{h(y)/D_r}dy$$

As we have found a distribution that solves the stationary forwards Kolmogorov equation, we obtain a unique stationary distribution for  $\alpha_t$ . We conclude by checking that this distribution is the stationary distribution for  $\alpha_t$  by demonstrating that it is the symmetrizing measure for our space  $L^2(S)$ .

Consider the Hilbert Space  $L^2(S,\mu)$ , where  $\mu$  is the above stationary distribution and  $S = [0,\pi]/\sim$ . Then with the usual  $L^2$  inner product,  $\forall f, g \in L^2(S,\mu)$ ,

$$(\mathcal{L}f,g)_{\mu} = \int_0^{\pi} \mathcal{L}f(x)g(x)r(x)dx$$
(36)

$$=K\int_0^{\pi} (D_r\partial_{xx}^2 f(x) + b(x)\partial_x f(x))g(x)e^{h(x)/D_r}dx$$
(37)

(38)

Note that

$$\partial_x (e^{h(x)/D_r} D_r \partial_x f(x)) = \frac{b(x)}{D_r} D_r \partial_x f(x) e^{h(x)/D_r} + e^{h(x)/D_r} D_r \partial_{xx}^2 f(x) = e^{h(x)/D_r} \mathcal{L}f(x) \quad (39)$$

And so

$$(36) = K \int_0^\pi \partial_x (e^{h(x)/D_r} D_r \partial_x f(x)) g(x) dx$$
(40)

$$= KD_r[g(x)e^{h(x)/D_r}\partial_x f(x)]_0^{\pi} - KD_r \int_0^{\pi} \partial_x g(x)\partial_x f(x)e^{h(x)/D_r}dx \quad (41)$$

$$= KD_r[g(x)e^{h(x)/D_r}\partial_x f(x)]_0^{\pi} - KD_r[\partial_x g(x)e^{h(x)/D_r}f(x)]_0^{\pi} +$$
(42)

$$K \int_{0}^{\pi} f(x)(b(x)\partial_{x}g(x)e^{h(x)/D_{r}} + e^{h(x)/D_{r}}D_{r}\partial_{xx}^{2}g(x)dx$$
(43)

$$=K\int_0^{\pi} e^{h(x)/D_r} \mathcal{L}g(x)f(x)dx \tag{44}$$

$$= (f, \mathcal{L}g)_{\mu} \tag{45}$$

Thus our stationary distribution is in fact the symmetrizing measure.  $\Box$ 

**Lemma 4.2.** The transition density function for  $\alpha_t$ , p(t, x, y) such that

$$\mathbb{P}^x(\alpha_t \in dy) = p(t, x, y)\mu(dy),$$

exists and has a eigen-function expansion

$$p(t,x,y) = 1 + \sum_{i=2}^{\infty} e^{\lambda_i t} \phi_i(x) \phi_i(y)$$

$$\tag{46}$$

With

$$0 = \lambda_1 < \lambda_2 \le \lambda_3 \le \dots \le \infty$$

and  $\{\phi_i\}$  forming an orthnonormal basis for  $L^2(S,\mu)$ .

Proof. As our generator for  $\alpha_t$ ,  $\mathcal{L}$  is defined on a periodic domain, it has periodic boundary conditions and thus regular boundary conditions. By [4] Theorem (3.1), the Green operators corresponding to  $\mathcal{L}$  are compact self-adjoint, and have a discrete spectrum  $0 \leq \mu_1 < \mu_2 \leq \mu_3 \leq \ldots \leq -\infty$  and a corresponding orthonormal basis of  $L^2(S,\mu)$  of eigenfunctions  $\{\phi_i\}$ . Further, we know that 0 is an eigenvalue as the stationary density is a solution corresponding to eigenvalue 0. For all of the non-zero eigenvalues  $\mu_i$  of the Green operator, the eigenvalues of the generator  $\mathcal{L}$  are  $1/\mu_i$ , but with the same eigenfunctions  $\phi_i$ . Thus we have a spectral decomposition for the transition density, which converges uniformly and absolutely in space and exponentially in time to the stationary density  $r(y) = e^{h(y)/D_r}$ ,

$$p(t, x, y) = 1 + \sum_{i=2}^{\infty} e^{\lambda_i t} \phi_i(x) \phi_i(y)$$

$$(47)$$

 $\operatorname{So}$ 

$$\mathbb{P}^x(\alpha_t \in dy) = p(t, x, y)\mu(dy) = \left(1 + \sum_{i=2}^{\infty} e^{\lambda_i t} \phi_i(x)\phi_i(y)\right) K e^{h(y)/D_r} dy \quad (48)$$

**Theorem 4.1.** The effective diffusion coefficients in the x and y coordinates,  $D_{xx}$  and  $D_{yy}$  respectively, for the Anisotropic model specified by equations (27), (28), and (29) is given by the identical expression for both  $D_{xx}$  and  $D_{yy}$ :

$$D_{xx} = \bar{D} + \frac{K_1 U^2 \lambda_2}{2(\lambda_2^2 + w_h^2)} = D_{yy}$$

Where  $\lambda_2$  is the second eigenvalue in the eigen-function decomposition of p(t, x, y) defined in Lemma (4.2) and  $K_1$  is given by the expression

$$K_1 = \int_0^{\pi} \int_0^{\pi} \cos(x - y)\phi_2(x)\phi_2(y)\mu(dy)\mu(dx)$$

With  $\phi_2$  being the eigenfunction corresponding to  $\lambda_2$ .

*Proof.* Recall that our effective diffusion coefficients for both coordinates are given by the expressions:

$$D_{xx} := \lim_{t \to \infty} \frac{U^2}{t} \left( \int_0^t \int_{s_1}^t \mathbb{E}_{\theta_0} [\cos \theta_{s_1} \cos \theta_{s_2}] \, ds_2 \, ds_1 \right. \\ \left. + 2\bar{D}t + \Delta D \int_0^t \mathbb{E}_{\theta_0} [\cos(2\theta(s))] \, ds \right)$$
(49)

$$D_{yy} := \lim_{t \to \infty} \frac{U^2}{t} \left( \int_0^t \int_{s_1}^t \mathbb{E}_{\theta_0} [\sin \theta_{s_1} \sin \theta_{s_2}] \, ds_2 \, ds_1 \right. \\ \left. + 2\bar{D} t - \Delta D \int_0^t \mathbb{E}_{\theta_0} [\cos(2\theta(s))] \, ds \right)$$
(50)

In order to make these integrals in terms of  $\alpha_t$ , we introduce the functions

$$g_t(x) = \cos(x + w_h t)$$
$$h_t(x) = \sin(x + w_h t)$$

So that  $g_t(\alpha_t) = \cos(\theta_t)$  and  $h_t(\alpha_t) = \sin(\theta_t)$ . Further, observe that  $\alpha_t = \theta_t - w_h t \pmod{\pi}$ , so at t = 0,  $\alpha_0 = \theta_0 \pmod{\pi}$ , and so on our new state space S,  $\alpha_0 = \theta_0$ .

Thus we can rewrite 49 and 50 as

$$D_{xx} = \lim_{t \to \infty} \frac{1}{2t} \left( 2 \int_0^t \int_{s_1}^t U(s_1) U(s_2) \mathbb{E}_{\alpha_0} [g_{s_1}(\alpha_{s_1}) g_{s_2}(\alpha_{s_2})] ds_2 ds_1 + 2\bar{D}t + \Delta D \int_0^t \mathbb{E}_{\alpha_0} [g_{2s}(2\alpha_s)] ds \right)$$
(51)

$$D_{yy} = \lim_{t \to \infty} \frac{1}{2t} \left( 2 \int_0^t \int_{s_1}^t U(s_1) U(s_2) \mathbb{E}_{\alpha_0} [h_{s_1}(\alpha_{s_1}) h_{s_2}(\alpha_{s_2})] ds_2 ds_1 + 2\bar{D}t - \Delta D \int_0^t \mathbb{E}_{\alpha_0} [g_{2s}(2\alpha_s)] ds \right)$$
(52)

Let's now first calculate the single time integral in 51 and 52:

$$\lim_{t \to \infty} \frac{1}{2t} \Delta D \int_0^t \mathbb{E}_{\alpha_0}[g_{2s}(2\alpha_s)] ds$$
(53)

By Lemma 2.1,

$$(53) = \lim_{t \to \infty} \frac{\Delta D}{2t} \int_0^t \int_0^\pi g_{2s}(2x)\mu(dx)ds$$
 (54)

By trig identities, we can write  $g_{2s}(2x) = \cos(2x)\cos(2w_h s) - \sin(2x)\sin(2w_h s)$ and use Fubini to perform the time integration first in 53. Thus,

$$(53) = \lim_{t \to \infty} \left( \frac{\Delta D \sin(2w_h t)}{4w_h t} \int_S \dots dx + \frac{\Delta D \sin(w_h t)^2}{2w_h t} \int_S \dots dx \right)$$
(55)

$$=0$$
(56)

Consequently, the single integrals do not contribute to the long term effective diffusivity of the particle. We are now left to calculate the expressions

$$\lim_{t \to \infty} \frac{U^2}{t} \left( \int_0^t \int_{s_1}^t \mathbb{E}_{\alpha_0}[g_{s_1}(\alpha_{s_1})g_{s_2}(\alpha_{s_2})]ds_2ds_1 \right) + \bar{D}$$
$$\lim_{t \to \infty} \frac{U^2}{t} \left( \int_0^t \int_{s_1}^t \mathbb{E}_{\alpha_0}[h_{s_1}(\alpha_{s_1})h_{s_2}(\alpha_{s_2})]ds_2ds_1 \right) + \bar{D}$$

By Lemma 2, we write

$$\mathbb{E}_{\alpha_0}[g_{s_1}(\alpha_{s_1})g_{s_2}(\alpha_{s_2})] = \int_0^\pi \int_0^\pi g_{s_1}(x)g_{s_2}(y)p(s_2 - s_1, x, y)\mu(dx)\mu(dy) \quad (57)$$

$$\mathbb{E}_{\alpha_0}[h_{s_1}(\alpha_{s_1})h_{s_2}(\alpha_{s_2})] = \int_0^\pi \int_0^\pi h_{s_1}(x)h_{s_2}(y)p(s_2 - s_1, x, y)\mu(dx)\mu(dy) \quad (58)$$

Now, we use the fact that p(t, x, y) converges exponentially fast to the stationary density, and that the rate of convergence is governed by  $\lambda_2$  to approximate  $p(t, x, y) \sim 1 + e^{-\lambda_2 t} \phi_2(x) \phi_2(y)$  Thus we write 57 as

$$\mathbb{E}_{\alpha_0}[g_{s_1}(\alpha_{s_1})g_{s_2}(\alpha_{s_2})] = \int_0^\pi \int_0^\pi g_{s_1}(x)g_{s_2}(y)\mu(dx)\mu(dy) + \int_0^\pi \int_0^\pi g_{s_1}(x)g_{s_2}(y)e^{-\lambda_2(s_2-s_1)}\phi_2(x)\phi_2(y)\mu(dy)\mu(dx)$$
(59)

$$\mathbb{E}_{\alpha_0}[h_{s_1}(\alpha_{s_1})h_{s_2}(\alpha_{s_2})] = \int_0^\pi \int_0^\pi h_{s_1}(x)h_{s_2}(y)\mu(dx)\mu(dy) + \int_0^\pi \int_0^\pi h_{s_1}(x)h_{s_2}(y)e^{-\lambda_2(s_2-s_1)}\phi_2(x)\phi_2(y)\mu(dy)\mu(dx) \quad (60)$$

We now use the specific trig properties of  $g_s(x)$  and  $h_s(x)$  and write them as

$$g_{s_1}(x)g_{s_2}(y) = \frac{1}{2} \bigg( \cos(x+y)\cos(w_h(s_1+s_2)) - \sin(x+y)\sin(w(s_1+s_2)) + \cos(x-y)\cos(w(s_1-s_2)) - \sin(x-y)\sin(w(s_1-s_2)) \bigg)$$
(61)

$$h_{s_1}(x)h_{s_2}(y) = \frac{1}{2} \left( -\cos(x+y)\cos(w_h(s_1+s_2)) + \sin(x+y)\sin(w(s_1+s_2)) + \cos(x-y)\cos(w(s_1-s_2)) - \sin(x-y)\sin(w(s_1-s_2)) \right)$$
(62)

Splitting 61 and 62 and distributing among the integrals, we write

$$D_{xx} = \bar{D} + \lim_{t \to \infty} \frac{U^2}{2t} \left( \sum_{k=1}^4 I_k + J_k \right)$$
(63)

$$D_{yy} = \bar{D} + \lim_{t \to \infty} \frac{U^2}{2t} \left( \sum_{k=1}^4 I'_k + J'_k \right)$$
(64)

Where

$$I_{k} := \int_{0}^{t} \int_{s_{1}}^{t} \int_{0}^{\pi} \int_{0}^{\pi} c_{k}(x, y, s_{1}, s_{2}) \mu(dy) \mu(dx) ds_{2} ds_{1}$$
(65)  
$$J_{k} := \int_{0}^{t} \int_{s_{1}}^{t} \int_{0}^{\pi} \int_{0}^{\pi} c_{k}(x, y, s_{1}, s_{2}) e^{-\lambda_{2}(s_{2} - s_{1})} \phi_{2}(x) \phi_{2}(y) \mu(dy) \mu(dx) ds_{2} ds_{1}$$
(66)

$$I'_{k} := \int_{0}^{t} \int_{s_{1}}^{t} \int_{0}^{\pi} \int_{0}^{\pi} d_{k}(x, y, s_{1}, s_{2}) \mu(dy) \mu(dx) ds_{2} ds_{1}$$

$$(67)$$

$$J'_{k} := \int_{0}^{t} \int_{s_{1}}^{s} \int_{0}^{s} \int_{0}^{s} d_{k}(x, y, s_{1}, s_{2}) e^{-\lambda_{2}(s_{2}-s_{1})} \phi_{2}(x) \phi_{2}(y) \mu(dy) \mu(dx) ds_{2} ds_{1}$$

$$\tag{68}$$

$$c_1(x, y, s_1, s_2) := \cos(x+y)\cos(w_h(s_1+s_2)) \tag{69}$$

$$c_2(x, y, s_1, s_2) := -\sin(x+y)\sin(w_h(s_1+s_2))$$
(70)

$$c_3(x, y, s_1, s_2) := \cos(x - y)\cos(w_h(s_1 - s_2))$$
(71)

$$c_4(x, y, s_1, s_2) := -\sin(x - y)\sin(w_h(s_1 - s_2))$$
(72)

$$l_1(x, y, s_1, s_2) := -c_1(x, y, s_1, s_2) \tag{73}$$

$$d_2(x, y, s_1, s_2) := -c_2(x, y, s_1, s_2) \tag{74}$$

$$d_3(x, y, s_1, s_2) := c_3(x, y, s_1, s_2) \tag{75}$$

$$d_4(x, y, s_1, s_2) := c_4(x, y, s_1, s_2) \tag{76}$$

However, we claim that  $\forall k \in \{1, 2, 3, 4\}$ ,  $\lim_{t\to\infty} \frac{U^2}{2t}I_k = \lim_{t\to\infty} \frac{U^2}{2t}I'_k = 0$ . Note that  $I'_k$  differs from  $I_k$  by at most a sign, so if we can show that  $\forall k$ ,  $\lim_{t\to\infty} \frac{U^2}{2t}I_k = 0$ , then our claim is true.  $\forall k \in \{1, 2, 3, 4\}$ , note that by Fubini we can perform the time integration first, and treat the double space integral as constant with regards to t.

Observe that

$$\lim_{t \to \infty} \frac{U^2}{2t} I_1 = \lim_{t \to \infty} \frac{U^2 \sin^2\left(\frac{w_h t}{2}\right) \cos(w_h t)}{t w_h^2} \int_0^\pi \int_0^\pi \cos(x+y) \mu(dy) \mu(dx) = 0$$
(77)

$$\lim_{t \to \infty} \frac{U^2}{2t} I_2 = \lim_{t \to \infty} -\frac{U^2(\sin(2w_h t) - 2\sin(w_h t))}{4tw_h^2} \int_0^\pi \int_0^\pi \sin(x+y)\mu(dy)\mu(dx) = 0$$
(78)

$$\lim_{t \to \infty} \frac{U^2}{2t} I_3 = \lim_{t \to \infty} \frac{U^2 (1 - \cos(w_h t))}{2t w_h^2} \int_0^\pi \int_0^\pi \cos(x - y) \mu(dy) \mu(dx) = 0$$
(79)

$$\lim_{t \to \infty} \frac{U^2}{2t} I_4 = \lim_{t \to \infty} -\frac{U^2(w_h t - \sin(w_h t))}{2tw_h^2} \int_0^\pi \int_0^\pi \sin(x - y)\mu(dy)\mu(dx)$$
(80)

$$= \frac{-U^2}{2w_h} \int_0^{\pi} \int_0^{\pi} \sin(x-y)\mu(dy)\mu(dx)$$
(81)  
$$-U^2 \left(\int_0^{\pi} \int_0^{\pi} \int_0^{\pi} (x-y)\mu(dy)\mu(dx) - \int_0^{\pi} \int_0^{\pi} \int_0^{\pi} (x-y)\mu(dy)\mu(dx) - \int_0^{\pi} \int_0^{\pi} \int_0^{\pi} (x-y)\mu(dy)\mu(dx) - \int_0^{\pi} \int_0^{\pi} \int_0^{\pi} \int_0^{\pi} \int_0^{\pi} (x-y)\mu(dy)\mu(dx) - \int_0^{\pi} \int_0^{\pi}$$

$$= \frac{-U^2}{2w_h} \left( \int_0^{\pi} \int_0^{\pi} \sin(x) \cos(y) \mu(dy) \mu(dx) - \int_0^{\pi} \int_0^{\pi} \cos(x) \sin(y) \mu(dy) \mu(dx) \right)$$
(82)  
= 0 (83)

Thus  $\forall k \in \{1, 2, 3, 4\}$ ,  $\lim_{t \to \infty} \frac{U^2}{2t} I_k = 0 = \lim_{t \to \infty} \frac{U^2}{2t} I'_k$ , as claimed.

We now further claim that  $\lim_{t\to\infty} \frac{U^2}{2t}J_1 = 0 = \lim_{t\to\infty} \frac{U^2}{2t}J_2$ . This consequently would imply that  $\lim_{t\to\infty} \frac{U^2}{2t}J_1' = 0 = \lim_{t\to\infty} \frac{U^2}{2t}J_2'$ , as  $J_1'$ ,  $J_2'$  differ

And

from  $J_1$  and  $J_2$  by only a sign. To see this, we again use Fubini's theorem to perform the time integration first, and obtain

$$\lim_{t \to \infty} \frac{U^2}{2t} J_1 = \lim_{t \to \infty} \frac{U^2(\cos(tw_h) \left(w_h e^{-\lambda_2 t} + \lambda_2 \sin(tw_h) - w_h \cos(tw_h)\right))}{2tw_h \left(\lambda_2^2 + w_h^2\right)} \int_0^{\pi} \int_0^{\pi} \dots \mu(dy) \mu(dx)$$

$$= 0 \tag{85}$$

$$\lim_{t \to \infty} \frac{U^2}{2t} J_2 = \lim_{t \to \infty} \frac{U^2(\lambda_2 \sin^2(tw_h) + w_h e^{-\lambda_2 t} \sin(tw_h) - \frac{1}{2}w_h \sin(2tw_h))}{2t(\lambda_2^2 w_h + w_h^3)} \int_0^{\pi} \int_0^{\pi} \dots \mu(dy) \mu(dx)$$

$$= 0 \tag{86}$$

$$= 0 \tag{87}$$

Thus as  $t \to \infty$ ,  $\frac{U^2}{2t}J_1 = \frac{U^2}{2t}J_2 = \frac{U^2}{2t}J'_1 = \frac{U^2}{2t}J'_2 = 0$  Further, one can immediately observe that  $J_3 = J'_3$  and  $J_4 = J'_4$ , and so  $D_{xx} = D_{yy}$ ! It remains to calculate

$$\lim_{t \to \infty} \frac{U^2}{2t} J_3 \tag{88}$$

$$\lim_{t \to \infty} \frac{U^2}{2t} J_4 \tag{89}$$

Again we Fubini the quadruple integrals to first perform the time integration and obtain the expressions

$$\lim_{t \to \infty} \frac{U^2}{2t} J_3 = \lim_{t \to \infty} \frac{U^2}{2t} \frac{e^{-\lambda_2 t} \left( \left(\lambda_2^2 - w_h^2\right) \cos(tw_h) \right)}{\left(\lambda_2^2 + w_h^2\right)^2} + \frac{e^{\lambda_2 t} \left(\lambda_2^3 t - \lambda_2^2 + \lambda_2 tw_h^2 + w_h^2\right) - 2\lambda_2 w_h \sin(tw_h)\right)}{\left(\lambda_2^2 + w_h^2\right)^2} K_1 \quad (90)$$

 $\operatorname{So}$ 

$$\lim_{t \to \infty} \frac{U^2}{2t} J_3 = \frac{K_1 U^2}{2} \frac{\lambda_2^3 + \lambda_2 w_h^2}{(\lambda_2^2 + w_h^2)^2}$$
(91)

$$\lim_{t \to \infty} \frac{U^2}{2t} J_4 = \lim_{t \to \infty} \frac{U^2}{2t} \frac{-e^{-\lambda_2 t} \left( w_h \left( -e^{\lambda_2 t} \right) \left( \lambda_2^2 t - 2\lambda_2 + t w_h^2 \right) \right)}{\left( \lambda_2^2 + w_h^2 \right)^2} + \frac{\left( w_h^2 - \lambda_2^2 \right) \sin(t w_h) - 2\lambda_2 w_h \cos(t w_h) \right)}{\left( \lambda_2^2 + w_h^2 \right)^2} K_2 \quad (92)$$

So, likewise

$$\lim_{t \to \infty} \frac{U^2}{2t} J_4 = \frac{K_2 U^2}{2} \frac{\lambda_2^2 w_h + w_h^3}{(\lambda_2^2 + w_h^2)^2}$$
(93)

Where

$$K_1 = \int_0^\pi \int_0^\pi \cos(x - y)\phi_2(x)\phi_2(y)\mu(dy)\mu(dx)$$
(94)

$$K_2 = \int_0^\pi \int_0^\pi \sin(x - y)\phi_2(x)\phi_2(y)\mu(dy)\mu(dx)$$
(95)

And so,

$$D_{xx} = \bar{D} + \frac{K_1 U^2}{2} \frac{\lambda_2^3 + \lambda_2 w_h^2}{(\lambda_2^2 + w_h^2)^2} + \frac{K_2 U^2}{2} \frac{\lambda_2^2 w_h + w_h^3}{(\lambda_2^2 + w_h^2)^2} = D_{yy}$$
(96)

We conclude by claiming that  $K_2 = 0$ . To see this, observe that

$$K_1 + iK_2 = \int_0^{\pi} \int_0^{\pi} (\cos(x - y) + i\sin(x - y))\phi_2(x)\phi_2(y)\mu(dy)\mu(dx)$$
(97)

$$= \int_{0}^{\pi} \int_{0}^{\pi} e^{ix} e^{-iy} \phi_2(x) \phi_2(y) \mu(dy) \mu(dx)$$
(98)

$$= \int_{0}^{\pi} e^{ix} \phi_{2}(x) \mu(dx) \int_{0}^{\pi} e^{-iy} \phi_{2}(y) \mu(dy)$$
(99)

$$= \int_{0}^{\pi} e^{ix} \phi_{2}(x) \mu(dx) \overline{\int_{0}^{\pi} e^{iy} \phi_{2}(y) \mu(dy)}$$
(100)

$$= \left| \int_0^{\pi} e^{ix} \phi_2(x) \mu(dx) \right|^2 \in \mathbb{R}$$
(101)

And thus as  $K_1 + iK_2$  is real,  $K_2$  must be 0. Finally,

$$D_{xx} = \bar{D} + \frac{K_1 U^2 \lambda_2}{2(\lambda_2^2 + w_h^2)} = D_{yy}$$
(102)

### 5 Discussion and Further Directions

Observe that up to the constant  $K_1$ , and noticing that  $w_h$  controls the rotation of the anisotropic particle, exactly like w controls the rotation of the spherical particle in section (3), the effective diffusivity of the anisotropic particle is very similar to the effective diffusivity of the previous spherical particle. This is a fascinating result as it states that in large time, these two very different particles act almost exactly alike. This will hopefully result in simplified computations for future projects.

For further directions, one could investigate whether particles with other geometries and propulsions still behave like the self-propelled spherical particle. One also could add a reflecting or absorbing boundary to the processes  $X_t$  and  $Y_t$  and observe the long term behaviour of a bounded particle, i.e. one in a cell or organ.

### References

- Ebbens S, et al. (2010). Self-assembled autonomous runners and tumblers. Phys Rev E Stat Nonlin Soft Matter Phys. 2010 Jul;82(1 Pt 2):015304.
- [2] Fan, Wai-Tong & Pak, On Shun & Sandoval, Mario. (2017). Ellipsoidal Brownian self-driven particles in a magnetic field. Physical Review E. 95. 10.1103/PhysRevE.95.032605.
- [3] Kouloumpou, Dimitra & Papanicolaou, Vassilis. (2012). The random motion on the sphere generated by the Laplace-Beltrami operator. Journal of Applied Functional Analysis. 7.
- [4] McKean, Henry. (1956). Elementary Solutions for Certain Parabolic Partial Differential Equations. Transactions of The American Mathematical Society - TRANS AMER MATH SOC. 82. 519-519. 10.1090/S0002-9947-1956-0087012-3.

## Bernoulli Percolation on Hyperbolic Graphs

Christine Sullivan Advisor: Chris Connell

July 2019

#### Abstract

A *p*-Bernoulli bond percolation on an infinite, Gromov hyperbolic graph X is performed by considering each edge in X and removing the edge with probability 1 - p. This results in a random subgraph of X consisting of finite components and possibly infinite components. Gromov hyperbolic graphs have a well-defined boundary at infinity with a natural topology. We observe properties of the percolation subgraph, especially the behavior of the boundary at infinity of the infinite components of the percolation subgraph.

### 1 Introduction

All graphs discussed in this paper can be assumed to be connected, infinite, and locally finite. Given a *p*-Bernoulli bond percolation on a graph X, the percolation subgraph  $\omega$  is a subgraph of X such that  $V(\omega) = V(X)$  and each edge of Xis in  $\omega$  with probability p. Bernoulli percolation is an *invariant percolation* on a graph X, meaning that it is invariant under the automorphism group of X, Aut(X). The graph X is *transitive* if Aut(G) acts transitively on the vertices of X. The graph X is *quasi-transitive* if the action of Aut(X) on V(X) has finitely many orbits.

By Kolmogrov's 0-1 law, the probability of a percolation subgraph containing an infinite component is either 0 or 1, depending on the value of p and the graph X. Let  $p_c(X)$  be the infimum of the set of  $p \in [0,1]$  such that the p-Bernoulli percolation on X has an infinite component a.s. Let  $p_u(X)$  be the infimum of the set of  $p \in [0,1]$  such that the p-Bernoulli percolation on X has exactly 1 infinite component a.s.

Given a *p*-Bernoulli percolation  $\omega$  on a graph X,  $\hat{\omega}$  is a (1 - p)-Bernoulli percolation on X such that  $E(\hat{\omega}) = E(X) \setminus E(\omega)$ .

A graph X is *Gromov hyperbolic* if all triangles in G are  $\delta$ -thin. A triangle is  $\delta$ -thin if there exists a  $\delta > 0$  such that each side lies in the  $\delta$ -thickening of the other two sides. For some  $\delta$ , if all triangles in X are  $\delta$ -thin, then X is called  $\delta$ -hyperbolic and  $\delta$  is called the hyperbolicity constant.

Given a hyperbolic graph X, the boundary of X at infinity the is set of equivalence classes of geodesic rays to infinity. It is denoted  $\partial_{\infty} X = \{\gamma :$ 

 $[0,\infty) \to X\}/\sim$ . For geodesics  $\gamma_1$  and  $\gamma_2$ ,  $\gamma_1 \sim \gamma_2$  if  $d(\gamma(t), \gamma_2(t)) < C$ , for all  $t \ge 0$  and some constant C > 0.

Given some point  $o \in X$ , the *Gromov Product* of x and y in X is as follows:

$$(x \cdot y)_o = \frac{1}{2} [d(x, o) + d(y, o) - d(x, y)]$$

Then the Gromov product of two points  $\xi$  and  $\eta$  in the boundary at infinity of X is

$$(\xi,\eta)_o = \lim_{\substack{x \to \xi \\ y \to \eta}} = \frac{1}{2} [d(x,o) + d(y,o) - d(x,y)]$$

We define a quasi-metric  $d_o(\xi, \eta)$  on  $\partial_{\infty} X$  as  $d_o(\xi, \eta) = e^{-(\xi \cdot \eta)_o}$ .

Fix a point  $o \in X$ . Then  $B(\xi, \epsilon) = \{\eta : d_o(\xi, \eta) < \epsilon\}$  is an open ball of radius  $\epsilon$ . Let the subbasis for a topology on the boundary of X be the set of open balls:

$$\{B(\xi,\epsilon):\xi\in\partial\infty X,\epsilon>0\}$$

This topology is independent of the choice of the point o.

One important characterization of hyperbolic spaces is that balls grow exponentially. The size of balls of radius r in a hyperbolic graph is approximately  $e^{hr}$  where h is the Hausdorff-dimension of X. Lemma 1.1 uses this exponential growth and in fact characterizes hyperbolic graphs [2].

**Lemma 1.1** Let X be a hyperbolic graph. Let  $y \in S(o, R)$  be the endpoint of a path of length s from  $x \in S(o, R)$ . Assume that R is much greater than s. Then there are constants c and b that depend on R and the hyperbolicity constant of X, such that  $d(x, y) \leq c \log_b(s)$ .

### 2 Boundary at Infinity

The following theorem is a generalization from Theorem 4.1 from [1]

**Theorem 2.1** Let X be a unimodular, quasi-transitive, hyperbolic graph. Let  $\omega$  be a Bernoulli percolation of X. Then a.s. every infinite component of  $\omega$  contains a path that has a unique limit point  $\partial_{\infty} X$ .

The following lemma is from Theorem 1.3 in [5]. Theorem 1.3 is stated for Cayley graphs, but it is remarked that the theorem applies more generally to transitive graphs as well. This trivially applies to quasi-transitive

**Lemma 2.2** Let X be a unimodular, quasi-transitive graph. Let  $\omega$  be a p-Bernoulli percolation of X such that  $p_c(X) . Then a.s. every infinite$  $component of <math>\omega$  is transient and any simple random walk on it has positive drift.

*Proof* [Proof of Theorem 2.1] Let X be a unimodular quasi-transitive hyperbolic graph. Let  $\omega$  be a p-Bernoulli percolation of X such that  $p_c(X) .$ 

Let Z(t) be a simple random walk on an infinite component  $K_i$  of  $\omega$ . By Lemma 2.2, there is some  $\lambda > 0$  such that for sufficiently large t,

$$d(Z(t), Z(0)) \ge \lambda t$$

Let R(t) be the distance between Z(0) and Z(t). Then let u be the point on the line [Z(0), Z(t+s)] such that d(u, Z(0)) = R(t). Then by triangle inequality,

$$d(u, Z(t+s)) \ge d(Z(t+s), Z(t)) - d(u, Z(t))$$

Also note:

$$d(Z(0), Z(t+s)) = d(Z(0), u) + d(u, Z(t+s))$$
  
=  $R(t) + d(u, Z(t+s))$   
=  $d(Z(0), Z(t)) + d(u, Z(t+s)).$ 

The path from Z(t) to Z(t+s) has maximum length s. For a path of length s, the maximum lateral distance moved is  $c \log_b(s)$  for some constants c and b that depend on the hyperbolicity constant  $\delta$ . Hence,

$$\begin{aligned} d(u, Z(t)) &\leq c \log_b(s), \\ d(u, Z(t+s)) &\geq d(Z(t+s), Z(t)) - c \log_b(s), \\ (Z(t) \cdot Z(t+s))_{Z(0)} &= \frac{1}{2} [d(Z(t), Z(0)) + d(Z(t+s), Z(0)) - d(Z(t), Z(t+s))] \\ &= \frac{1}{2} [d(Z(t), Z(0)) + d(Z(0), u) + d(Z(t+s), u) - d(Z(t), Z(t+s))] \\ &\geq \frac{1}{2} [2d(Z(t), Z(0)) + d(Z(t+s), Z(t)) - c \log_b(s) - d(Z(t), Z(t+s))] \\ &= d(Z(t), Z(0)) - \frac{c}{2} \log_b(s) \\ &\geq \lambda t - \frac{c}{2} \log_b(s) \end{aligned}$$

Then we have an upper bound on the distance between Z(t) and Z(t+s) with respect to Z(0):

$$d_{Z(0)}(Z(t), Z(t+s)) = e^{-(Z(t), Z(t+s))_{Z(0)}} \le e^{\lambda t + \frac{c}{2}\log_b(s)} = e^{-\lambda t} s^{\frac{c}{2\log(b)}}$$

Let  $C = \frac{c}{2\log(b)}$ . Fix some big t and consider the sequence  $Z(t), Z(2t), Z(3t), \dots$ 

Let  $[a, b]_{\infty}$  denote the point in  $\partial_{\infty} X$  that is touched by the ray to infinity extended from [a, b].

Then the distance with respect to  $d_{Z(0)}$  between  $[Z(0), Z(t)]_{\infty}$  and  $[Z(0), Z(kt)]_{\infty}$ is:

$$d_{Z(0)}([Z(0), Z(t)]_{\infty}, [Z(0), Z(kt)]_{\infty}) \le \sum_{i=1}^{k-1} d_{Z(0)}([Z(0), Z(it)]_{\infty}, [Z(0), Z(it+t)]_{\infty})$$
$$\le \sum_{i=1}^{k-1} t^{C} e^{-i\lambda t}$$

$$\lim_{k \to \infty} d_{Z(0)}([Z(0), Z(t)]_{\infty}, [Z(0), Z(kt)]_{\infty}) = \sum_{i=1}^{\infty} t^{C} e^{-i\lambda t} < \infty$$

Given some > 0, pick T such that  $t^C e^{-\lambda tT} < \frac{\epsilon}{2}$ . Assume that t is large enough such that  $e^{-\lambda t} < \frac{1}{2}$ . Then:

$$d_{Z(0)}([Z(0), Z(tT)]_{\infty}, [Z(0), Z(\infty)]_{\infty}) \le \sum_{i=T}^{\infty} t^{C} e^{-i\lambda t} < \sum_{i=0}^{\infty} \frac{t^{C} e^{-\lambda tT}}{2^{i}} = t^{C} e^{-\lambda tT} < \frac{\epsilon}{2}$$

Hence for any  $\epsilon > 0$ , there is a T such that for n, m > T,

$$|d_{Z(0)}([Z(0)Z(nt)]_{\infty}, [Z(0)Z(mt)]_{\infty})| < \epsilon.$$

So the sequence  $Z(t), Z(2t), Z(3t), \dots$  is a Cauchy sequence that converges to  $Z(\infty)$ .

The following lemma is a generalization of Lemma 4.3 from [1]

**Lemma 2.3** Let X be a unimodular, quasi-transitive, hyperbolic graph. Let  $\omega$  be a Bernoulli percolation of X. Let  $Y = \partial_{\infty} \bigcup_{i} K_{i}$  be the set of points  $\xi$  in  $\partial_{\infty} X$  such that there is a path to  $\xi$  in  $\omega$  Then a.s. Y = 0 or Y is dense in  $\partial_{\infty} X$ .

*Proof* Given some  $x \in X$ , let  $d_x$  be the Gromov metric with respect to x. Let a(v) be the radius of the biggest ball in  $\partial_{\infty}X \setminus Y$  in the  $d_v$  metric. Note that a(v) does not depend on the choice of v. Let  $o \in X$ ,  $\epsilon \in (0, 1)$ , and let  $\delta$  be the probability that  $\epsilon < a(o) < 1 - \epsilon$ .

Suppose  $\delta > 0$ . Let R > 0 be very large and let  $x \in X$  be a uniform random point on the sphere of radius R about o in X. Given that  $\epsilon < a(o) < 1 - \epsilon$ , the event that the geodesic ray from o containing x hits  $\partial_{\infty} X$  at a point  $\xi_0$  with  $d_o(\xi_0, Y) \ge \frac{\epsilon}{4}$  happens with probability greater than  $\frac{\epsilon}{4}$ . This event happens when  $\xi_0$  is in a ball of radius greater than  $\frac{\epsilon}{4}$  inside the largest ball of  $\partial_{\infty} X \setminus Y$ with respect to  $d_o$ . On that event, if R is very large we can make a(x) as close to 1 as we want, with  $a(x) \neq 1$ . But we also have

$$P[a(x) \in (1-t,1)] = P[a(o) \in (1-t,1)] \to 0$$

as t decreases to zero. It follows that  $P[a(x) \in (\epsilon, 1 - \epsilon)] = 0$ . Hence a.s.  $a(x) \in \{0, 1\}$ . The set Y cannot be a single point, because then there would be a finite measure on X which is automorphism invariant.

The following lemma gives the result of Lemma 2.3 with a different argument.

**Lemma 2.4** Let X be a hyperbolic graph such that for all open sets  $U \subseteq X$  the open cone  $C_o(U) = \{\gamma = [o, \xi] : \xi \in U\}$  has critical percolation  $p_c(C_o(U)) \leq p_c(X)$ . Let  $\omega$  be a p-Bernoulli percolation on X with  $p > p_c(X)$ . Let  $K_i$  be a component of  $\omega$  such that  $|K_i| = \infty$ . Then a.s.  $\partial_{\infty} X = \bigcup_i \partial_{\infty} K_i$ .

*Proof* Suppose for contradiction that  $\partial_{\infty} X \setminus \bigcup_{i} \partial_{\infty} K_{i} \neq \emptyset$ .

Then  $\partial_{\infty}X \setminus \overline{\bigcup \partial_{\infty}K_i} = O$  for some open set  $O \subset \partial_{\infty}X$ . Let U be a proper

open subset of O such that  $\overline{U} \subset O$ . Let  $C_o(U)$  be the open cone consisting of geodesic rays  $[o, \theta]$  for all  $\theta \in U$ . Consider  $C_o(U)$  as a separate graph. Since no  $K_i$  has boundary in U, there is no infinite component in  $C_o(U)$ . But  $p > p_c(X) \ge p_c(C_o(U))$ . Then a.s. the cone  $C_o(U)$  has an infinite component in  $\omega.$  Contradiction. 

**Lemma 2.5** Let X be a hyperbolic graph. Let  $\omega$  be a Bernoulli percolation on X with p < 1. Let  $K_i$  be a component of  $\omega$  such that  $|K_i| < \infty$ . Then a.s.  $\partial_{\infty} X = \overline{\bigcup_{i} \partial_{\infty} K_{i}}.$ 

Proof Suppose for contradiction that  $\partial_{\infty} X \setminus \overline{\bigcup_{i} \partial_{\infty} K_{i}} \neq \emptyset$ . Then  $\partial_{\infty} X \setminus \overline{\bigcup_{i} \partial_{\infty} K_{i}} = O$  for some open set  $O \subset \partial_{\infty} X$ . Let U be a proper

open subset of O such that  $\overline{U} \subset O$ . Let  $C_o(U)$  be the open cone consisting of geodesic rays  $[o, \theta]$  for all  $\theta \in U$ . Since the points in U are not limit points of the finite components of  $\omega$ , there must be some ball centered at o with radius R such that there are no finite components of  $\omega$  in  $C_o(U)$  past radius R. But there is a finite component N in the cone past radius R with probability  $p^{|N|} \cdot (1-p)^{|\partial N|}$ . Where  $\partial N$  is the boundary of N, i.e. the edges removed from X that disconnect the component. Since  $0 and both N and <math>\partial N$  are finite, this probability is positive. So a.s. there is a finite component in the cone  $C_o(U)$  past radius R. Contradiction.  $\square$ 

**Lemma 2.6** Let X be a hyperbolic graph. Let  $\omega$  be a p-Bernoulli percolation with  $p > p_u(X)$ . Suppose  $p_u(X) < 1 - p_c(X)$ . Then a.s.  $\partial_{\infty} K = \partial_{\infty} X$  where K is the infinite component of  $\omega$ .

*Proof* Suppose there is some point  $\xi \in \partial_{\infty} X$  such that  $\xi \notin \partial_{\infty} K$ . Then there is no sequence in K that limits to  $\xi$ . Then all sequences to  $\xi$  in X must have subsequences in  $\omega^c$ . At least some of these subsequences would be infinite. But  $\omega^c$  is a p'-Bernoulli percolation with  $p' < p_c(X)$ . So there are no infinite components in  $\omega^c$ . Contradiction. 

**Lemma 2.7** Let X be quasi-transitive, planar and hyperbolic. Let  $\omega$  be a p-Bernoulli percolation with  $p > p_c(X)$ . Then for two distinct infinite components  $K_i, K_j$  in  $\omega$ , the intersection of  $\partial_{\infty} K_i$  and  $\partial_{\infty} K_j$  must be finite.

*Proof* Suppose the boundary of two distinct components  $K_i, K_j$  have a nonempty intersection. Then there is a point  $\xi$  in  $K_i \cap K_j$  such that there are sequences  $\hat{a_n} \in K_i$  and  $\hat{b_n} \in K_j$  that both have  $\xi$  as a limit point. Let  $\eta$  be a point on either side of  $\xi$ . A path to  $\eta$  from  $K_j$  would have to cross  $\hat{a_n}$  or vice versa, which is impossible in a planar graph. Hence  $\eta \notin \partial_{\infty} K_i \cap \partial_{\infty} K_j$  unless  $\eta$  is on the edge of the boundary of one of the components. In a planar graph, the edges of the boundary must be finite. Hence  $\partial_{\infty} K_i \cap \partial_{\infty} K_j < \infty$ .

**Theorem 2.8** Let X be quasi-transitive, planar and hyperbolic. Let  $\omega$  be a p-Bernoulli percolation with  $p > p_c(X)$ . Then a.s. the boundary of the infinite components have empty interior.

Proof Suppose there is an open set U in the boundary at infinity of some infinite component  $K_0 \subseteq \omega$ . Consider the open cone  $C_o(U)$  consisting of geodesic rays  $[o, \theta] \in X$  for all  $\theta \in U$ . Since X is quasi-transitive,  $p_c(C_o(U)) = p_c(X)$ . Then a.s.  $\omega$  restricted to the cone has infinitely many infinite components. Since an infinite component of X has non-empty interior, an infinite component of  $C_o(U)$ a.s. has an infinite component with non-empty interior. By Lemma 2.7, The boundary of any infinite component  $K_i \in C_o(U), i \neq 0$  must have finite intersection with  $\partial_{\infty}K_0 = U$ . Then  $K_0$  is the only infinite component in  $C_o(U)$  that has non-empty interior. Since X is quasi-transitive, if one component has nonempty interior, then infinitely many components have non-empty interior. But  $C_o(U)$  only has one infinite component with non-empty interior. Contradiction.  $\Box$ 

For non-planar graphs, the boundary at infinite of infinite components can overlap much more. The following is a simple example with the potential for infinite overlap between components.

Example 2.9 Let T be a tree. Then let X be the graph consisting of two copies of T with added edges between a node and its copy. Formally,  $V(X) = (0, 1) \times T$  and  $[(i,j),(k,l)] \in E(X)$  if and only if  $\begin{cases} i < k, \text{ and } j = l \\ i = k \text{ and } [j,l] \in E(T) \end{cases}$ 

Then for each point  $\xi \in \partial_{\infty} X$ , there are infinitely many geodesic rays  $\gamma$  such that  $\gamma(\infty) = \xi$ .

Suppose  $\xi \in \partial_{\infty} K_i \cap \partial_{\infty} K_j$ . Then  $K_i$  and  $K_j$  have disjoint geodesics to  $\xi$ . The only way for this to happen is  $\alpha_1 \in K_i$ ,  $\alpha_2 \in K_j$  where  $\alpha_1$  is the only geodesic to  $\xi$  in  $\{0\} \times T$  and  $\alpha_2$  is the only geodesic to  $\xi$  in  $\{1\} \times T$ . Since  $K_i \neq K_j$ , there must be no edges between  $\alpha_1$  and  $\alpha_2$ . We will call this a disjoint ladder. Figure 1 shows an example of a disjoint ladder. Let B(o, n) be the ball of radius n. There are  $e^{hn}$  branches in B(o, n) where h is the Hausdorff dimension of  $\partial_{\infty}T$  and  $e^h$  is the branching number of T, br(T). The probability of a disjoint ladder is  $p^n(1-p)^n p^n$ . We will calculate a lower bound of the probability that the sides of the disjoint ladder are in distinct components. At some level n, there are  $\approx br(T) - 1$  branches, the probability that at least one of the edge or its copy is removed is  $2p(1-p) + (1-p)^2 = 1-p^2$ . Then the probability of there being a disjoint ladder in a ball of radius n in distinct components is greater than  $(p^2(1-p))^n(1-p^2)^{nbt(t)}$ .



Figure 1: The first 5 levels of a graph X with consisting of two copies of the binary tree. The red lines are an example of a disjoint ladder. At each level, our probability calculation required that at least one of the green edges be removed. Note that X would have disjoint ladders with probability 0, since the branching number of T is small.

In order for this ladder to occur at radius n, we need

$$1 \le [p^2(1-p)(1-p^2)^{br(T)}e^h]^n \Rightarrow 1 \le p^2(1-p)(1-p^2)^{br(T)}e^h \Rightarrow p^2(1-p)(1-p^2)^{br(T)} \ge \frac{1}{e^h}$$

Hence if the branching number is big enough and p is small enough, then infinitely many of the disjoint ladders will occur a.s. Note that this is feasible since for any tree T,  $p_c(T) = \frac{1}{br(T)}$ . A proof of this fact can be found in [4].

**Lemma 2.10** Let X a non-planar, hyperbolic graph. Let  $\omega$  be a p-Bernoulli percolation with  $p > p_c(X)$ . Then the boundary of the infinite components may have non-empty interior.

#### Example 2.11

Let  $G_i$  be a hyperbolic Cayley graph such that  $p_u(G_i) < 1$  and  $p_u(G_i) < 1 - p_c(G_i)$ . Let Z be the graph with  $V(Z) = \mathbb{Z}$  and  $E(Z) = \{uv : v = u + 1\}$ . Then Z is the line of integers. Let X be the graph that replaces each vertex i of Z with  $G_i$  and connect  $G_i$  and  $G_{i+1}$  only by their identity elements. Figure 2 shows a diagram of the graph X. Note that  $p_u(X) = 1$  since removing a



Figure 2: The graph of Example 2

finite number of edges would result in multiple components. By Lemma 2.6, the boundary at infinity of  $\omega$  restricted to some  $G_i$  is the entire boundary  $\partial_{\infty}G_i$ . Since  $\partial_{\infty}G_i$  is an open set there are infinitely many components of  $\omega$  with non-empty interior.

### 3 The Structure of Infinite Components

#### 3.1 Quasi-Geodesics

Given two metric spaces  $(M_1, d_1), (M_2, d_2)$ , a function f is a quasi-isometric embedding if there exists constants A > 0, B > 0 such that  $\forall x, y \in M_1$ :

$$\frac{1}{A}d_1(x,y) - B \le d_2(f(x), f(y)) \le Ad(x,y) + B$$
(1)

**Lemma 3.1** Let X be a hyperbolic, one ended, quasi-transitive graph. Let  $\omega$  be a p-Bernoulli percolation subgraph of X with  $p \in (0, 1)$ . Then a.s. each infinite component  $K_i \subset \omega$  is not quasi-isometrically embedded in X.

Proof For any t > 0, there is a positive probability that an entire ball of radius t will be removed from  $\omega$  and that a connected component in  $\omega$  will surround it. Let  $B(x_0,t)$  be the ball of radius t centered at  $x_o$ . There is some constant c such that the set  $[B(x_0,t+c) \setminus B(x_0,t)] \cap \omega$  is connected. Let  $CS(x_o,t) = [B(x_0,t+c) \setminus B(x_0,t)] \cap \omega$  be the connected shell where c is minimal and  $B(x_o,t) \cap \omega = \emptyset$ . Then  $CS(x_o,t+1)$  is in some component  $K_i$  of  $\omega$ .

If Aut(X) acts quasi-transitively on  $K_i$  then all geodesics of  $K_i$  are a finite uniform distance from geodesics of X. But there are arbitrarily large t-configurations  $CS(x_o, t)$ . By Lemma 1.1, geodesics between antipodal points of  $CS(x_o, t)$  must have length greater than  $e^t$  which in X they have length t. Since t can be arbitrarily big, there are no constants A, B which satisfy equation (1).  $\Box$ 

**Lemma 3.2** There are infinitely many configurations  $CS(\cdot, t)$  in a given component K of  $\omega$  for any t > 0.

*Proof* Define the function

$$F(x,y;\omega) = \begin{cases} \frac{1}{e^{ht}} : y \in K(x), y \in CS(x_o,t) \subset \omega, B(x_o,t) \subset \omega^c \\ 0 : \text{otherwise} \end{cases}$$

If there are a finite number of t-configurations in K then

$$\sum_{y \in X} F(x, y; \omega) < \infty.$$

If  $x \in S(x_o, t+1)$  with  $B(x_o, t) \subset \omega^c$ , then  $\sum_{y \in X} F(y, x; \omega) = \infty$ . So  $\sum_{y \in X} f(y, x) = \infty$ . By mass transport principle,  $\sum_{y \in X} f(x, y) = \infty$ . Hence, there are infinitely many *t*-configurations in a component K.

#### 3.2 Balls within an Infinite Component

**Lemma 3.3** Let  $\omega$  be a p-Bernoulli percolation subgraph of a hyperbolic graph X, with  $p_c(X) . Let <math>K_i \subset \omega$ . Then the ratio

$$q(x, y, r) = \frac{|B(x, R) \cap K_i|}{|B(y, R) \cap K_i|}$$

is exponentially large in R as  $d(x, y) \to \infty$ .

**Proof** Suppose that  $x, y \in X$  are in the same infinite component K of  $\omega$ . Given some radius R, there is a ball of radius  $R - \frac{d(x,y)}{2}$  contained inside the intersection of B(x, R) and B(y, R).

$$|B(x,R) \setminus B(y,R)| \ge e^{hR} - e^{h(R - \frac{d(x,y)}{2})} = e^{hR} (1 - e^{-h\frac{d(x,y)}{2}})$$

Then  $K \cap |B(x,R) \setminus B(y,R)|$  could be as big as  $e^{hR}(1-e^{-h\frac{d(x,y)}{2}})$ . In the case that K grows towards infinity in the  $\overrightarrow{xy}$  direction and not in the  $\overrightarrow{yx}$  direction, we could have  $K \cap |B(x,R) \setminus B(y,R)|$  close to  $e^{hR}(1-e^{-h\frac{d(x,y)}{2}})$  and  $K \cap |B(x,R) \setminus B(y,R)|$  close to 0. Then,  $q(x,y,R) \approx e^{hR}(1-e^{-h\frac{d(x,y)}{2}})$  and,

$$d(x,y) \to \infty \Rightarrow q(x,y,R) \to e^{hR}$$

### 4 Mass Transport Principle

The mass transport principle states that the expected total mass transported into any vertex  $x \in X$  is equal to the the expected total mass transported out of x, in an invariant percolation. The statement of mass transport principle can be simplified depending on the type of automorphism group of the graph X.

Let  $\Gamma$  be a group of automorphisms of a graph X. The *stabalizer* of  $x \in X$  is the set  $S(x) = \{\gamma \in \Gamma : \gamma x = x\}$ . Note that  $\gamma$  preserves the distance between x and y, so all  $\gamma y \in S(x)y$  are the same distance from x. Since X is locally finite and connected, it follows that the set  $S(x)y = \{\gamma y = \gamma \in S(x)\}$  is always finite.

**Theorem 4.1 (MTP)** Let  $\Gamma$  be a transitive group of automorphisms of X. Let  $F(x, y; \omega)$  be a function that is nonnegative and diagonally invariant under the action of  $\Gamma$ . That means for all  $\gamma \in \Gamma$ ,  $F(\gamma x, \gamma y; \gamma \omega) = F(x, y; \omega)$ . Let  $f(x, y) = E[F(x, y; \omega)]_{\omega}$ . Then,

$$\sum_{x \in X} f(o, x) = \sum_{x \in X} f(x, o) \frac{|S(x)y|}{|S(y)x|}$$

If |S(x)y| = |S(y)x| for all x, y such that  $y \in \Gamma x$ , then  $\Gamma$  is unimodular. Note that  $y \in \Gamma x$  means that there is some  $\gamma \in \Gamma$  such that  $x = \gamma y$ .

**Corollary 4.2** If  $\Gamma$  is both unimodular and transitive, then:

$$\sum_{x \in X} f(o, x) = \sum_{x \in X} f(x, o)$$

An example of the use of mass transport is in the proof of Lemma 4.3.

Given a point x, the diameter of K(x) is the maximum distance between two points on the boundary at infinity of K(x).

$$diam_x(\partial_{\infty}K(x)) = \max_{\xi,\eta\in\partial_{\infty}K(X)} d_x(\xi,\eta)$$

Recall that  $d_x(\xi,\eta) = e^{-(\xi\cdot\eta)_x}$  where  $(\xi\cdot\eta)_x$  is the Gromov product. Then clearly  $0 \le d_x(\xi,\eta) \le 1$  where  $d_x(\xi,\eta) = 0$  if  $\xi = \eta$  and  $d_x(\xi,\eta) = 1$  if x is on the geodesic between  $\xi$  and  $\eta$ .

**Lemma 4.3** For a unimodular, hyperbolic graph X and any point  $x \in K_i \subset X$ , there are infinitely many points for which the minimum diameter is achieved.

**Proof** Let X be a quasi-transitive, hyperbolic graph. For a vertex  $x \in X$ , let K(x) denote the component of  $\omega$  containing x. The minimal diameter of K(x) is the smallest diam<sub>x</sub>(K(x) for all  $x \in K(x)$ . Let N(y) be the number of points  $x \in K(y)$  where the diam<sub>x</sub>( $\partial_{\infty}K(y)$ ) is minimal. Suppose N(y) is finite. Define the function:

$$F(x,y;\omega) = \begin{cases} \frac{1}{N(x)} : x \in K(y), \operatorname{diam}_{x}(\partial_{\infty}K(y)) \text{ minimal} \\ 0 : \text{ otherwise} \end{cases}$$

Let  $f(x,y) = E_{\omega}[F(x,y;\omega)]$ . Then  $\sum_{x \in G} f(x,y) \leq 1$ . Also,  $\sum_{x \in X} f(y,x) = \infty$  since  $\sum_{x \in X} F(y,x;\omega) = \infty$  when  $\operatorname{diam}_{y}(\partial_{\infty}K(y))$  is minimal. By mass transport principle,

1 ~ ( ) 1

$$\sum_{x \in X} f(x, y) = \sum_{x \in X} f(y, x) \frac{|S(x)y|}{|S(y)x|}$$

Then the relative stabilizers must tend to infinity for all y. If X is unimodular, then  $\frac{|S(x)y|}{|S(y)x|} = 1$  and we have a contradiction. Hence, if X is unimodular, a.s. there are infinitely many points in  $K_i$  where the minimum diameter of  $\partial_{\infty}K_i$  is achieved.

We can also use mass transport to say something about the density of certain properties of a graph X within an infinite component  $K_i$  of the percolation subgraph. But first, we need the following lemma.

**Lemma 4.4** Let  $\hat{a}_k$  be a sequence such that  $\hat{a}_k$  converges to zero. Then there is some sequence  $\hat{b}_k$  such that  $\sum_{k=0}^{\infty} b_k = \infty$  and  $\sum_{k=0}^{\infty} a_k b_k < \infty$ .

*Proof* There is a sequence  $k_0, k_1, k_2, \dots$  such that for  $k \in [k_i, k_{i+1}], |a_k| < \frac{1}{2^i}$ . Choose  $b_k$  such that  $\sum_{k=k_i}^{k_{i+1}} b_k = 1$ . Then

$$\sum_{k=0}^{\infty} b_k = \sum_{k=0}^{k_0} b_k + \sum_{i=0}^{\infty} \sum_{k=k_i}^{k_j} b_k = C + \sum_{i=0}^{\infty} 1 = \infty,$$

$$\sum_{k=0}^{\infty} |a_k b_k| = \sum_{k=0}^{k_0} |a_k b_k| + \sum_{i=0}^{\infty} \sum_{k=k_i}^{k_{i+1}} |a_k b_k| < C' + \sum_{i=0}^{\infty} \frac{1}{2^i} < \infty$$

**Proposition 4.5 (Density Argument)** Let P be a diagonally-invariant property of vertices or edges of a graph quasi-transitive, unimodular and hyperbolic graph X. Let  $\omega$  be a p-Bernoulli percolation of X with  $p_c(X) and$  $<math>K_i$  be an infinite component of  $\omega$ . Then the asymptotic density of vertices with property P in a ball of finite radius r in  $K_i$  cannot approach 0 as r approaches infinity.

Let  $P_y = \{x \in K(y) : P \text{ holds for } x\}$ . We are defining the asymptotic density of vertices with property P in  $K_i$  as:

$$\lim_{r \to \infty} |P_y \cap B(y,r)| E_{x \in P_y \cap S(y,r)} \left[ \frac{1}{|B(x,r) \cap K(x)|} \right]$$

*Proof* Using mass transport principle, we can look at the ratio of vertices on some sphere  $S(x_o, r)$  with property P as r approaches infinity. Define  $a_r$  as:

$$a_r = |P_y \cap B(y,r)| E_{x \in P_y \cap B(y,r)} \left\lfloor \frac{1}{|B(x,r) \cap K(x)|} \right\rfloor$$

Then let  $\hat{b_r}$  be the sequence constructed in Lemma 4.4 such that  $\sum_{r=0}^{\infty} a_r b_r < \infty$ . Let  $l_x(r) = \frac{|B(x,r) \cap K(x)|}{b_r}$  in the  $d_K$  metric.

$$F(x,y;\omega) = \begin{cases} \frac{1}{l_x(d(x,y))} : x \in K(y), x \text{ satisfies } P\\ 0 : \text{ otherwise} \end{cases}$$

If x satisfies P, then we have:

$$\sum_{y \in X} F(x, y; \omega) = \sum_{y \in K(x)} \frac{1}{l_x(d(x, y))} = \sum_{r=0}^{\infty} \sum_{y \in B(x, r) \cap K(x)} \frac{1}{l_x(r)} = \sum_{r=0}^{\infty} \frac{b_r |B(x, r) \cap K(x)|}{|B(x, r) \cap K(x)|} = \sum_{r=0}^{\infty} b_r = \infty$$

So  $\sum_{y \in X} f(x, y) = \infty$ . Taking the sum over x, we have:

$$\sum_{x \in X} F(x, y; \omega) = \sum_{x \in K(y)} \frac{1}{l_x(d(x, y))}$$
$$= \sum_{r=0}^{\infty} \sum_{x \in P_y \cap B(y, r)} \frac{1}{l_x(r)}$$
$$= \sum_{r=0}^{\infty} b_r |P_y \cap B(y, r)| E_{x \in P_y \cap B(y, r)} \left[ \frac{1}{|B(x, r) \cap K(x)|} \right]$$
$$= \sum_{r=0}^{\infty} a_r b_r$$

Thus if the asymptotic density  $a_r$  approaches 0 then  $\sum_{x \in X} f(x, y)$  wound be finite. Hence the asymptotic density would approach a constant.

Define the property P such that x satisfies P if  $\operatorname{diam}_x(\partial_\infty K(x))$  is minimal. Then by the argument above, we can conclude that number of points that achieve minimum diameter in a ball of radius r has the same exponential order of growth as the ball inside K(x).

We speculate that Lemma 2.8 holds for all non-planar, quasi-transitive, hyperbolic graphs. If we assume otherwise, we can define a diagonally-invariant property using the radius of open balls at infinity. Given a point x, the let R(x) be the maximum radius of an open ball on the boundary at infinity of K(x) with respect to x.

$$R(x) = \max_{B(\xi,\epsilon)\subset \partial_\infty K(X)} \epsilon$$

If we can prove that the asymptotic density of the set  $\{x \in B(x,r) : R(x) \text{ is minimal}\}$  approaches 0 as r approaches infinity, then we can use the density argument to show a contradiction.

### References

- I. Benjamini and O. Schramm, Percolation in the hyperbolic plane, J. Amer. Math. Soc. 14 (2000), no. 2, 487-507.
- [2] M. Bridson and A. Haeiger, *Metric Spaces of Non-positive Curvature*, Springer, Berlin (1999).
- [3] I. Benjamini, R. Lyons, Y. Peres, and O. Schramm, Critical Percolation on Any Nonamenable Group Has No Infinite Clusters, The Annals of Probability, vol. 27, (1999),no. 3, 1347-1356.
- [4] Lyons, Russell, and Peres, Y, Probability on Trees and Networks. 2016.

[5] I. Benjamini, R. Lyons, and O. Schramm, Percolation perturbations in potential theory and random walks, in M. Picardello and W. Woess, editors, Random Walks and Discrete Potential Theory, Sympos. Math., Cambridge University Press, Cambridge, 1999, pp. 5684. Papers from the workshop held in Cortona, 1997.

# Spectral properties of reversible indel models

Linden Yuan Mentor: Wai-Tong (Louis) Fan

# July 30, 2019

#### Abstract

W.-T. Fan and S. Roch [FR17] considered the ancestral state reconstruction problem on a sequence of trees with uniformly bounded height under the Thorne–Kishino–Felsenstein 1991 model [TKF91, TKF92] of DNA sequence evolution. They provided explicit and consistent root state estimators that achieve the optimal rate of convergence. In this paper, we study the rate of convergence of a reversible indel chain to its equilibrium distribution.

# Contents

1	Intr	oduction	2	
	1.1	New results	3	
	1.2	Binary indel chain	3	
	1.3	Total variation distance	5	
	1.4	$L^2(\mu)$ convergence vs. $d_{TV}$ convergence	5	
2	Sequence length chain			
	2.1	Length chain: Generator and Dirichlet form	7	
	2.2	Not the Poincaré inequality	7	
	2.3	Poincaré inequality	8	
	2.4	Explicit formula for transition density: The Karlin–McGregor theorem	8	
3	Cou	pling	9	
	3.1	Coupling of the length chain	9	
	3.2	Indel coupling	11	
4	Fur	ther directions	15	
	4.1	Improvement of $d_{TV}$ coupling bound	15	
	4.2	Poincaré inequality: canonical paths	16	
	4.3	Spectral decomposition	17	
	4.4	Other models	17	

5	Acknowledgements	18
A	List of theorems	18
Re	ferences	20

# **1** Introduction

In computational biology and in phylogenetics, the problems of sequence alignment and phylogeny reconstruction have become important. A sequence alignment is an array of nucleotides, A, C, G, T and blank characters (which we may take to be dashes, for instance). From a collection of unaligned sequences, researchers construct an alignment and try to infer the phylogeny (see Figure 1).



Figure 1: Top: phylogenetic tree. Middle: sequence alignment. Bottom: collection of unaligned sequences. Source: [Fle04].

The TKF91 indel (insertion-deletion) model models DNA sequence evolution via a continuoustime Markov chain. The state space has not been well-studied, and this fact poses a difficulty in the analysis of the chain. The TKF91 indel model has been criticized for being unrealistic [MT01] and for being computationally unwieldy [BCJ13]. The TKF91 indel model is more realistic, but it is still unrealistic [MT01]. The Poisson indel model [BCJ13] is a variant that is computationally wieldy. In the context of *d*-ary trees, the Kesten–Stigum threshold [Mos01, Mos04, KS66, Roc12] for reconstruction relates the rate of convergence, via the spectral gap, of a certain 2-state Markov chain to the problem of ancestral reconstruction. If the mixing rate is above a certain threshold, then reconstruction is always possible; if the mixing rate is below the threshold, then reconstruction is impossible (success can only happen with probability at most 1/2). Research is still being conducted on similar results in more general scenarios. This result and this area of research suggest deep connection between the spectral gap of certain Markov chains and ancestral reconstruction.

In Subsection 1.1, we summarize our results. In Subsection 1.2, we introduce a simplified version of the TKF91 model. In Subsections 1.3 and 1.4, we provide a refresher of some results on certain types of convergence.

### **1.1** New results

We prove exponential ergodicity in total variation distance for the TKF91 indel chain, and furthermore we show the rate is at least as good as  $\mu - \lambda + \varepsilon$  for all  $\varepsilon > 0$ . The challenges we overcame were finding a good coupling and by bounding the coupling time. We addressed these two challenges, respectively, by drawing inspiration from previous analysis of random walk on a hypercube, and by separating the coupling time  $\tau_{\text{couple}}$  into two easily analyzable random variables,  $\tau_1$  and  $\tau_2$  (see Lemma 8).

### **1.2 Binary indel chain**

We consider the following model of binary sequence evolution with insertion and deletion, which is a simplified version of the TKF91 model. C. Daskalakis and S. Roch introduced a related model that gets rid of the reversibility condition [DR10].

**Definition 1** (binary indel chain). *The* **binary indel chain** *is a continuous-time Markov chain*  $\mathcal{I} = (\mathcal{I}_t)_{t>0}$  *on the space* 

$$\mathcal{S} := \bigcup_{M \ge 0} (\{\bullet\} \times \{0, 1\}^M) \tag{1}$$

of binary sequences appended to an **immortal link** "•". We also refer to the positions of a sequence (including digit and the immortal link) as sites. Let  $(\nu, \lambda, \mu) \in (0, \infty)^3$  with  $\lambda < \mu$  be given parameters. In addition, let  $\pi_0 = 1/2, \pi_1 = 1/2$ . The continuous-time Markovian dynamic is described as follows: if the current state is the sequence  $\vec{x}$ , then the following events occur independently:

- (Substitution) Each digit (but not the immortal link) is substituted independently at rate  $\nu > 0$ . When a substitution occurs, the corresponding digit is replaced by 0 and 1 with probabilities  $\pi_0$  and  $\pi_1$  respectively.
- (Deletion) Each digit (but not the immortal link) is removed independently at rate  $\mu > 0$ .
- (Insertion) Each site gives birth to a new digit independently at rate  $\lambda > 0$ . When a birth occurs, a digit is added immediately to the right of its parent site. The newborn site has digit 0 and 1 with probabilities  $\pi_0$  and  $\pi_1$  respectively.
The length of a sequence  $\vec{x} = (\bullet, x_1, x_2, \cdots, x_M)$  is defined as the number of digits in  $\vec{x}$  and is denoted by  $|\vec{x}| = M$  (with the immortal link alone corresponding to M = 0). When  $M \ge 1$  we make the identification  $\vec{x} = (\bullet, x_1, x_2, \cdots, x_M) = (x_1, x_2, \cdots, x_M)$ .

The binary indel chain is reversible. Further, let  $r = \lambda/\mu$  (ratio), and suppose that

$$0 < \lambda < \mu,$$

an assumption we make throughout. Then the chain has a stationary distribution  $\Pi$ , given by

$$\Pi(\vec{x}) = (1-r) \left(\frac{r}{2}\right)^{|\vec{x}|}.$$
(2)

Under  $\Pi$ , the sequence length is geometrically distributed and, conditioned on the sequence length, all sites are independent with distribution  $\pi_{sim} := (\pi_0, \pi_1)$ .

Throughout this paper, we let  $\mathbb{P}_{\vec{x}}$  be the probability measure when the root state is  $\vec{x}$ . If the root state is chosen according to a distribution  $\Pi$ , then we denote the probability measure by  $\mathbb{P}_{\Pi}$ . From the previous paragraph, we have for all  $\vec{a}, \vec{z} \in S$  and all  $t \ge 0$ ,

$$\mathbb{P}_{\Pi}(\mathcal{I}_t = \vec{z}) = \Pi(\vec{z}) \quad \text{and} \tag{3}$$

$$\frac{\mathbb{P}_{\vec{a}}(\mathcal{I}_t = \vec{z})}{\Pi(\vec{z})} = \frac{\mathbb{P}_{\vec{z}}(\mathcal{I}_t = \vec{a})}{\Pi(\vec{a})}.$$
(4)

We also denote by  $\mathbb{P}_M$  the conditional probability measure, under  $\mathbb{P}_{\Pi}$ , for the event that the root state has length  $M \ge 1$ . The next lemma says that conditioned on having length  $\ge N$ , the distribution of the first N digits of  $\mathcal{I}_t$  under  $\mathbb{P}_M$  are independent with the same distribution  $\pi_{sim}$ .

Remark 2. (Infinitesimal generator and Dirichlet form)

For all  $\vec{x}, \vec{y}$ , we write  $\lambda(\vec{x}, \vec{y})$  for the transition rate from  $\vec{x}$  to  $\vec{y}$ . The **infinitesimal generator** of the indel chain  $\mathcal{I} = (\mathcal{I}_t)_{t \ge 0}$  is given as follows: for  $\vec{x} = (\bullet, x_1, \cdots, x_M) \in \{0, 1\}^M$  where  $M \ge 0$ ,

$$\mathcal{A}f(\vec{x}) = \sum_{\vec{y}\in\mathcal{S}} \lambda(\vec{x},\vec{y})f(\vec{y})$$

$$= \frac{\lambda}{2} \sum_{j=0}^{M} \sum_{z\in\{0,1\}} f(\vec{x}_{0}^{+z}) + \frac{\nu}{2} \sum_{j\in[M]} f(\vec{x}_{j}) + \mu \sum_{j\in[M]} f(\vec{x}_{j}^{-})$$

$$- f(\vec{x}) \left(\nu \sum_{j\in[M]} \pi_{1-x_{j}} + M\mu + (M+1)\lambda\right),$$
(5)

where

$$\vec{x}_j := (x_1, \cdots, x_{j-1}, 1 - x_j, x_{j+1}, \cdots, x_M) \in \{0, 1\}^M$$
$$\vec{x}_j^- := (x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_M) \in \{0, 1\}^{M-1}$$
$$\vec{x}_j^{+z} := (x_1, \cdots, x_j, z, x_{j+1}, \cdots, x_M) \in \{0, 1\}^{M+1}$$

are obtained respectively by substitution, deletion and insertion of  $z \in \{0, 1\}$  for site  $x_j$  where  $j \ge 1$  (for the case j = 0, we define formally  $\vec{x}_j^{+z} = (z, x_1, \dots, x_M)$ ). The **Dirichlet form** of  $\mathcal{I} = (\mathcal{I}_t)_{t \ge 0}$  on  $L^2(\Pi)$  is symmetric and is given by

$$\mathcal{E}(f,g) = -\langle \mathcal{A}f,g \rangle_{\Pi} = -\sum_{\vec{x} \in \mathcal{S}} \sum_{\vec{y} \in \mathcal{S}} f(\vec{y})g(\vec{x})\lambda(\vec{x},\vec{y})\Pi(\vec{x})$$
(6)

$$= \frac{1}{2} \sum_{\vec{x} \in \mathcal{S}} \sum_{\vec{y} \in \mathcal{S}} (f(\vec{y}) - f(\vec{x})) (g(\vec{y}) - g(\vec{x})) \lambda(\vec{x}, \vec{y}) \Pi(\vec{x}).$$
(7)

Thus we have

$$\mathcal{E}(f,f) = \frac{1}{2} \sum_{\vec{x}\in\mathcal{S}} \sum_{\vec{y}\in\mathcal{S}} (f(\vec{y}) - f(\vec{x}))^2 \lambda(\vec{x},\vec{y}) \Pi(\vec{x}).$$
(8)

#### **1.3** Total variation distance

We use the notation  $a \wedge b = \min\{a, b\}$  and  $a \vee b = \max\{a, b\}$ . For two probability measures  $\mu_1$ ,  $\mu_2$  on a countable measure space S, let

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \sum_{\sigma \in \mathcal{S}} |\mu_1(\sigma) - \mu_2(\sigma)| = \sup_{\mathcal{A} \subseteq \mathcal{S}} |\mu_1(\mathcal{A}) - \mu_2(\mathcal{A})|$$
$$= 1 - \sum_{\sigma \in \mathcal{S}} \mu_1(\sigma) \wedge \mu_2(\sigma)$$
(9)

be the total variation distance between  $\mu_1$  and  $\mu_2$ . The last equality follows from noticing that

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \sum_{\sigma \in \mathcal{S}} [\mu_1(\sigma) \lor \mu_2(\sigma) - \mu_1(\sigma) \land \mu_2(\sigma)] \text{ and}$$
$$1 = \frac{1}{2} \sum_{\sigma \in \mathcal{S}} [\mu_1(\sigma) \lor \mu_2(\sigma) + \mu_1(\sigma) \land \mu_2(\sigma)].$$

# **1.4** $L^2(\mu)$ convergence vs. $d_{TV}$ convergence

In this subsection we show  $L^2(\mu)$  ergodicity is better than  $d_{TV}$  ergodicity. This result can be applied to the indel chain, with  $\mu = \Pi$ , and to the length process (see Section 2), with  $\mu = \vec{\gamma}$ . The Poincaré inequality for a reversible Markov chain with Dirichlet form  $\mathcal{E}$  is

$$\mathcal{E}(f,f) \ge \frac{1}{c} \operatorname{Var}_{\mu}(f), \tag{10}$$

where  $\mu$  is the stationary measure,

$$\operatorname{Var}_{\mu}(f) = \int f^2 d\mu - \left(\int f d\mu\right)^2 = \|f - \int f d\mu\|_{\mu}^2$$

is the variance of f and  $c \in (0, \infty)$  is a constant independent of f and  $\|\cdot\|_{\mu} = \|\cdot\|_{L^2(\mu)}$  is the  $L^2$  norm with respect to measure  $\mu$ . It is well known (Theorem 11) that for reversible ergodic Markov chains, (10) is equivalent to the  $L^2$  exponential ergodicity

$$\operatorname{Var}_{\mu}(P_t f) \le e^{-2t/c} \operatorname{Var}_{\mu}(f), \tag{11}$$

or equivalently

$$\|P_t f - \int f \, d\mu\|_{\mu} \le e^{-t/c} \, \|f - \int f \, d\mu\|_{\mu}.$$
(12)

It can be checked also that the  $L^2$  exponential ergodicity (12) implies exponential ergodicity *in* total variation distance for all probability measures f (i.e.  $\sum_x f(x) = 1$ ):

$$d_{TV}(P_t f, \mu) \le e^{-t/c} \cdot \frac{1}{2} \left\| f/\mu - 1 \right\|_{L^2(\mu)}.$$
(13)

*Proof for* (12) *implies* (13). Denote by  $p_t(x, y) := \mathbb{P}_x(X_t = y)$  the transition probability. Then we have symmetry  $\mu_x p_t(x, y) = \mu_y p_t(y, x)$  and

$$\begin{aligned} d_{TV}(P_t f, \mu) &= \frac{1}{2} \sum_x |P_t f(x) - \mu_x| \\ &= \frac{1}{2} \sum_x \left| \sum_y (f(y) - \mu_x) p_t(x, y) \right| \\ &= \frac{1}{2} \sum_x \mu_x \left| \sum_y \left( \frac{f(y)}{\mu_x} - 1 \right) p_t(x, y) \right| \\ &= \frac{1}{2} \left\| \sum_y \left( \frac{f(y)}{\mu} - 1 \right) p_t(\cdot, y) \right\|_{L^1(\mu)} \\ &\leq \frac{1}{2} \left\| \sum_y \left( \frac{f(y)}{\mu} - 1 \right) p_t(\cdot, y) \right\|_{L^2(\mu)} \\ &\leq \frac{1}{2} \left\| P_t(f/\mu - 1) \right\|_{L^2(\mu)} \\ &\leq \frac{1}{2} e^{-t/c} \left\| f/\mu - 1 \right\|_{L^2(\mu)}. \end{aligned}$$

The assumption  $\sum_{x} f(x) = 1$  guarantees that inequality (12) is applicable in the last inequality.

Fixing x and taking  $f = 1_x$ , we obtain

$$||p_t(\cdot, x) - \mu||_{TV} \le e^{-t/c} ||1_x/\mu - 1||_{L^2(\mu)}$$

# 2 Sequence length chain

For each  $t \ge 0$ , we write  $L_t = |\mathcal{I}_t|$  to mean the sequence length at time t, and we write  $p_t(i, j) = P_i(L_t = j)$  for the transition function, for all  $i, j \ge 0$ .

It is clear that the sequence length  $L := (L_t)_{t \ge 0}$  evolves as a birth-death process with equilibrium distribution  $\vec{\gamma} := (\gamma_M)_{M \ge 0}$ , where

$$\forall M \ge 0 \quad \gamma_M = (1 - r)r^M. \tag{14}$$

Convergence rate towards equilibrium for L was studied by A. Mitrophanov and M. Borodovsky [MB07].

# 2.1 Length chain: Generator and Dirichlet form

The generator of the length chain is given by

$$Q_{i,i+1} = (1+i)\lambda \ (i \ge 0), \quad Q_{i,i-1} = i\mu \ (i \ge 1), \quad Q_{i,j} = 0 \ (|j-i| \ge 2).$$

The Dirichlet form of the length chain is

$$\mathcal{E}(f,g) = -\sum_{i=0}^{\infty} \left(\sum_{j=0}^{\infty} Q_{i,j}f(j)\right) g(i) \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^{i}$$
$$= \frac{\mu - \lambda}{2\mu} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{i,j}(f(j) - f(i))(g(j) - g(i)) \left(\frac{\lambda}{\mu}\right)^{i}$$
$$= \frac{1}{2}(1-r) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{i,j}(f(j) - f(i))(g(j) - g(i))r^{i}.$$

Note we have

$$\mathcal{E}(f,f) = \frac{1}{2}(1-r)\left(\sum_{i\geq 0}(1+i)\lambda(f_{i+1}-f_i)^2r^i + \sum_{i\geq 1}i\mu(f_i-f_{i-1})^2r^i\right).$$

We have

$$\sum_{i\geq 1} i\mu(f_i - f_{i-1})^2 r^i = \sum_{i\geq 0} (i+1)\mu(f_{i+1} - f_i)^2 r^{i+1}$$

and

$$\forall i \ge 0 \quad (i+1)\mu(f_{i+1} - f_i)^2 r^{i+1} = (i+1)\lambda(f_{i+1} - f_i)^2 r^i,$$

so

$$\mathcal{E}(f,f) = \frac{1}{2}(1-r)\left(\sum_{i\geq 0}(1+i)\lambda(f_{i+1}-f_i)^2r^i + \sum_{i\geq 1}i\mu(f_i-f_{i-1})^2r^i\right)$$
  
=  $(1-r)\sum_{i\geq 0}(1+i)\lambda(f_{i+1}-f_i)^2r^i = (1-r)\lambda\sum_{i\geq 0}(1+i)r^i(f_{i+1}-f_i)^2.$  (15)

# 2.2 Not the Poincaré inequality

By a previous result (Theorem 10), we have

$$\forall M \ge 0 \quad \sum_{n=0}^{\infty} n |p_t(M, n) - \gamma_n| \le \left(M + \frac{\lambda}{\mu - \lambda}\right) e^{-(\mu - \lambda)t}.$$

The triangle inequality gives the following result.

### **Proposition 3.**

$$\forall M \ge 0 \quad d_{TV}(p_t(M, \cdot), \vec{\gamma}) \le \left(M + \frac{\lambda}{\mu - \lambda}\right) e^{-(\mu - \lambda)t}.$$

*Proof.* We have

$$\sum_{n=1}^{\infty} |p_t(M,n) - \gamma_n| \le \sum_{n=1}^{\infty} n |p_t(M,n) - \gamma_n| = \sum_{n=0}^{\infty} n |p_t(M,n) - \gamma_n|$$

and

$$|p_t(M,0) - \gamma_0| = \left| \left( 1 - \sum_{n=1}^{\infty} p_t(M,n) \right) - \left( 1 - \sum_{n=1}^{\infty} \gamma_n \right) \right| = \left| \sum_{n=1}^{\infty} p_t(M,n) - \sum_{n=1}^{\infty} \gamma_n \right|$$
$$= \left| \sum_{n=1}^{\infty} (p_t(M,n) - \gamma_n) \right| \le \sum_{n=1}^{\infty} |p_t(M,n) - \gamma_n|,$$
$$d_{TV}(p_t(M,\cdot),\vec{\gamma}) = \frac{1}{2} \sum_{n=0}^{\infty} |p_t(M,n) - \gamma_n| \le \left( M + \frac{\lambda}{\mu - \lambda} \right) e^{-(\mu - \lambda)t}.$$

so

### 2.3 Poincaré inequality

For the length process, we that the Poincaré inequality holds, and further that the best constant in the Poincaré inequality is

$$\inf_{f \in L^2(\vec{\gamma}): \vec{\gamma} f = 0, \operatorname{Var}_{\vec{\gamma}} f = 1} \mathcal{E}(f, f) =: \operatorname{gap}(\mathcal{E}) = \mu - \lambda$$

[Che04, Section 9.3, second example under heading "Examples 9.27"].

### 2.4 Explicit formula for transition density: The Karlin–McGregor theorem

Karlin and J. McGregor proved [KM58] that the transition function  $p_t(i, j)$  for the birth-death process L can be represented as

$$p_t(i,j) = r^j \int_0^\infty e^{-xt} Q_i(x) Q_j(x) d\phi(x), \quad i,j \in \mathbb{Z}_+, \ t \ge 0,$$
(16)

where  $\phi$  is the spectral measure of the transition matrix and  $\{Q_n\}_{n \in \mathbb{Z}_+}$  are orthogonal polynomials. More explicit expressions are given below according to different cases:

Case 1 ( $\lambda < \mu$ ). There is a stationary distribution given by  $\gamma_n = (1 - r)r^n$ . Moreover, in (16), the orthogonal polynomials are given by Meixner polynomials

$$Q_n(x) = M_n\left(\frac{x}{\mu - \lambda}; 1, r\right), \quad n \in \mathbb{Z}_+,$$

and  $\phi$  is the probability distribution assigning mass  $w_n := \gamma_n$  at the points  $(\mu - \lambda)n$ .

Case 2 ( $\lambda = \mu$ ). In (16), the orthogonal polynomials are given by Laguerre polynomials

$$Q_n(x) = L_n^{(0)}\left(\frac{x}{\lambda}\right), \quad n \in \mathbb{Z}_+,$$

and  $\phi$  is the probability density function of the Gamma distribution  $\Gamma(1, \lambda)$  (exponential with intensity  $1/\lambda$ ), that is,

$$d\phi(x) = \frac{e^{-x/\lambda}}{\lambda} dx.$$

The case  $\lambda > \mu$  is not needed here so we do not type it out. Interested readers can consult Chapter 3 of Schoutens's book [Sch00, Ch. 3, *Birth and Death Processes, Random Walks, and Orthogonal Polynomials*] and take  $\beta = 1$  there. Anderson's continuous-time Markov chain book is also a good reference on the Karlin–McGregor theorem (see [And91, Ch. 8, *Birth and Death Processes*]).

By (16), we have the explicit formula

$$p_{t}(i,j) = r^{j} \sum_{n \ge 0} e^{-(\mu - \lambda)nt} M_{i}(n;1,r) M_{j}(n;1,r) \gamma_{n}$$
  
$$= (1-r)r^{j} \sum_{n \ge 0} e^{-(\mu - \lambda)nt} M_{i,n} M_{j,n} r^{n}$$
  
$$= (1-r)r^{j} \sum_{n \ge 0} [re^{-(\mu - \lambda)t}]^{n} M_{i,n} M_{j,n}, \qquad (17)$$

where  $M_{i,n}$  are Meixner polynomials defined by

$$M_{i}(n;1,r) = \sum_{k=0}^{i} (-1)^{k} {i \choose k} {n \choose k} k! (n+1)_{i-k} r^{-k},$$

where the Pochhammer symbol denotes the rising factorial.

See the Appendix of W.-T. Fan and S. Roch's 2017 preprint [FR17] for more properties of the indel chain  $\mathcal{I}$  and the sequence length chain L.

# **3** Coupling

### **3.1** Coupling of the length chain

In this subsection, we establish the following result.

**Proposition 4.** Say  $M \in \{0\} \cup \mathbb{N}$ , and say  $t \in [0, \infty)$ . Then

$$d_{TV}(p_t(M,\cdot),\vec{\gamma}) \le e^{-(\mu-\lambda)t} \left(M + \frac{2r^{M+1}-r}{1-r}\right).$$

Remark 5. Note

$$M + \frac{2r^{M+1} - r}{1 - r} \le M + \frac{r}{1 - r}$$

for all  $M \ge 0$ , with strict inequality for all  $M \ge 1$ . Thus this bound is better than the bound we obtained in Subsection 2.2 in Proposition 3.

Proof of Proposition 4. Say  $M \in \{0\} \cup \mathbb{N}$ , and say  $t \in [0, \infty)$ . Define the following coupling of  $p_t(M, \cdot)$  and  $\vec{\gamma} : (X_t)_{t \in [0,\infty)}, (Y_t)_{t \in [0,\infty)}$  are birth-death processes with rate parameters  $\lambda_n = (n+1)\lambda$   $(n \ge 0), \mu_n = n\mu$   $(n \ge 1)$  with  $X_0 \sim \delta_M$  and  $Y_0 \sim \vec{\gamma}$  that run independently until they meet, from which point they stay together. By the coupling theorem (Theorem 12), we see

$$d_{TV}(p_t(M, \cdot), \vec{\gamma}) \le \mathbb{P}(X_t \neq Y_t) = \sum_{i=0}^{\infty} \gamma_i \mathbb{P}(X_t \neq Y_t \mid Y_0 = i) \le \sum_{i=0}^{\infty} \gamma_i \mathbb{E}|X_t - Y_t|$$
$$= \sum_{i=0}^{\infty} \gamma_i |i - M| e^{-(\mu - \lambda)t}.$$

We see

$$\sum_{i=0}^{\infty} \gamma_i |i - M| e^{-(\mu - \lambda)t} = (1 - r) \sum_{i=0}^{\infty} r^i |i - M| e^{-(\mu - \lambda)t} = e^{-(\mu - \lambda)t} (1 - r) \sum_{i=0}^{\infty} r^i |i - M|.$$

We have

$$\sum_{i=0}^{\infty} r^{i} |i - M| = \sum_{i=0}^{M} r^{i} (M - i) + \sum_{i=M+1}^{\infty} r^{i} (i - M).$$

If  $M \geq 1$ , we have

$$\sum_{i=0}^{M} r^{i}(M-i) = 1r^{M-1} + \dots + Mr^{0} = (r^{M-1} + \dots + r^{0}) + \dots + r^{0} = \frac{1-r^{M}}{1-r} + \dots + \frac{1-r}{1-r}$$
$$= \frac{1}{1-r}((1-r) + \dots + (1-r^{M})) = \frac{1}{1-r}(M - (r + \dots + r^{M})).$$
We have  $r + \dots + r^{M} = r(1-r^{M})/(1-r)$ , so

$$\sum_{i=0}^{M} r^{i}(M-i) = \frac{1}{1-r} \left( M - \frac{r(1-r^{M})}{1-r} \right).$$

In addition, we have

$$\sum_{i=M+1}^{\infty} r^{i}(i-M) = r^{M} \sum_{i=M+1}^{\infty} r^{i-M}(i-M)e^{-(\mu-\lambda)t} = r^{M} \sum_{i=1}^{\infty} r^{i}i.$$

We have

$$\sum_{i=1}^{\infty} r^{i}i = r + 2r^{2} + 3r^{3} + \dots = (r + r^{2} + r^{3} + \dots) + (r^{2} + r^{3} + \dots) + (r^{3} + \dots) + \dots$$
$$= \frac{r}{1 - r} + \frac{r^{2}}{1 - r} + \frac{r^{3}}{1 - r} + \dots = \frac{r/(1 - r)}{1 - r} = \frac{r}{(1 - r)^{2}}.$$

Thus

$$\sum_{i=M+1}^{\infty} r^{i}(i-M) = \frac{r^{M+1}}{(1-r)^{2}}.$$

Thus

$$e^{-(\mu-\lambda)t}(1-r)\sum_{i=0}^{\infty}r^{i}|i-M| = e^{-(\mu-\lambda)t}\left(M + \frac{2r^{M+1}-r}{1-r}\right).$$

## 3.2 Indel coupling

In this subsection, we establish the following result.

**Theorem 6.** Say  $\vec{x} \in S$ , and say  $t \in [0, \infty)$ . Let  $M = |\vec{x}|$ . Then

$$d_{TV}(p_t(\vec{x},\cdot),\Pi) \le e^{-(\mu-\lambda)t} \left(M + \frac{2r^{M+1} - r}{1-r}\right) \left((\mu - \lambda) \left(\frac{\mu + \nu}{\lambda + \nu} \cdot t - \frac{\mu - \lambda}{(\lambda + \nu)^2}\right) + \frac{\mu + \nu}{\lambda + \nu} \cdot M + 1\right)$$

Before we prove this result, we record the following corollary, which summarizes our result more succinctly.

**Corollary 7.** Say  $\vec{x} \in S$ . For all  $\varepsilon > 0$ , there is a finite constant C such that

$$\forall t \ge 0 \quad d_{TV}(p_t(\vec{x}, \cdot), \Pi) \le e^{(\varepsilon - (\mu - \lambda))t} C$$

(and moreover, C can be chosen to depend only on  $|\vec{x}|$ ).

Proof of Theorem 6. Let  $X = (X_t)_{t \in [0,\infty)}$ ,  $Y = (Y_t)_{t \in [0,\infty)}$  be two indel chains coupled as follows:  $X_0 \sim \delta_{\vec{x}}$  for some  $\vec{x}$ ,  $Y_0 \sim \Pi$ ; X and Y run independently until they have the same length; once X, Y have the same length, X, Y perform insertion and deletion together (putting in the same new element, for insertion), and in matched coordinates they substitute together (putting in the same new element) but in unmatched coordinates they substitute independently (independent times and independent choices of new elements); and once the processes meet, they stay together. Let  $M = |\vec{x}|$ .

Let  $\tau_{\text{couple}} = \inf\{t \in [0, \infty) : X_t = Y_t\}$  denote the coupling time. Let  $\tau_1 = \inf\{t \in [0, \infty) : |X_t| = |Y_t|\}$  denote the time it takes for the lengths to couple, and let  $\tau_2 = \tau_{\text{couple}} - \tau_1$ . Then  $\tau_{\text{couple}} = \tau_1 + \tau_2$ , so we have  $\mathbb{P}(\tau_{\text{couple}} \ge t) = \mathbb{P}(\tau_1 + \tau_2 \ge t)$ .

We have

$$\mathbb{P}(\tau_1 + \tau_2 \ge t) = \int_{[0,\infty)} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) \mathbb{P}(\tau_1 \in ds)$$
$$= \int_{[0,t)} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) \mathbb{P}(\tau_1 \in ds) + \int_{[t,\infty)} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) \mathbb{P}(\tau_1 \in ds).$$

We can bound the second term by

$$\int_{[t,\infty)} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) \mathbb{P}(\tau_1 \in ds) = \int_{[t,\infty)} \mathbb{P}(\tau_1 \in ds) = \mathbb{P}(\tau_1 \ge t)$$
$$\le e^{-(\mu - \lambda)t} \left( M + \frac{2r^{M+1} - r}{1 - r} \right),$$

using the inequality from Subsection 3.1.

We record the following result, which we will prove at the end of this subsection.

**Lemma 8.** Say  $\vec{y} \in \{\bullet\} \times \{0,1\}^M$ , and let  $T = \tau_{\text{couple}}$ . Let  $d = |\{|i \in [M] : x_i \neq y_i\}$  denote the number of spots at which  $\vec{x}$  and  $\vec{y}$  differ (alternatively, the Hamming distance of  $\vec{x}$  and  $\vec{y}$ ). Then  $T = \tau_2$  and

$$\mathbb{P}(T \ge t \mid Y_0 = \vec{y}) = 1 - (1 - e^{-(\mu + \nu)t})^d \le 1 - (1 - e^{-(\mu + \nu)t})^M \le M e^{-(\mu + \nu)t}.$$

We will now derive an upper bound for  $\mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s)$  via Lemma 8.

Let  $\Lambda$  denote the common length of X and Y when they obtain the same length (so  $\Lambda = |X_{\tau_1}| = |Y_{\tau_1}|$ ). For each  $s \in [0, t)$ , we have

$$\mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) = \sum_{j=0}^{\infty} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s, \Lambda = j) \mathbb{P}(\Lambda = j \mid \tau_1 = s)$$

$$\leq \text{(Lemma 8 \& Markov property)} \sum_{j=0} j e^{-(\mu+\nu)(t-s)} \mathbb{P}(\Lambda = j \mid \tau_1 = s)$$
(18)

$$= e^{-(\mu+\nu)(t-s)} \sum_{j=0}^{\infty} j \mathbb{P}(\Lambda = j \mid \tau_1 = s) = e^{-(\mu+\nu)(t-s)} \mathbb{E}(\Lambda \mid \tau_1 = s) = e^{-(\mu+\nu)(t-s)} (M + (\mu-\lambda)s).$$

Thus we have

$$\int_{[0,t)} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) \mathbb{P}(\tau_1 \in ds) \le \int_{[0,t)} e^{-(\mu + \nu)(t - s)} (M + (\mu - \lambda)s) \mathbb{P}(\tau_1 \in ds)$$
$$= e^{-(\mu + \nu)t} \int_{[0,t)} e^{(\mu + \nu)s} (M + (\mu - \lambda)s) \mathbb{P}(\tau_1 \in ds).$$

We have

$$\int_{[0,t)} e^{(\mu+\nu)s} (M+(\mu-\lambda)s) \mathbb{P}(\tau_1 \in ds)$$
$$= M \int_{[0,t)} e^{(\mu+\nu)s} \mathbb{P}(\tau_1 \in ds) + (\mu-\lambda) \int_{[0,t)} s e^{(\mu+\nu)s} \mathbb{P}(\tau_1 \in ds).$$

We have

$$\begin{split} \int_{[0,t)} e^{(\mu+\nu)s} \mathbb{P}(\tau_1 \in ds) &= \mathbb{E}(e^{(\mu+\nu)\tau_1} \mathbf{1}_{\tau_1^{-1}([0,t])}) \stackrel{*}{=} \int_{-\infty}^t (\mu+\nu)e^{(\mu+\nu)s} \mathbb{P}(s \le \tau_1 \le t) ds \\ &\leq \int_{-\infty}^t (\mu+\nu)e^{(\mu+\nu)s} e^{-(\mu-\lambda)s} \left(M + \frac{2r^{M+1} - r}{1-r}\right) ds \\ &= \left(M + \frac{2r^{M+1} - r}{1-r}\right)(\mu+\nu) \int_{-\infty}^t e^{(\lambda+\nu)s} ds = \frac{\mu+\nu}{\lambda+\nu} \left(M + \frac{2r^{M+1} - r}{1-r}\right) e^{(\lambda+\nu)t} \end{split}$$

(thus

$$e^{-(\mu+\nu)t}M\int_{[0,t)}e^{(\mu+\nu)s}\mathbb{P}(\tau_1 \in ds) \le \frac{\mu+\nu}{\lambda+\nu} \cdot M\left(M + \frac{2r^{M+1}-r}{1-r}\right)e^{-(\mu-\lambda)t}$$

).

Justification of the equality marked with an asterisk (\*): By Fubini's theorem, we see

$$\int_{-\infty}^{t} (\mu + \nu) e^{(\mu + \nu)s} \int_{\Omega} \mathbf{1}_{\tau_{1}^{-1}([s,t])}(\omega) \mathbb{P}(d\omega) ds = \int_{\Omega} \int_{-\infty}^{t} (\mu + \nu) e^{(\mu + \nu)s} \mathbf{1}_{\tau_{1}^{-1}([s,t])}(\omega) ds \mathbb{P}(d\omega).$$

For each  $\omega \in \Omega$ , we have  $1_{(-\infty,t]}(\tau_1(\omega)) = 1_{[0,t]}(\tau_1(\omega))$  and

$$\int_{-\infty}^{t} (\mu + \nu) e^{(\mu + \nu)s} \mathbf{1}_{\tau_{1}^{-1}([s,t])}(\omega) ds = \mathbf{1}_{(-\infty,t]}(\tau_{1}(\omega)) \int_{-\infty}^{\tau_{1}(\omega)} (\mu + \nu) e^{(\mu + \nu)s} ds$$
$$= \mathbf{1}_{(-\infty,t]}(\tau_{1}(\omega)) \left[ e^{(\mu + \nu)s} \right]_{-\infty}^{\tau_{1}(\omega)} = \mathbf{1}_{(-\infty,t]}(\tau_{1}(\omega)) e^{(\mu + \nu)\tau_{1}(\omega)}.$$

Thus

$$\int_{\Omega} \int_{-\infty}^{t} (\mu + \nu) e^{(\mu + \nu)s} \mathbf{1}_{\tau_{1}^{-1}([s,t])}(\omega) ds \mathbb{P}(d\omega) = \int_{\Omega} \mathbf{1}_{[0,t]}(\tau_{1}(\omega)) e^{(\mu + \nu)\tau_{1}(\omega)} \mathbb{P}(d\omega)$$
$$= \mathbb{E} \left( e^{(\mu + \nu)\tau_{1}} \mathbf{1}_{\tau_{1}^{-1}([0,t])} \right).$$

We now desire an upper bound for  $\int_{[0,t)} s e^{(\mu+\nu)s} \mathbb{P}(\tau_1 \in ds)$ . We have

$$\begin{split} \int_{[0,t)} s e^{(\mu+\nu)s} \mathbb{P}(\tau_1 \in ds) &= \mathbb{E}(\tau_1 e^{(\mu+\nu)\tau_1} \mathbf{1}_{\tau_1^{-1}([0,t])}) \stackrel{*}{=} \int_{-\infty}^t ((\mu+\nu)s+1) e^{(\mu+\nu)s} \mathbb{P}(s \le \tau_1 \le t) ds \\ &\leq \int_{-\infty}^t ((\mu+\nu)s+1) e^{(\mu+\nu)s} e^{-(\mu-\lambda)s} \left(M + \frac{2r^{M+1} - r}{1-r}\right) ds \\ &= \left(M + \frac{2r^{M+1} - r}{1-r}\right) \int_{-\infty}^t ((\mu+\nu)s+1) e^{(\lambda+\nu)s} ds. \end{split}$$

Justification of the equality marked with an asterisk (\*): By Fubini's theorem, we see

$$\int_{-\infty}^{t} ((\mu + \nu)s + 1)e^{(\mu + \nu)s} \int_{\Omega} 1_{\tau_{1}^{-1}([s,t])}(\omega)\mathbb{P}(d\omega)ds$$
$$= \int_{\Omega} \int_{-\infty}^{t} ((\mu + \nu)s + 1)e^{(\mu + \nu)s} 1_{\tau_{1}^{-1}([s,t])}(\omega)ds\mathbb{P}(d\omega).$$

For each  $\omega \in \Omega$ , we have  $1_{(-\infty,t]}(\tau_1(\omega)) = 1_{[0,t]}(\tau_1(\omega))$  and

$$\int_{-\infty}^{t} ((\mu+\nu)s+1)e^{(\mu+\nu)s} \mathbf{1}_{\tau_{1}^{-1}([s,t])}(\omega)ds = \mathbf{1}_{(-\infty,t]}(\tau_{1}(\omega)) \int_{-\infty}^{\tau_{1}(\omega)} ((\mu+\nu)s+1)e^{(\mu+\nu)s}ds$$
$$= \mathbf{1}_{(-\infty,t]}(\tau_{1}(\omega)) \left[se^{(\mu+\nu)s}\right]_{-\infty}^{\tau_{1}(\omega)} = \mathbf{1}_{(-\infty,t]}(\tau_{1}(\omega))\tau_{1}(\omega)e^{(\mu+\nu)\tau_{1}(\omega)}.$$

Thus

$$\int_{\Omega} \int_{-\infty}^{t} ((\mu + \nu)s + 1)e^{(\mu + \nu)s} \mathbf{1}_{\tau_{1}^{-1}([s,t])}(\omega)ds\mathbb{P}(d\omega) = \int_{\Omega} \mathbf{1}_{[0,t]}(\tau_{1}(\omega))\tau_{1}(\omega)e^{(\mu + \nu)\tau_{1}(\omega)}\mathbb{P}(d\omega)$$

$$= \mathbb{E}(\tau_1 e^{(\mu+\nu)\tau_1} \mathbf{1}_{\tau_1^{-1}([0,t])}).$$

Next, we have

$$\int_{-\infty}^t ((\mu+\nu)s+1)e^{(\lambda+\nu)s}ds = \int_{-\infty}^t e^{(\lambda+\nu)s}((\lambda+\nu)s+1)ds + \int_{-\infty}^t e^{(\lambda+\nu)s}((\mu-\lambda)sds.$$

Then, we have

$$\int_{-\infty}^{t} e^{(\lambda+\nu)s}((\lambda+\nu)s+1)ds = \left[se^{(\lambda+\nu)s}\right]_{-\infty}^{t} = te^{(\lambda+\nu)t}$$

and

$$\int_{-\infty}^{t} e^{(\lambda+\nu)s}(\mu-\lambda)sds = \frac{\mu-\lambda}{(\lambda+\nu)^2} \left[ ((\lambda+\nu)s-1)e^{(\lambda+\nu)s} \right]_{-\infty}^{t} = \frac{\mu-\lambda}{(\lambda+\nu)^2} ((\lambda+\nu)t-1)e^{(\lambda+\nu)t},$$

so

$$\int_{-\infty}^{t} ((\mu+\nu)s+1)e^{(\lambda+\nu)s}ds = e^{(\lambda+\nu)t} \left( \left(\frac{\mu-\lambda}{\lambda+\nu}+1\right)t - \frac{\mu-\lambda}{(\lambda+\nu)^2} \right)$$
$$= e^{(\lambda+\nu)t} \left(\frac{\mu+\nu}{\lambda+\nu} \cdot t - \frac{\mu-\lambda}{(\lambda+\nu)^2}\right),$$

so

$$e^{-(\mu+\nu)t}(\mu-\lambda)\int_{[0,t)}se^{(\mu+\nu)s}\mathbb{P}(\tau_1\in ds)$$
(19)

$$\leq e^{-(\mu+\nu)t}(\mu-\lambda)\left(M+\frac{2r^{M+1}-r}{1-r}\right)e^{(\lambda+\nu)t}\left(\frac{\mu+\nu}{\lambda+\nu}\cdot t-\frac{\mu-\lambda}{(\lambda+\nu)^2}\right)$$
(20)

$$= (\mu - \lambda) \left( M + \frac{2r^{M+1} - r}{1 - r} \right) e^{-(\mu - \lambda)t} \left( \frac{\mu + \nu}{\lambda + \nu} \cdot t - \frac{\mu - \lambda}{(\lambda + \nu)^2} \right).$$
(21)

Thus

$$\begin{aligned} \int_{[0,t)} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) \mathbb{P}(\tau_1 \in ds) \\ \le \frac{\mu + \nu}{\lambda + \nu} \cdot M \left( M + \frac{2r^{M+1} - r}{1 - r} \right) e^{-(\mu - \lambda)t} \\ + (\mu - \lambda) \left( M + \frac{2r^{M+1} - r}{1 - r} \right) e^{-(\mu - \lambda)t} \left( \frac{\mu + \nu}{\lambda + \nu} \cdot t - \frac{\mu - \lambda}{(\lambda + \nu)^2} \right) \\ = e^{-(\mu - \lambda)t} \left( M + \frac{2r^{M+1} - r}{1 - r} \right) \left( \frac{\mu + \nu}{\lambda + \nu} \cdot M + (\mu - \lambda) \left( \frac{\mu + \nu}{\lambda + \nu} \cdot t - \frac{\mu - \lambda}{(\lambda + \nu)^2} \right) \right). \end{aligned}$$

Thus we have

$$d_{TV}(p_t(\vec{x},\cdot),\Pi) \leq \mathbb{P}(X_t \neq Y_t) \leq \mathbb{P}(\tau_{\text{couple}} \geq t)$$
  
$$\leq e^{-(\mu-\lambda)t} \left(M + \frac{2r^{M+1} - r}{1 - r}\right) \left((\mu - \lambda) \left(\frac{\mu + \nu}{\lambda + \nu} \cdot t - \frac{\mu - \lambda}{(\lambda + \nu)^2}\right) + \frac{\mu + \nu}{\lambda + \nu} \cdot M + 1\right). \quad (22)$$

Proof of Lemma 8. Let

$$D: \mathcal{S}^2 \to \mathcal{P}([M]): (\vec{a}, \vec{b}) \mapsto \{i \in [M]: a_i \neq b_i\}.$$

Note for the continuous-time Markov chain  $((X_t, Y_t))_{t \in [0,\infty)}$ , the non-zero distance-changing transition rates are

$$Q((\vec{a}, \vec{b}), (\vec{a}_j, \vec{b})) = \nu \pi_{1-a_j}, \ Q((\vec{a}, \vec{b}), (\vec{a}, \vec{b}_j)) = \nu \pi_{1-b_j} \quad (j \in D(\vec{a}, \vec{b})),$$
$$Q((\vec{a}, \vec{b}), (\vec{a}_j^{-}, \vec{b}_j^{-})) = \mu, \quad (j \in [M]).$$

In addition, note  $d = D(\vec{x}, \vec{y})$ . We can view deleting coordinates as matching them. For each initially unmatched coordinate  $i \in [M]$ , let

$$Z_i = \inf\{t \in [0, \infty) : (X_t)_i = (Y_t)_i\}$$

denote the time it takes to match the coordinates or delete them. Then the random variables  $Z_i$   $(i \in D(\vec{x}, \vec{y}))$  are independent exponential random variables, with rates  $\nu \pi_0 + \nu \pi_1 + \mu = \mu + \nu$ . If  $D(\vec{x}, \vec{y}) \neq \emptyset$ , then we have  $T = \max_{i \in D(\vec{x}, \vec{y})} Z_i$ . Thus  $\mathbb{P}(T \leq t) = (1 - e^{-(\mu + \nu)t})^{d_0}$ .

Thus

$$\mathbb{P}(T \ge t) = 1 - (1 - e^{-(\mu + \nu)t})^d \le 1 - (1 - e^{-(\mu + \nu)t})^M \le M e^{-(\mu + \nu)t}.$$
(23)

(For the last inequality, use Bernoulli's inequality (Theorem 13).)

(Note: We drew inspiration from Aldous and Fill's *Reversible Markov Chains and Random Walks on Graphs* [AF02, Subsection 12.1.4, *Continuous-time random walk on the* d-cube].)

# **4** Further directions

We obtained exponential  $d_{TV}$  ergodicity via coupling, and further we found the rate to be at least as good as  $\mu - \lambda + \varepsilon$  for all  $\varepsilon > 0$ . In this section, we detail further directions of research, namely improving our bound to get rid of the  $\varepsilon$ , proving a Poincaré inequality, proving a spectral decomposition and investigating similar models.

#### 4.1 Improvement of $d_{TV}$ coupling bound

We suspect we can improve our  $\mu - \lambda + \varepsilon$  term to  $\mu - \lambda$ . For instance, the estimate

$$1 - (1 - e^{-(\mu + \nu)(t-s)})^j < j e^{-(\mu + \nu)(t-s)}$$

may be good when we fix j, but it is not good when we fix t and s and vary j, as  $je^{-(\mu+\nu)(t-s)} \to \infty$  as  $j \to \infty$  and  $1 - (1 - e^{-(\mu+\nu)(t-s)})^j \le 1$  for all j.

A new proof might replace the sentence with equation 18 by the following sentence. For each  $s \in [0, t]$ , we have

$$\mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s) = \sum_{j=0}^{\infty} \mathbb{P}(\tau_2 \ge t - s \mid \tau_1 = s, \Lambda = j) \mathbb{P}(\Lambda = j \mid \tau_1 = s)$$

$$\leq \sum_{j=0}^{\infty} (1 - (1 - e^{-(\mu+\nu)(t-s)})^j) \mathbb{P}(\Lambda = j \mid \tau_1 = s) = 1 - \sum_{j=0}^{\infty} (1 - e^{-(\mu+\nu)(t-s)})^j \mathbb{P}(\Lambda = j \mid \tau_1 = s)$$
$$= 1 - \mathbb{E}((1 - e^{-(\mu+\nu)(t-s)})^{\Lambda} \mid \tau_1 = s).$$

Can we find a good bound for the expression  $1 - \mathbb{E}((1 - e^{-(\mu+\nu)(t-s)})^{\Lambda} \mid \tau_1 = s)$ ?

## 4.2 Poincaré inequality: canonical paths

We use an idea from Theorem 3.2 in Berestycki's notes [Ber16] and Theorem 3.2.1 in Saloff-Coste's notes [SC97] that applies for irreducible, reversible Markov chains. Let  $\Gamma$  denote the set of directed paths. Let  $\gamma : S^2 \to \Gamma$  such that for all  $\vec{x}, \vec{y}$ , the directed path  $\gamma(\vec{x}, \vec{y})$  starts at  $\vec{x}$  and ends at  $\vec{y}$ . For all  $\vec{u}, \vec{v} \in S$  such that  $e := (\vec{u}, \vec{v})$  is an edge, let  $Q(e) = 2^{-1}\lambda(\vec{u}, \vec{v})\Pi(\vec{u})$ . Let

$$C = \sup_{e} \left( \frac{1}{Q(e)} \sum_{\substack{\vec{x}, \vec{y} \in \mathcal{S}:\\\gamma(\vec{x}, \vec{y}) \ni e}} |\gamma(\vec{x}, \vec{y})| \Pi(\vec{x}) \Pi(\vec{y}) \right).$$

Define the function R from the edge set to  $\overline{\mathbb{R}}$  by

$$R(e) = \frac{2(1-r)}{\lambda(\vec{u},\vec{v})(r/2)^{|\vec{u}|}} \sum_{\substack{\vec{x},\vec{y}\in\mathcal{S}:\\\gamma(\vec{x},\vec{y})\ni e}} |\gamma(\vec{x},\vec{y})| \left(\frac{r}{2}\right)^{|\vec{x}|+|\vec{y}|},$$

where  $\vec{u}, \vec{v}$  are the unique elements of S such that  $e = (\vec{u}, \vec{v})$ . ("R" is the next letter after "Q".) Then

$$C = \sup_{e} R(e).$$

**Proposition 9.** For all f,

 $\operatorname{Var}_{\Pi}(f) \leq C\mathcal{E}(f, f).$ 

(As usual, we interpret  $\infty \cdot a = \infty$   $(a \in (0, \infty))$  and  $\infty \cdot 0 = 0$ .)

*Proof.* For all  $\vec{u}, \vec{v} \in S$  such that  $e := (\vec{u}, \vec{v})$  is an edge, we define the operator  $\Delta_e$  by  $\Delta_e(f) = f(\vec{v}) - f(\vec{u})$ . Using this notation, we see

$$\forall f \quad \mathcal{E}(f, f) = \sum_{e} \Delta_{e}(f)^{2} Q(e),$$

by equation 8. We see for all f,

$$\operatorname{Var}_{\Pi}(f) = (\text{by definition}) \sum_{\vec{x} \in \mathcal{S}} (f(\vec{x}) - \sum_{\vec{y} \in \mathcal{S}} \Pi(\vec{y}) f(\vec{y}))^2 \Pi(\vec{x})$$
$$= (\operatorname{since} \sum_{\vec{y} \in \mathcal{S}} \Pi(\vec{y}) = 1) \sum_{\vec{x} \in \mathcal{S}} \left( \sum_{\vec{y} \in \mathcal{S}} (f(\vec{x}) - f(\vec{y})) \Pi(\vec{y}) \right)^2 \Pi(\vec{x})$$

$$\leq (\operatorname{Cauchy-Schwarz}) \sum_{\vec{x}\in\mathcal{S}} \left( \sum_{\vec{y}\in\mathcal{S}} (f(\vec{x}) - f(\vec{y}))^2 \Pi(\vec{y}) \right) \Pi(\vec{x})$$

$$= \sum_{\vec{x}\in\mathcal{S}} \sum_{\vec{y}\in\mathcal{S}} \left( (f(\vec{x}) - f(\vec{y}))^2 \Pi(\vec{x}) \Pi(\vec{y}) \right) = \sum_{\vec{x},\vec{y}\in\mathcal{S}} \left( \sum_{e\in\gamma(\vec{x},\vec{y})} \Delta_e(f) \right)^2 \Pi(\vec{x}) \Pi(\vec{y})$$

$$\leq (\operatorname{Cauchy-Schwarz}) \sum_{\vec{x},\vec{y}\in\mathcal{S}} |\gamma(\vec{x},\vec{y})| \sum_{e\in\gamma(\vec{x},\vec{y})} \Delta_e(f)^2 \Pi(\vec{x}) \Pi(\vec{y})$$

$$= \sum_{\vec{x},\vec{y}\in\mathcal{S}} \sum_{e} [e \in \gamma(\vec{x},\vec{y})] |\gamma(\vec{x},\vec{y})| \Delta_e(f)^2 \Pi(\vec{x}) \Pi(\vec{y}) = \sum_{e} \sum_{\substack{\vec{x},\vec{y}\in\mathcal{S}:\\\gamma(\vec{x},\vec{y})\ni e}} |\gamma(\vec{x},\vec{y})| \Delta_e(f)^2 \Pi(\vec{x}) \Pi(\vec{y})$$

$$= \sum_{e} \left( \frac{1}{Q(e)} \sum_{\substack{\vec{x},\vec{y}\in\mathcal{S}:\\\gamma(\vec{x},\vec{y})\ni e}} |\gamma(\vec{x},\vec{y})| \Pi(\vec{x}) \Pi(\vec{y}) \right) \Delta_e(f)^2 Q(e) \leq C\mathcal{E}(f,f).$$

## 4.3 Spectral decomposition

In this section, we abbreviate  $p_t(\vec{x}, \vec{y}) = \mathbb{P}_{\vec{x}}(X_t = \vec{y})$ , overloading the symbol p.

By Kendall's representation theorem (Theorem 1.6.5 in [And91]), there exists a doubly-indexed set  $\{\gamma_{\vec{x},\vec{y}}\}_{\vec{x},\vec{y}\in\mathcal{S}}$  of finite signed measures  $\mathcal{B}([0,\infty)) \to \overline{\mathbb{R}}$  such for each  $\vec{x}, \gamma_{\vec{x},\vec{x}}$  is a probability measure and such that the transition function P has the representation

$$\forall \vec{x}, \vec{y} \in \mathcal{S} \quad \forall t \ge 0 \quad p_t(\vec{x}, \vec{y}) = \left( r^{|\vec{y}| - |\vec{x}|} \cdot \frac{\prod_{i \in [|\vec{y}|]} \pi_{y_i}}{\prod_{i \in [|\vec{x}|]} \pi_{x_i}} \right)^{1/2} \int_{[0,\infty)} e^{-tx} d\gamma_{\vec{x}, \vec{y}}(x).$$

The Karlin–McGregor theorem (Theorem 8.2.1 in [And91]; see also Subsubsection 2.4) concerns continuous-time birth-death processes, and it is more refined than Kendall's theorem (Theorem 1.6.5 in [And91]). The Karlin–McGregor theorem not only states existence of some finite signed measures, but also tells us we can choose the measures such that satisfy a certain relation involving a sequence of polynomials.

Perhaps we can derive an analogue of the Karlin–McGregor theorem for the indel chain. Questions to consider: Does the chain have discrete spectrum? Does it have a spectral gap?

### 4.4 Other models

Do similar results hold for other indel models, such as the Poisson indel model of Bouchard-Côté and Jordan [BCJ13]?

## **5** Acknowledgements

Thanks to Prof. Wai-Tong Fan for guidance, and thanks to Bat Dejean for spending a lot of time helping me. Thanks to Prof. Chris Connell for organizing this REU program, and thanks to the NSF for funding this REU program.

# A List of theorems

Theorem 10 (see Theorem 3 in [MB07]).

$$\forall i \ge 0 \quad \sum_{n=0}^{\infty} n |p_t(i,n) - \gamma_n| \le \left(i + \frac{\lambda}{\mu - \lambda}\right) e^{-(\mu - \lambda)t}.$$

**Theorem 11** (Poincaré inequality; Theorem 2.18 in [vH16]). Say  $P_t$  is a reversible ergodic Markov semigroup with stationary measure  $\mu$ . Say  $c \in (0, \infty)$ . The following statements are equivalent:

(Poincaré inequality);

**Theorem 12** (Coupling; Proposition 4.7 in [LP17]). Say  $\mathcal{X}$  is a countable (i.e., finite or denumerable (i.e., countably infinite)) set, and say  $\mu, \nu : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$  are two probability measures. Then

$$d_{TV}(\mu,\nu) = \inf\{\mathbb{P}(X \neq Y) : (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\}$$

**Theorem 13** (Bernoulli's inequality [Wik19]). (See Figure 2.) For all real numbers  $x \ge -2$  and integers  $r \ge 0$ ,  $(1 + x)^r \ge 1 + rx$ .

**Theorem 14** (Variance identity). Say X is a countable (i.e., finite or denumerable (i.e., countably infinite)) set, and say  $\mu : \mathcal{P}(S) \to \mathbb{R}$  is a finite measure such that  $\forall x \in X \quad \mu(\{x\}) > 0$ . Say  $f \in L^2(\mu)$ . For each x, let  $\mu_x = \mu(\{x\})$ . Then

$$\operatorname{Var}_{\mu}(f) = \frac{1}{2} \sum_{x} \sum_{y} (f(x) - f(y))^2 \mu_x \mu_y.$$

Proof. We have

$$\frac{1}{2} \sum_{x} \sum_{y} (f(x) - f(y))^{2} \mu_{x} \mu_{y} = \frac{1}{2} \sum_{x} \mu_{x} \sum_{y} ((f(x)^{2} - 2f(x)f(y) + f(y)^{2})\mu_{y}$$

$$= \frac{1}{2} \sum_{x} \mu_{x} (\sum_{y} f(x)^{2} \mu_{y} - 2\sum_{y} f(x)f(y)\mu_{y} + \sum_{y} f(y)^{2} \mu_{y} = \frac{1}{2} \sum_{x} \mu_{x} (f(x)^{2} - 2f(x)\mu f + \mu f^{2})$$

$$= \frac{1}{2} \sum_{x} \mu_{x} f(x)^{2} - 2\sum_{x} \mu_{x} f(x)\mu f + \mu f^{2} = \mu f^{2} - (\mu f)^{2} = \operatorname{Var}_{\mu}(f).$$



Figure 2: Bernoulli's inequality. Here, r = 3.

**Theorem 15** (Variance inequality; see [Fan19, Lemma 1] or [SC97, First two sentences, Proof of Theorem 3.2.1]). Say S is a countable (i.e., finite or denumerable (ie., countably infinite)) set, and say X is an irreducible Markov chain. Let A denote the set of arcs with positive transition probability. Let  $\Gamma$  denote the set of directed paths in the digraph (X, A), and let  $\gamma : S^2 \to \Gamma$  such that for all  $x, y \in S \ \gamma(x, y)$  begins at x and ends at y. Say  $\mu : \mathcal{P}(S) \to \mathbb{R}$  is a finite measure such that  $\forall x \in S \ \mu(\{x\}) > 0$ . Say  $f \in L^2(\mu)$ , and for each arc  $e \text{ let } \Delta_f(e) = f(e_2) - f(e_1)$ . (Note: We treat arcs as ordered pairs.) Then

$$\operatorname{Var}_{\mu}(f) \leq \frac{1}{2} \sum_{x,y \in S} \mu_{x} \mu_{y} |\gamma(x,y)| \sum_{e \in \gamma(x,y)} \Delta_{f}(e)^{2}.$$

*Proof.* For all  $x, y \in S$ , we have  $f(y) - f(x) = \sum_{e \in \gamma(x,y)} \Delta_f(e)$ . and so, applying the Cauchy–Schwarz inequality, we have

$$(f(y) - f(x))^2 \le |\gamma(x, y)| \sum_{e \in \gamma(x, y)} \Delta_f(e)^2.$$

Thus

$$\frac{1}{2}\sum_{x,y\in S} (f(x) - f(y))^2 \mu_x \mu_y \le \frac{1}{2}\sum_{x,y\in S} \mu_x \mu_y |\gamma(x,y)| \sum_{e\in\gamma(x,y)} \Delta_f(e)^2.$$

By Theorem 14, the left-hand side equals  $\operatorname{Var}_{\mu}(f)$ .

19
----

# References

- [AF02] David Aldous and James Allen Fill. Reversible Markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014, retrieved from http://www.stat.berkeley.edu/~aldous/RWG/book.html.
- [And91] William J Anderson. *Continuous-time Markov chains: An applications-oriented approach*. Springer, 1991.
- [BCJ13] Alexandre Bouchard-Côté and Michael I Jordan. Evolutionary inference via the Poisson indel process. *Proceedings of the National Academy of Sciences*, 110(4):1160–1166, 2013.
- [Ber16] Nathanaël Berestycki. Mixing times of Markov chains: Techniques and examples, November 22, 2016. Lecture notes retrieved from http: //www.statslab.cam.ac.uk/~beresty/Articles/mixing3.pdf.
- [Che04] Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific, second edition, 2004.
- [DR10] Constantinos Daskalakis and Sebastien Roch. Alignment-free phylogenetic reconstruction. In Annual International Conference on Research in Computational Molecular Biology, pages 123–137. Springer, 2010.
- [Fan19] Wai-Tong Fan. Poincaré inequalities for detailed balanced biochemical models, 2019. Nine pages of unpublished notes.
- [Fle04] Roland Fleißner. Sequence alignment and phylogenetic inference. Logos Verlag Berlin, 2004.
- [FR17] Wai-Tong Fan and Sebastien Roch. Statistically consistent and computationally efficient inference of ancestral DNA sequences in the TKF91 model under dense taxon sampling. *arXiv preprint arXiv:1707.05711*, 2017.
- [KM58] Samuel Karlin and James McGregor. Linear growth, birth and death processes. *Journal* of Mathematics and Mechanics, pages 643–662, 1958.
- [KS66] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. Ann. Math. Statist., 37:1463–1481, 1966.
- [LP17] David A Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, second edition, 2017.
- [MB07] Alexander Y Mitrophanov and Mark Borodovsky. Convergence rate estimation for the TKF91 model of biological sequence length evolution. *Mathematical biosciences*, 209(2):470–485, 2007.
- [Mos01] E. Mossel. Reconstruction on trees: beating the second eigenvalue. *Ann. Appl. Probab.*, 11(1):285–300, 2001.

- [Mos04] E. Mossel. Survey: Information flow on trees. In J. Nestril and P. Winkler, editors, *Graphs, morphisms and statistical physics*, pages 155–170. Amer. Math. Soc., 2004.
- [MT01] István Miklós and Zoltán Toroczkai. An improved model for statistical alignment. In *International Workshop on Algorithms in Bioinformatics*, pages 1–10. Springer, 2001.
- [Roc12] Sebastien Roch. Lecture notes on stochastic processes in evolutionary genetics, 2012. Retrieved from http://www.math.wisc.edu/~roch/evol-gen.
- [SC97] Laurent Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics*, pages 301–413. Springer, 1997.
- [Sch00] Wim Schoutens. Stochastic processes and orthogonal polynomials. Springer, 2000.
- [TKF91] Jeffrey L Thorne, Hirohisa Kishino, and Joseph Felsenstein. An evolutionary model for maximum likelihood alignment of DNA sequences. *Journal of Molecular Evolution*, 33(2):114–124, 1991.
- [TKF92] Jeffrey L Thorne, Hirohisa Kishino, and Joseph Felsenstein. Inching toward reality: an improved likelihood model of sequence evolution. *Journal of Molecular Evolution*, 34(1):3–16, 1992.
- [vH16] Ramon van Handel. Probability in high dimension, December 21, 2016. Lecture notes retrieved from https://web.math.princeton.edu/~rvan/APC550.pdf.
- [Wik19] Wikipedia contributors. Bernoulli's inequality Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Bernoulli's\_inequality, 2019. Accessed June 24, 2019.