DOUBLE BUBBLES IN SPACES OF CONSTANT CURVATURE

BY
JOSEPH A. CORNELI

A Thesis

Submitted to the Division of Natural Sciences
New College of Florida,
in partial fulfillment of the requirements for the degree
of Bachelor of Arts
Under the sponsorship of Patrick T. McDonald

Sarasota, Florida
May, 2002

Dedication

To Bea Canchester and Helen Corneli, who encouraged me to pursue my education and have made significant material contributions that have made it possible.

Acknowledgements

My father's mother, to whom, together with my mother's mother, this thesis is jointly dedicated, told me at an early age that "it takes a village to raise a child". If this is true of my personal life (and it is) it is even more true of my academic life. This leaves me with many people to thank.

First of all there is my academic sponsor and thesis advisor, Pat McDonald. Pat taught me lots of math and has been very generous with both time and energy spent talking to me about my thesis, and about many other things. It is no coincidence that while I was his student I began to think seriously about how my mind works, to have solid hope for the future, to give respect where it is due and to value intellectual freedom. In short, Pat showed me what a mathematician is and opened the way for me to try it out for myself. In addition to all of this, he has a number of remarkable personal qualities, including his own special blend of modesty and confidence, that I have learned an immense amount by trying (and usually failing) to emmulate.

Then there is Frank Morgan of Williams College. Frank was my devoted advisor during a full year of research that lead to three and a half papers, that, in revised form, constitute the body of this thesis. While I was in his Geometry Group at Williams, Frank made daily comments on daily revisions of our papers every weekday morning (starting on our second day). Afterwards, he continued to advise me and my collaborators, mostly through an extensive email correspondence, in which he made hundreds of thoughtful suggestions about both research and writing. Working for Frank was always fun for me, in part because I was learning so much, but also because of his pleasant enthusiasm and congeniality. In addition to this, he arranged a number of special opportunities for the Geometry Group, including the unprecidented invitation of a group of undergraduates to the Mathematical Sciences Research Institute in Berkeley, California, and full participation in his special session on soap bubble geometry at MathFest'01 in Madison, Wisconsin. By observing Frank in action, I learned an immense amount about how to interact positively with people.

Without the hard work of these two men, both exemplary scholars and teachers, this thesis would not exist. In addition to making that acknowledgement, I want to take this moment to thank each of them for taking a friendly interest in my personal and moral development as well as in my academics.

In addition to being jointly responsible for the majority of the content of this thesis, my collaborators during this past year were both my peers and my teachers, and they formed an invaluable asset to my general education in mathematics.

My collaborators in the 2001 Geometry Group at the Williams REU were Nick Leger of the University of Texas at Austin, and Paul Holt and Eric Schoenfeld of Williams College. Ben Steinhurst from Williams and George Lee from Harvard University will be in the 2002 Geometry Group and they worked with my group during 2001-2002 academic year. Revised versions of our results appear in Chapters 4 and 6.

For three weeks of the summer, the 2001 Geometry Group attended the Clay Mathematics Institute's Summer School on the Global Theory of Minimal Surfaces at the Mathematical Sciences Research Institute in Berkeley, California. Collaborators I met at the summer school were Miguel Carriòn Àlvarez of the University of California at Riverside, Genevieve Walsh of the University of California at Davis, and Shabnam Beheshti of Texas Tech University. A revised version of our joint paper appears in Chapter 7.

Aside from these people who I worked intimately with, there are many others in the world of mathematics at large who made important contributions. In Berkeley, I had a number of helpful conversations with Michael Hutchings, whose theory of double bubble structure underlies the current understanding of the double bubble problem in three and more dimensions, and is treated here in Chapter 3. There I also met David Futer of Stanford University, with whom I had a lengthy correspondence with about the Euclidean version of the Curvature Conjecture appearing in Chapter 4. David's careful work and thoughtful letters contributed greatly to my understanding of the subject

of double bubbles and mathematics in general. During the Summer School, conversations with Matt Kudzin of Indiana State were also very illuminating. I also had several helpful and fun conversations with Manuel Ritoré of the University of Grenada about bubbles in the three-torus and John Sullivan of the University of Illinois at Champaign/Urbana about tilings (and later, by email, about foams).

Andy Cotton and David Freeman, both graduating from Harvard this year, were members of the 2000 Geometry Group, and my group continued their work on double bubbles in spherical and hyperbolic space. I had the pleasure of meeting Andy at MathFest'01 in Madison, and then met again with Andy and David in Cambridge, Massachusetts this past fall, where we talked about double bubbles. At MathFest I also met Wacharin Wichiramala of the University of Illinois Champagne-Urbana, whose results on planar clusters include a lemma that was very important for developing the theory of double bubbles in the two-torus. Wacharin made a number of very helpful suggestions by email during the academic year. At MathFest I also met Joel Foisy, whose Geometry Group in 1992 solved the double bubble problem in the plane, and who proved in his Williams honors thesis a result that is important in Hutchings' structure theory, and whose talk on double bubbles in corners contributed to the Geometry Group's understanding of bubbles on the cylinder and torus.

Notable contributions to the three torus project were made in Berkeley by three of the faculty participants, Michael Dorff, Denise Halverson, and Gary Lawlor, of Brigham Young University, who got us started with *Surface Evolver* computations, and also by David Futer and Matt Kudzin (who I already mentioned), Baris Coskunuzer, Brian Dean, Tom Fleming, Jim Hoffman, Jon Hofmann, Jesse Ratzkin, and Jean Steiner, as well as the entire Geometry Group.

Back at New College, where lots of the writing and tons of the editing took place, Mary Whelan and was a big overall help. In particular, going well beyond her call of duty as an employee of the New College Writing Resource Center, made detailed comments on the Geometry appendix. Conversations with her helped me set the tone for the thesis, and I think enormously contributed to its overall comprehesibility. Yair Kagan produced the illustration of the Hexagonal Honeycomb

appearing at the end of Chapter 7. Holly Barone is the person in charge of interlibrary loans and article copies at the New College library. She worked hard to obtain copies of the books and articles for me. References that were too hard for her to find I gave up on. Tim Teräväinen of New College has had lots of helpful conversations with me about everything; we've lived around each other in three cities and I appreciate his steadfastness. Rob Meyers and Caroline Arruda had the difficult task of being friends with me during my first two years at New College. They helped make things interesting for me here. In particular, Rob and I had lots of fun talking about math together, and it was out of conversations related to his thesis [31] that I began to have some conception of the direction I wanted to go with mine. I am most happy that my dreams about that have come true. Friends who were especially helpful in one way or another during the writing process are Kartina Amin, Justin Clarke-Doane, Selena Lee, Siggy Meek, Ya'el Morowati, Jake Silverstein, Tik Sun, and Stella Tinnirello.

Long-distance from Minneapolis, my sister Zoë Corneli has been a support to me throughout my time at New College.

My dad, Steve Corneli, has had many many wonderful conversations with me over the years. He has also generously sustained all of my living expenses this past semester while I was in the process of writing this thesis, and has covered various other expenses. I confess it here, that I first began to study math at his prompting. I also first began to love it under his tutelage (in geometry and basic calculus) when he asked me point blank whether I really wanted to understand the stuff or not (I did). My dad was also the first person to teach me anything memorable about writing. Judy Knapp is a close friend of my dad's. She has listened to me and encouraged me about many personal issues, and has also moderated my dad's influence on me.

During both the research and writing phases of this project my mother, Lynne Corneli, has made a number of suggestions at key times that improved the quality of my work or writing and that re-focussed my mind. Most of this thesis was composed on a Hewlett-Packard computer she gave

me.

My research at Williams College was supported by the NSF. The Mathematics and Statistics Department at Williams College generously gave me travel support. Expenses incurred in connection with the Summer School were covered by the Clay Mathematics Institute and the Mathematical Sciences Research Institute. I should like to personally thank the organizers of the Summer School: Joel Hass, Arthur Jaffe, Antonio Ros, Harold Rosenberg, Richard Schoen, and Michael Wolf, and especially David Hoffman, who bravely agreed to support the participation of me and three other undergraduates. The Mathematical Association of America provided travel support for my trip to MathFest'01 in Madison, Wisconsin.

A semester of advanced study at the Massachusetts Institute of Technology contributed greatly to this thesis and to my general education in mathematics. This endeavor was underwritten by my paternal grandparents, Kip and Helen Corneli. My professors at MIT were Victor Guillemin, Richard Melrose, and Andras Vašy.

My time at New College was underwritten by my maternal grandparents, Jack and Bea Canchester, and by what used to be called the Dean and Warden's Leadership Scholarship, as well as various other grants from the state of Florida. At New College my professors were Pat McDonald, David Mullins, and Erini Poimenidou (in mathematics), George Rupeiner and Don Colladay (in physics), Catherine Elliot (in economics), Sandra Gilchrist (in biology), and Mac Miller and John McDiarmid (in English).

In addition to all of these people, my thesis is heavily indebted to writings by Almgren [25], Chavel [8], Hutchings [22], Nomizu and Kobayashi [39], Morgan [34], Rudin [45], Spivak [49], and is somewhat better than metaphysically linked to writings by Vinberg [12], Wolf [59], and numerous others.

Mathematics is owned by no one, but writing most certainly is, and I endeavor to point out in the text of the thesis everywhere I borrowed material and to give credit to all of the authors I read

while I wrote. At places where a specific citation is not given, that does not necessarily imply that what I am writing about at that moment is in any way an original idea of my own. In general, I am fairly sure that the distinction between new and old ideas will be clear enough to those who care to think about such things.

for latest
Anti-Errata List

Email me.

Contents

List of Figures

There is in addition one table, entitled "Computing the form of the Levi-Civita connection" that appears in 8.1.

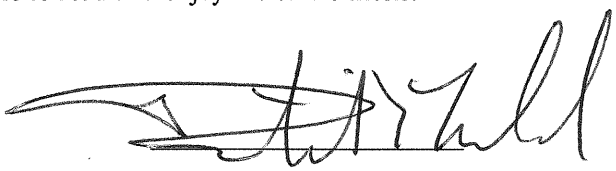
DOUBLE BUBBLES IN SPACES OF CONSTANT CURVATURE

Joseph A. Corneli

New College of Florida, 2002

ABSTRACT

The aim of this thesis is to be a comprehensive and self-contained treatment of the known results on area minimizing double bubbles in constant curvature spaces. Difficult proofs of existence and regularity theorems are omitted, as are several more readible proofs that are available (or soon to be made available) in the literature. Included in the thesis are a number of new results, including the reduction of the double bubble conjecture in spherical and hyperbolic space to a problem of asymptotic analysis, a proof of that the least-area way to divide Gauss space ($\mathbf{R}^n$ with Gaussian measure) into three equal pieces consists of three half-hyperplanes meeting at the origin at 120°, and results due to the author and various collaborators on double bubbles it two- and three-dimensional tori. The theoretical background of these developments is surveyed in some detail. A proof of the isoperimetric property of spheres in the classical constant curvature space forms is presented. Various detailed appendices are provided with the hope that persons with little or no background in differential geometry or variational methods will be able to read and enjoy the entire thesis.



Patrick T. McDonald

Division of Natural Sciences

'Perchè l' animo tuo tanto s' impiglia,'

Disse il Maestro, 'che l' andare allenti?'

Che ti fa ciò che quivi si pispiglia?

Vien rentro a me, e lascia dir le genti;

Sta come torre ferma, che non crolla

Giammai la cima per soffiar de' venti.

Chè sempre l' uomo in cui pensier rampolla

Sopra pensier, da sè dilunga il segno,

Perchè la foga l' un dell' altro insolla.'

Dante Alighieri, *Purgartorio* V, 10-18

Chapter 1

themes

- 'Double bubbles'?

- 'Spaces of constant curvature'?

Bubbles are regions in a geometric space (like Euclidean space $\mathbf{R}^n$) that minimize surface area given a volume constraint.

The single bubble or isoperimetric problem seeks the least area way to enclose one volume in a given geometric space.

The double bubble problem seeks the least area way to enclose and separate two given volumes in given geometric space – and so on.

Greek mathematicians knew the solution to the isoperimetric problem in Euclidean space well before the start of the common era. The answer is that the least-area enclosure of a given volume is a sphere, or in two dimensions, a circle. Euler is credited with the first rigorous proof of the isopermetry of circles in the plane. In the mid-eighteen-hundreds, analysts Schwarz (see [47]) and Steiner (see [50]) developed the methods that finally made possible a rigorous proof of the

isoperimetry of spheres in three dimensions. The key idea in their arguments is that symmetrization always reduces perimeter.

The idea of minimizing the surface area for given volume constraint(s) is a classic example of a variational problem. (We have some functional, namely area, defined on the space of all enclosures; we want to know which enclosures with a fixed volume are extrema for this functional.) This type of problem readily generalizes to spaces with non-Euclidean geometry. (The isoperimetry of spheres in the classical non-Euclidean spaces was first shown by Schmidt (see [46]) in the mid-nineteen-hundreds.)

The idea that bubbles or clusters of bubbles solve the least-area enclosure problem comes from our experience in the real world. Empiricism tells us that we should be able to restrict ourselves from the set of all enclosures, on which we impose very few constraints (say only that their boundaries be 'not too bad'), to a subset of enclosures that about which we can say something geometrically substantial. Thus, one of the primary tasks in this problem area was to characterize the geometry of minimizing idealized bubble clusters – an even higher-priority task was to say when a minimizer for the least area problem exists. The basic geometrical properties of the smooth (regular) and non-smooth (singular) sets of minimizing bubble clusters also characterize a class of 'potential minimizers', which are said to satisfy *regularity*. This term is common in variational calculus and partial differential equations, mainly because it turns out that the solutions (minimizers) are usually smooth, that is, regular, almost everywhere. This was proved for the least-area enclosure problem by Almgren [25]; a more specific characterization of the geometry of minimizers was given by Taylor ([53], [54]) and subsequently generalized by Brian White (unpublished).

Once we understand regularity, the problem becomes mostly geometric (and/or computational). Even with regularity, we do not know whether there are a finite number of components in a minimizing bubble cluster, which regular clusters are stable in the sense that they stay regular under volume preserving variationsm, and at the bottom of it all, what shapes are actually minimizing. In non-

standard geometries, the intuition gathered from every-day empiricism wears somewhat thin, and we may have to do some serious thinking to even come up with a reasonable conjecture. Computer simulations can be an aid to intuition.

Spaces with constant curvature are among the simplest things that a geometer can study. This class includes the plane (see Figure 1.1), the sphere (see Figure 1.2), the hyperbolic plane (see Figure1.3), the cylinder, and the flat torus (rectangle or box with identified sides, as in Figures 1.4 and 1.5), as well as the Euclidean, Spherical, and Hyperbolic spaces, and many other examples. Each of these geometrical spaces has the property of being 'pretty much the same' everywhere. This property gives us a relatively large set of tools. In particular, area minimizing bubble clusters in these spaces are described by regularity theorems similar to those proved for Euclidean space. However, a wide range of geometrical and topological complications unknown in Euclidean space make themselves felt in the more general category. These lead to new computational and theoretical challenges.

Figure 1.1: The Euclidean plane, or rather a segment of it, shown here tiled with regular hexagons.

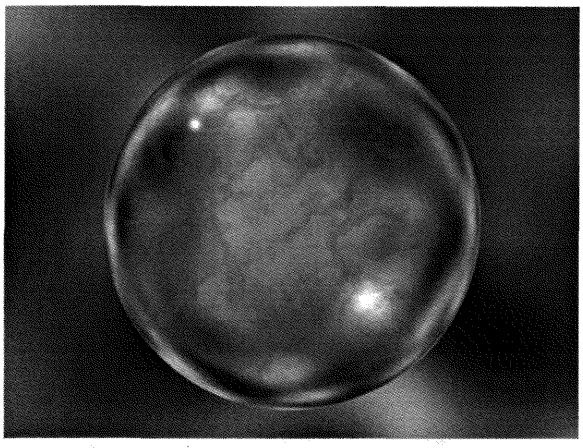


Figure 1.2: The sphere, filled with various media. This computer-generated image looks remarkably like a soap bubble.

Plan of the Work Chapter 2 is a review of the results relevant to the subsequent chapters and in other ways important to this thesis. A proof of the minimizing property of spheres for the single bubble problem in the classical constant curvature spaces (spheres, Euclidean space, and hyperbolic space) is presented, and the known results on the regularity properties of minimizing bubble clusters are described. Chapter 3 reviews the known results on the structure of area-minimizing double bubbles. In Chapter 4 we present a partial solution to the double bubble for spherical space $\mathbf{S}^3$ and hyperbolic space $\mathbf{H}^3$. In Chapter 5 we present a solution to a special case of the double bubble problem for n-dimensional Gauss space ($\mathbf{R}^n$ with gaussian measure), namely the case in which the two given volumes each take up exactly one third of the volume of the whole space. (This is sort of an 'interlude' since even though this is a perfectly fine space of constant curvature, we won't be directly concerned with curvature properties of the space in this chapter. We do however rely on the known theory for spheres.) In Chapter 6 we present a (more or less complete) solution to the double bubble problem for the flat two torus $\mathbf{T}^2$. In Chapter 7 we present a conjectured solution to the double bubble problem for the flat three torus $\mathbf{T}^3$. And at the end we include a short chapter (really an epilog) that puts together all these various examples and talks in a general way about

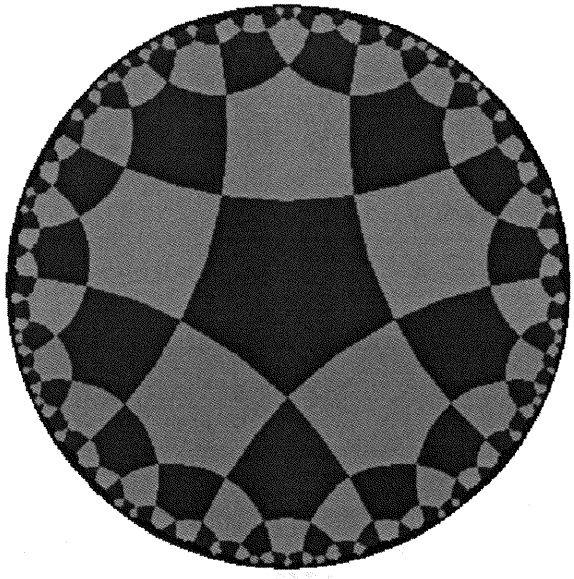


Figure 1.3: The Hyperbolic plane, shown here tiled with regular pentagons. It is not possible to tile the Euclidean plane with regular pentagons, but we can use them to tile the hyperbolic plane because of the way it curves. In particular, the curvature causes areas to expand quickly as you move away from any point – the pentagons look like they are different shapes because of hyperbolic perspective; they are actually exact isometric copies of one another.

what might be true about other related spaces.

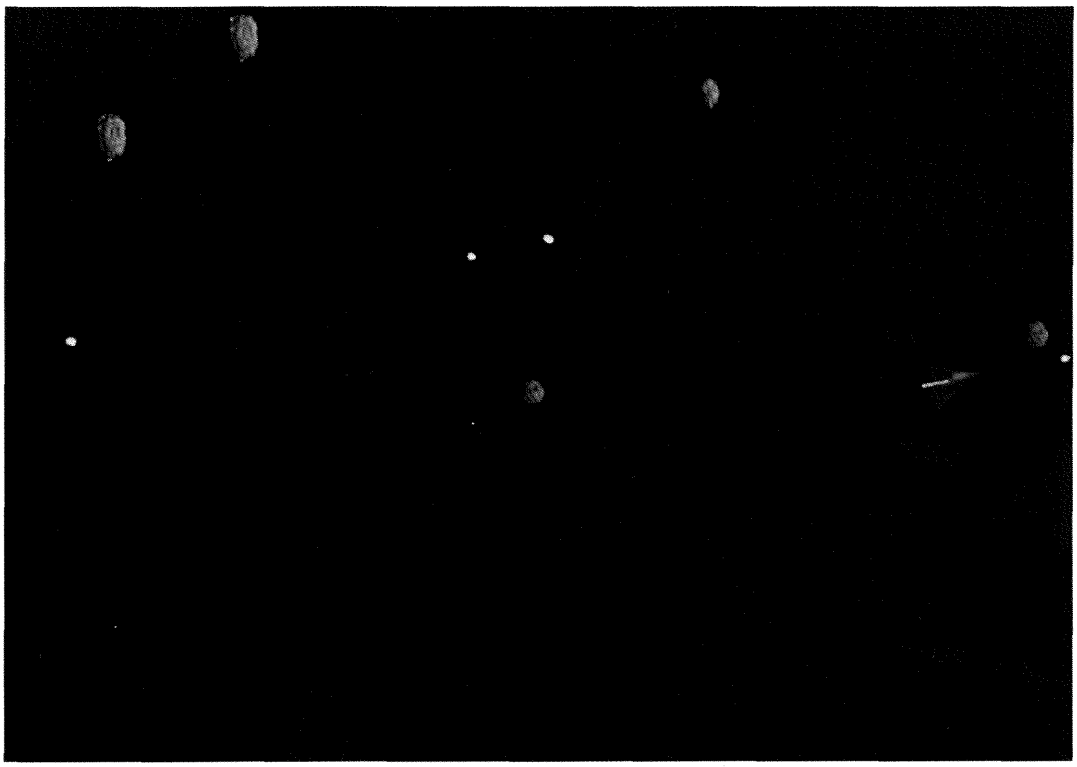


Figure 1.4: The two-torus, viewed from above, with a motif similar to the one in which it it made its debut in popular culture back in 1979. Shots, ships, and asteroids that fly off of the screen on one side instantly reappear on the opposite side (similarly with top and bottom).

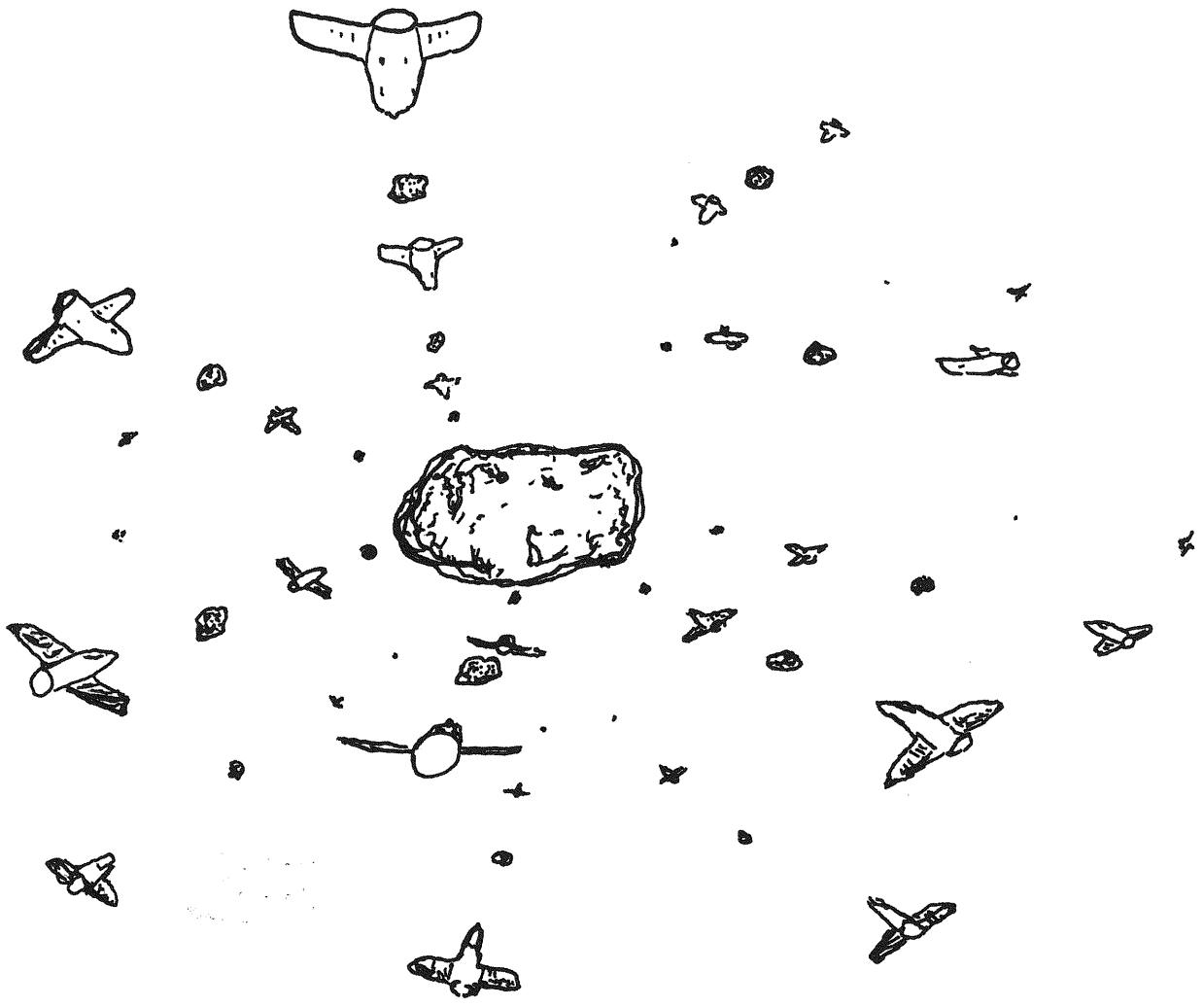


Figure 1.5: The three-torus, viewed from inside. What appears to be multiple ships bearing down on multiple asteroids are actually all just one ship (ours) bearing down on one asteroid, viewed in 'toroidal perspective'. Drawing based loosely on a screenshot from the Jeff Weeks/Geometry Center video, *The Shape of Space* [58]

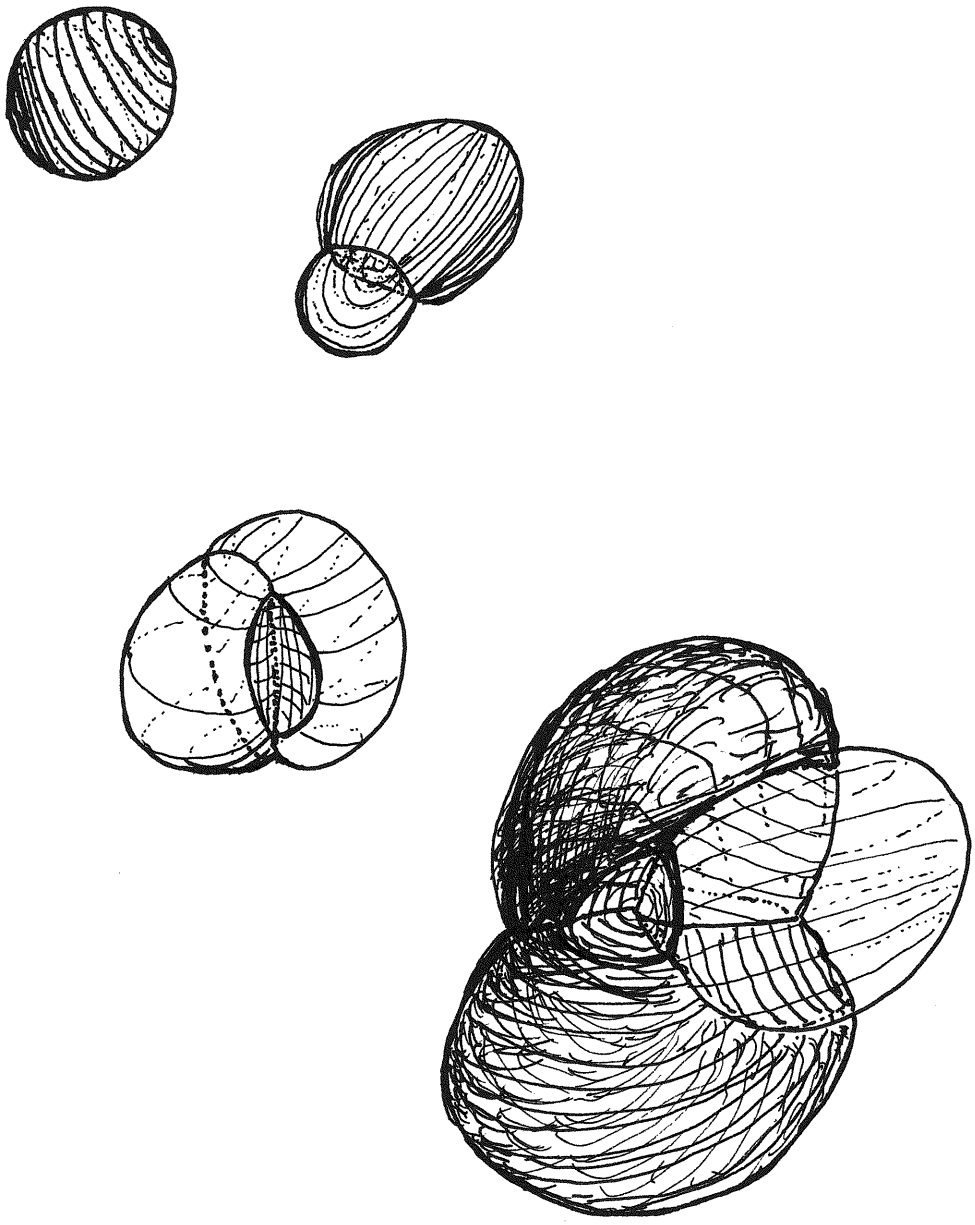


Figure 1.6: Bubble Clusters

Chapter 2

general background

Introduction

The beginning. We want to get some feeling for the known results. Other than Euclidean space, the best-understood geometric spaces are the 'classical' non-Euclidean geometries. Understanding the geometry of these spaces and the basic theory of area-minimizing bubbles in them is what is emphasized in this chapter.

The basics. We should know what the the simply connected spaces with constant curvature are and some of their basic properties, including their symmetries. These are the classical Euclidean and non-Euclidean geometries; they are collectively called *constant curvature space forms*. The basic properties of constant curvature space forms are reviewed in section 2.1. We should understand the important, well-known, result on the isoperimetric or 'single bubble' problem, which states that of regions with a fixed volume in constant curvature space forms, the ones with the least surface area are round balls. We will present a geometric proof of this fact in section 2.2. We should understand

the basic results of geometric measure theory on the existence and regularity properties of solutions to surface area minimization problems when there are multiple volumes to enclose and separate. These results are reviewed in section 2.3.

In this thesis we are primarily interested in understanding the case of two volumes, or the 'double bubble' problem, and the historical development of this problem up to the current date. It is known that a 'standard double bubble', like that pictured in Figure 2.1, is the least surface area way to enclose and separate two given volumes in $\mathbf{R}^3$ [23], that the equivalent of up one dimension minimizes the analogue of surface area in $\mathbf{R}^4$ [41], and that higher dimensional versions minimize higher dimensional equivalents of surface area when the two volumes to be enclosed are disparate from one another in size. The popularly accepted conjecture is that the standard double bubble is the least surface area way to enclose and separate two arbitrary volumes in Euclidean spaces of arbitrary dimension. We conjecture that standard double bubbles (after being subjected to a subtle redefiniton) give the least surface area way to enclose and separate two given volumes in constant curvature space forms as well. (In Chapter 4 we give a partial proof that this is indeed the case in three dimensions.) We will present the necessary background specific to double bubbles in Chapter 3.

There are some basic features of the double bubble problem that are worth keeping in mind when reading the current chapter however. For constant curvature spaces, one of the key differences between the double bubble problem (and all problems with a higher number of volume constraits) and the single bubble problem is that when there are multi-volume constraints each region in a perimeter minimizing bubble cluster might be disconnected into a number of fragments. In an arbitrary geometric space, a minimizing single bubble might disconnect as well (see Figure 2.2). One way understand why single bubbles in constant curvature spaces do not disconnect is the regularity theorem of Taylor that we discuss in subsection 1.3. This theorem gives explicit conditions on the singular sets present in perimeter minimizing bubble clusters. In particular, in three dimensions, the

boundary surfaces for an area-minimizing bubble cluster always meet at one hundred and twenty degrees along analytic curves. Given a disconnected candidate solution to the single bubble problem, we can use a series of local isometries to move one of the components so that it touches another at a point. *Since this transformation preserves area and perimeter and the final configuration violates regularity, neither the final configuration nor its preimage can be actually be minimizing*. The geometry of the double bubble problem is considerably more complicated than the geometry of the single bubble problem. It is not true, for example, that we can simply translate components of a non-standard double bubble to bring about a violation of regularity (see Figure 2.3 for a simple counter example).

Each of the sections of this chapter and the next requires a good-sized amount of 'modern' geometry and analysis. Many of the ideas (symmetrization, instability, etc.) are straightforward and the reader should be able to get a fairly clear understanding of the theorems without undo difficulty. However, in the hopes that readers will also be able to familarize themselves with the *proofs* of these known results (which I am not squeamish about presenting more or less in full here, to what I hope will be the greater glory of their originators), the standard definitions and results of geometry and analysis needed for the full appreciation of the results of this chapter and its sequel are presented in the appendix.

2.1 Constant Curvature Space Forms

Charter We should understand what *constant curvature space forms* are and some of their basic properties, including their symmetries. We should understand why there are in some sense only three simply connected spaces with constant curvature.

Three of the most useful geometric spaces in existence are Euclidean space, spherical space, and hyperbolic space (also sometimes called 'Lobachevski' space). These spaces are characterized by,

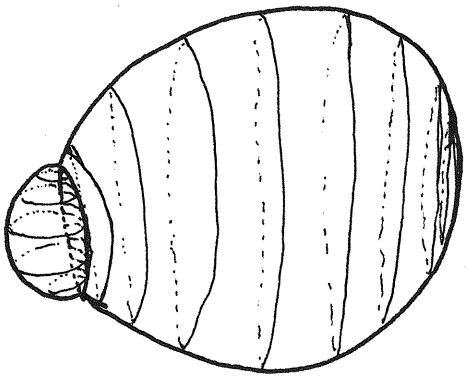


Figure 2.1: Standard Double Bubble

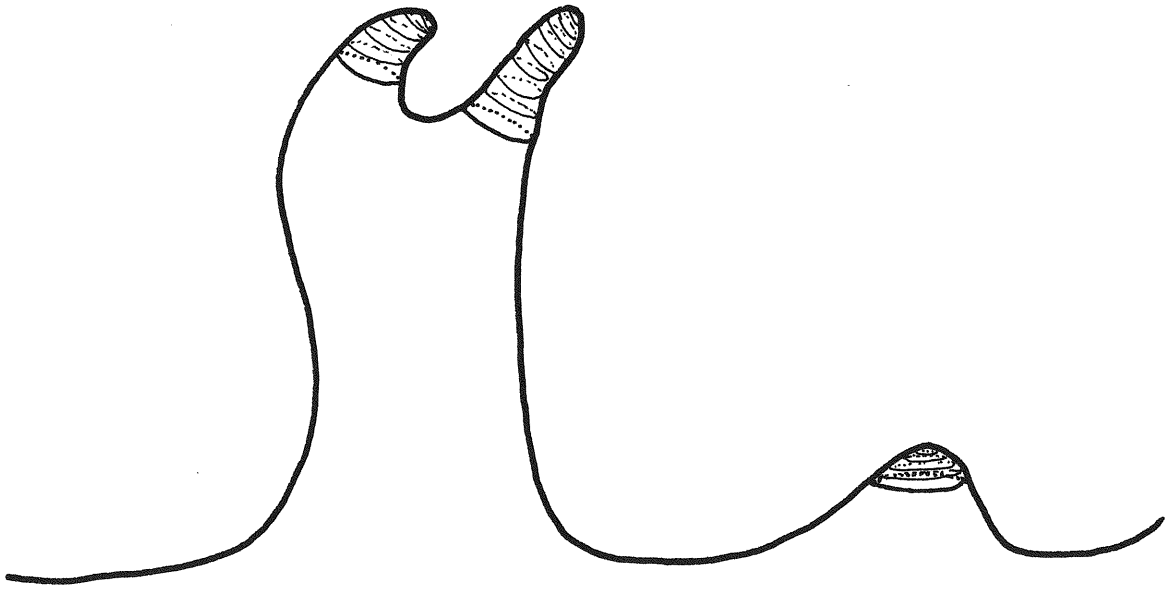


Figure 2.2: Disconnected Single Bubble Minimizer

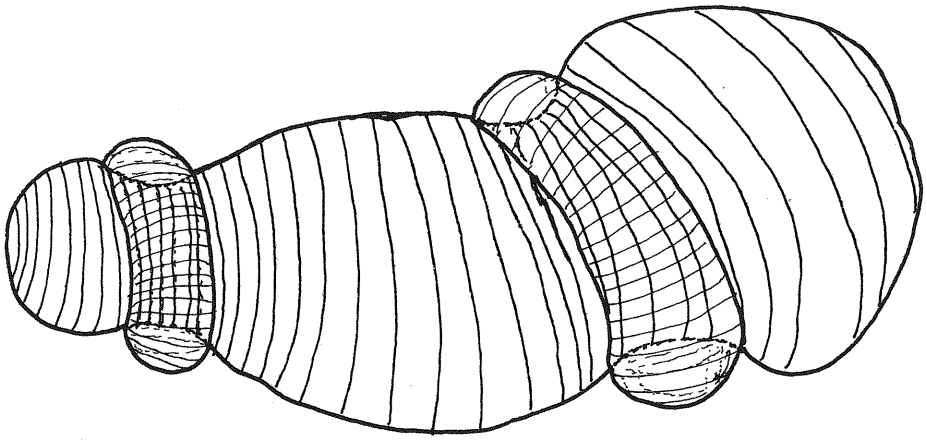


Figure 2.3: Non Standard Double Bubble

among other things, having *constant curvature* of zero, one, and negative one, respectively. By curvature we mean here the sectional curvature, which for the time being we will think of quite loosely as being 'a metric property of two-dimensional subspaces of a given space'. (To speak precisely, sectional curvature is a property of two dimensional subspaces of a tangent space, that is, a local property; we will come to this below. We won't reason in a circle like this in proofs, but in making a first approach to the subject of sectional curvature for constant curvature space forms, it wouldn't be too bad to think first about the curvature properties of subspaces, or, more accurately, submanifolds – such as lines, planes, meridians, and great circles for example – and then make the switch to the correct local formulation when we need it.)

We might mention at this point that there are very classical ways of characterizing these spaces that do not mention curvature. Non-Euclidean geometries were first developed at about the same time that modern geometry had its origins, and were studied using an amalgam of techniques. There are many books on non-Euclidean geometry, of which [3] seems to be very good for both mathematical and historical context.

Here we will work from the basic modern definitions, and we start with Euclidean space $\mathbf{R}^n$.

Definition 2.1. We define *n-dimensional Euclidean space* $\mathbf{R}^n$ to be the real vector space made up of

n-tuples of real numbers, such as $x = (x_1, x_2, ..., x_n)$, with addition and scalar multiplication defined element-wise, and with distance between points defined by

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}.$$

We understand $\mathbf{R}^n$ to carry the metric topology and standard smooth structure (cf. the appendices).

2.1.1 A detailed look at Euclidean space $\mathbf{R}^n$

In order to get anywhere in geometry we need to understand Euclidean space very well. The following paragraphs specialize the general treatment of topology and geometry found in the appendix to review the standard facts about Euclidean space.

1. We have *curves*, $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))$, where t lies in some interval of real numbers. Curves are frequently taken to be *smooth* in the sense that each of the component functions x_k is smooth, *i.e.* infinitely differentiable.

2. From this we get an extended (beyond the calculus 1 version) notion of differentiation, namely that of the *tangent vector to a curve*, written as $x'(t)$, and obtained by differentiating a curve component-wise, so that

$$x'(t) = (x_1'(t), x_2'(t), \ldots, x_n'(t)).$$

3. We also have maps between sets in $\mathbf{R}^n$ and $\mathbf{R}^m$, such as

$$x(t_1, \ldots, t_n) = (x_1(t_1, \ldots, t_n), x_2(t_1, \ldots, t_n), \ldots, x_m(t_1, \ldots, t_n))$$

and these maps also have derivatives, this time requiring matrices to represent (so we have for example, the Jacobian, the Hessian, etc.). The maps are said to be of class C^k if the first k of these matrices are non-singular.

4. We look at the set of curves through a point x and say that $\mathcal{C}_1 \sim_x \mathcal{C}_2$ if their tangent vectors are the same at x. If we identify curves that $\sim_x$ one another, the space of curves through the point x takes on a new identity, namely that of *the tangent space to* $\mathbf{R}^n$ *at* x, denoted $T_x\mathbf{R}^n$. So we have tangent spaces, and lots of linear algebra. (This definition of the tangent space varies slightly from the one given in the appendix, but as discussed in Remark C.3, the two definitions are really the same. There we show that the tangent space to $\mathbf{R}^n$ at each point is an isomorphic copy of $\mathbf{R}^n$.) The collection of all the tangent spaces is the *tangent bundle*, which we denote by $T\mathbf{R}^n$.

5. We have the Euclidean metric

$$dx_1^2 + dx_2^2 + \ldots + dx_n^2$$

defined on tangent spaces. The meaning of this is as follows: if $\langle \cdot, \cdot \rangle$ denotes the standard inner product of vectors in $\mathbf{R}^n$ (that is, $\langle \cdot, \cdot \rangle : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}, \langle v, w \rangle = \sum_{j=1}^n v_j w_j$) and if $\iota_p : T_p\mathbf{R}^n \to \mathbf{R}^n$ denotes the isomorphism from $T_p\mathbf{R}^n$ to $\mathbf{R}^n$, then *standard Riemannian metric on* $\mathbf{R}^n$ (the Euclidean metric) is defined by

$$T_p\mathbf{R}^n \times T_p\mathbf{R}^n \ni \xi, \eta \mapsto \langle \iota_p(\xi), \iota_p(\eta) \rangle.$$

We often omit the 'ι_p's. At each point, the metric can be written in local coordinates, which fix a basis for the tangent space. In terms of this basis, the metric becomes

$$\xi, \eta \mapsto \xi^T (g_{ij}) \eta,$$

where $g_{ij} = \delta_{ij}$. That is, at each point, the Euclidean metric can be represented by the identity matrix.

And, the point of all this? Broadly speaking, we understand curvature to be a smooth function (actually a quadratic form) on the tangent spaces of a manifold. To understand what this means,

we have to understand how to reduce questions about smoothness in a given geometric space (*i.e.* a manifold) to questions about smoothness in $\mathbf{R}^n$, and so we have to be comfortable doing the computations in $\mathbf{R}^n$, where the idea is that everything makes sense and is easy (or, as easy as its going to get).

2.1.2 Spherical and hyperbolic spaces

We define spherical and hyperbolic spaces as submanifolds of Euclidean space.

Definition 2.2. We define $(n-1)$-*dimensional spherical space* $\mathbf{S}^{n-1}$ to be the set of points a distance one from the origin in $\mathbf{R}^n$, carrying the (metric) topology induced by the metric topology on $\mathbf{R}^n$ and the smooth structure induced by the standard smooth structure on $\mathbf{R}^n$.

Definition 2.3. We define $(n-1)$-*dimensional hyperbolic space* $\mathbf{H}^{n-1}$ to be one sheet of a hyperboloid of two sheets

$$x_1^2 + x_2^2 + \ldots + x_{n-1}^2 - x_n^2 = -1 \tag{2.1}$$

(so, this, wherever $x_n > 0$) in the pseudo-Euclidean space $\mathbf{R}^{n-1,1}$ which has the same point-set as Euclidean space but is given the pseudo-Euclidean metric

$$dx_1^2 + dx_2^2 + \ldots + dx_{n-1}^2 - dx_n^2. \tag{2.2}$$

(The pseudo-Euclidean metric here is defined along exactly the same lines as the Euclidean metric above.) We understand $\mathbf{H}^{n-1}$ to carry the same topology and smooth structure as $\mathbf{R}^n$.

Remark 2.1. We can change the geometrical properties of spherical and hyperbolic space slightly by purturbing the above definitions; keeping the rest of the definitions the same, we put $\mathbf{S}^{n-1}(\rho)$ equal to the sphere of radius ρ, and $\mathbf{H}^{n-1}(\rho)$ equal to one sheet of the hyperboloid defined by the equation

$$x_1^2 + x_2^2 + \ldots + x_{n-1}^2 - x_n^2 = -\rho^2.$$

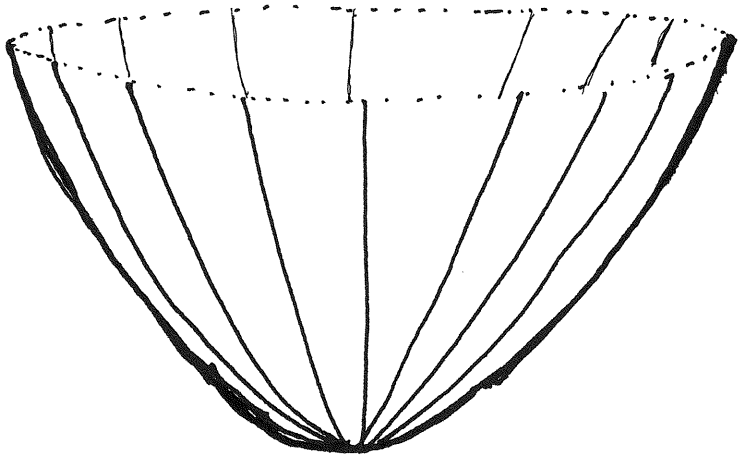


Figure 2.4: One sheet of the hyperboloid of two sheets

Visualizing higher dimensions and weird spaces. With all of this tricky abstraction at such an early stage in the game, it seems desirable to take a little time for reflection. The definitions above are, after all, supposed to be intuitive, in one way or another. The key is to figure out a way to actually think about everything geometrically. The basis for geometrical thinking is visualization and that is always lots of fun (one can do this lying on the couch with closed eyes). So for starters, how do we make sense of $\mathbf{R}^4$?

I like to think of this is as space-time – not anything tricky, just nice Cartesian space going on forever in all directions, and nice Galilean time (for lack of a better word) attached to each point in space. So, each point has a clock attached to it, say, and these clocks are syncronized. I like to use the fact that $\mathbf{R}$ is homeomorphic to an open interval to make everything happen in one infinitely long day. Right now its noon, everywhere, but in a few minutes, it will be 12:03. Time's running out! but don't worry, it will take forever to get to 6pm (the scheduled time for our infinite day comes to an end). On the other hand, 1:00 is coming up fairly soon, so we should be getting on with business.

More dimensions are easy: add more hands to the clock – say a hand indicating temperature. Lets say that temperature ranges from an extremely chilly -90 degrees to an unendurably hot 120000. Right now, its about 80 degrees in Sarasota, which is nicely described by – not to be self-centered

or anything – $(0, 0, 0, 1200, 80)$. Williamstown Massachusetts is located in $\mathbf{R}^5$ at $(1, 0, 0, 1200, 60)$. On the far side of the moon it is $(0, 15, 0, 1200, -60)$. Perpendicular in space to the plane determined by Williamstown, the moon, and Sarasota we find a point in the upper atmosphere of earth at $(0, 0, 1, 1200, -35)$.

Lets make things a little easier notationally and relabel the time and temperature axes so that 12:00 is 0, and 80 degrees is zero. Then Sarasota is at $(0, 0, 0, 0, 0)$, Williamstown is at $(1, 0, 0, 0, -.12)$, the far side of the moon is at $(0, 15, 0, 0, -3)$ and that special place in the upper atmosphere is $(0, 0, 1, 0, -1.5)$.

Ok, there. Now lets think about things geometrically. The most obvious thing is that Williamstown is pretty close to the north pole $(1, 0, 0, 0, 0)$ of the unit sphere in $\mathbf{R}^5$. The moon is not doing so well in this regard. The upper atmosphere is close to the sphere in most dimensions, but it would have to warm up a fair amount to make it. You might think about this as being sort of like the point $(0, 1, -1.5)$ in three dimensions – which lies due west of the equator of the unit sphere. If the unit sphere in three dimensions was being modeled by the surface of a pool ball, that special place way up in the sky might be stuck to the point on the edge of an 8-ball sitting between your and the corner pocket. The far side of the moon might show up on the tip of your nose at $(15, 0, -3)$, disturbing your concentration as you think about the shot. (If you are wondering how I chose the points in three space, the answer is that I projected onto the second, third, and fifth coordinates. I could have chosen a different model (a very loose sort of 'analogy') by projecting onto three different coordinates. On the other hand what you're probably wondering is how I chose the setting, and in that case you only get a hint – its not really noon in Sarasota.)

Anyway, fun and games aside, we have more exploring to do. We note that there are lots of different lower dimensional spheres. Everywhere the same distance from Sarasota as Sarasota is from Williamstown in the three-dimensional space given by the first three dimensions of $\mathbf{R}^5$ is a *two-sphere*, per usual. Everywhere that is this same distance from that special point in the upper

atmosphere is another two-sphere (and Sarasota is a point on its surface). If I worked on it a little, I could build a standard (2-)double bubble. Now, things get slightly more complicated as we note that the set of points every*when* a distance 1 from now (so, in terms of our original coordinate system, *everywhere* at 11:00 am together with *everywhere* at 1:00 pm) contains within it a set of points on the unit *three-sphere* – only two, actually – the point *here an hour ago* and the point *here an hour into the future*. These points are antipodal, being the poles of the sphere along the time axis. To interpolate between these points (*i.e.* to travel from one to the other and stay inside the sphere) we have to move just right. Let's say we were here an hour ago (not an unsafe assumption) and that we want to be here an hour from now (again, for most of us, this is a fair thing to say). What should we do? Should we sit around and wait? (Some people nod). Not if we want to traverse the three-sphere! Sitting here would take us through the center of the three-sphere, which would be fairly boring*. We have to travel to someplace a unit distance away in space in one hour and then get back in another hour. This means we are going to have to go fast. How can we get to Williamstown in an hour? Not much chance of that happening! Where can we go? Up, I suppose. But, thinking about this from a long-term perspective, that could be risky. The exercise would be better carried out starting in space, where we could designating some point clear of debris as (0,0,0,-1) and then come up on the point fast, and pass it, then gradually slow, turn, go back, and finally pass quickly through the point again at the designated time with no risk of bodily injury. Space also has the benefit of our being able to move in all directions, so we could visit any point on the two-sphere of directions in three-space that makes up the equator of the three-sphere.

So that is the three-sphere. It only a little trickier to build a standard 3-double bubble (which is made out of pieces of three-spheres meeting one another at 120 degrees), but it might take a long time to sort out the details if one wanted to go from scratch – so I'll leave that as a visualization exercise for the reader, should he or she wish to wrack his or her brain. There will be a hint at the

*Funny, am I not? – but in truth, the joke lacks content.

end of this paragraph for those who want the exercise to go more easily.

By a similar construction, everywhere-when-temperature that interpolates at the right rate between points that are one Sarasota-to-Williamstown-length off, one hour off, and one degree off makes up the unit 4-sphere we were talking about a few minutes ago. Materially speaking, this is much easier to traverse, since we can start at one pole in the time dimension, cool things down quickly and then more slowly so that we are one degree colder than we were an hour ago, and then reverse the operation over the course of the following hour. One could also easily combine cooling with driving, and move along a different trajectory from pole to pole. Again, we would be able to see much more of the 4-sphere if we did everything in outer space. However, it is nice to know that there is some higher-dimensional object that we can navigate in our lifetime (with little risk). *Indeed, such things actually show up all over the place when one is looking for them* – so the moral of the story *should* be that higher dimensions can be useful as well as entertaining.

As for hyperbolic space, I should say a few words about that here too. To understand this space in terms of our definition, we should first understand something about $\mathbf{R}^{n-1,1}$. To make things easy on ourselves at first we will restrict to the case that $n = 4$. We will think in a parallel way about the space $\mathbf{R}^{3,1}$ and the space $\mathbf{R}^4$. Indeed, $\mathbf{R}^{3,1}$ is often interpreted as another model for space-time – in this case, we have Cartesian space again, but Einsteinian or relativistic time [29]. Although we think of it as another model for space-time, we also need to keep the two different notions straight, so henceforth we will call $\mathbf{R}^{3,1}$ by its official name *Minkowski space*, The way Minkowski space models space-time is as follows: we have assume the speed of light c is constant (usually we chose this constant to be 1) and we think of vectors in the space as (x_1, x_2, x_3, ct). The pseudo-norm (Minkowski length) $|x|^2 = \sum_{i=1}^{3} x_i^2 - (ct)^2$ induced by the pseudo-Euclidean metric 2.2 gives us an immediate way to define the distance between points $x = (x_1, x_2, x_3, ct_1)$ and $y = (y_1, y_2, y_3, ct_2)$ in $\mathbf{R}^{3,1}$, namely $d(x, y) = |x - y| = \left(\sum_{i=1}^{3}(x_i - y_i)^2\right) - (ct_1 - ct_2)^2$; this is a global definition for distance, *i.e.* it works in much the same way as the Euclidean norm. Here however, distances

are allowed to be negative. The vectors with Minkowski length zero form a cone, since $|\cdot|^2$ is homogeneous. This cone is called the *light cone*, because points that move in straight lines away from the origin with the speed of light stay in the cone. It has two halves, the *forward light cone*, consisting of vectors with $t > 0$ and the *backward light cone*, consisting of vectors with $t < 0$ (similarly, there are forward light cones and backward light cones centered at each point in Minkowski space). All vectors in the light cone are called *light-like*. Vectors with positive Minkowski length are called *space-like* and vectors with negative Minkowski length are called *time-like*, where the names designate the coordinates that are dominant in the vector. The time-like vectors lie in region of $\mathbf{R}^{3,1}$ 'inside' the light cone. The physical interpretation is that particles can not travel faster than light, so tangent vectors must satisfy the inequality $|\frac{\partial x}{\partial t}|^2 = \sum_{i=1}^{3} \left(\frac{\partial x_i}{\partial t}\right)^2 - c \leq 0$. Continuing in the physical vein, we assume that only particles with zero mass have light-like tangent vectors, and particles with positive mass have time-like tangent vectors. The arclength $s = \int_a^b |\frac{\partial x}{\partial t}|^2 d\tau$ in the case of a particle moving with constant velocity becomes $\int_{t_1}^{t_2} \sqrt{c^2 - \sum_{i=1}^{3} \left(\frac{\partial x_i}{\partial t}\right)^2} d\tau = \int_{t_1}^{t_2} c\sqrt{1 - \frac{v^2}{c^2}} d\tau = ct\sqrt{1 - \frac{v^2}{c^2}}\Big|_{t_1}^{t_2}$ and s/c is interpreted as the amount of time the particle takes to travel between two points in space as seen by a fixed observer. If $v \neq 0$, this measurement is the amount of time a clock traveling with the particle takes, scaled by a factor of $\sqrt{1 - \frac{v^2}{c^2}}$. This is a basic property of special relativity; others follow similarly. Three-dimensional hyperbolic space consists of (half of) the vectors in $\mathbf{R}^{3,1}$ with Minkowski length -1. This set is asymptotic to the light cone as t increases. One interpretation might be that hyperbolic space consists of all of the positions of a set of particles that explode out of the origin in $\mathbf{R}^3$ at time $t = 1$, and their positions are then plotted with respect to time. The key idea is that the particles speed up as they fly outwards in space at a certain rate, and over time they approach the speed of light. They also spread out as they fly, so that the total mass along each direction goes to zero (assuming the particles were infinitesimal to begin with).

The basic idea is that hyperbolic space spreads out from any given point very quickly, and that it essentially the same as $\mathbf{R}^n$ in all other respects.

Figure 2.5 shows some of the important features of the hyperbolic plane in the Poincaré disk model (which does not look so terribly different from the hyperboloid - but it is different enough to warrent further investigation. Unfortunately we do not have time to show how to map between these representations in this thesis; we direct the interested reader to [16].

The first important feature of this model (which was also used to draw Figure 1.3) is that the boundary of the disk is at infinity, so we are looking at a truly infinite space, although it is captured in finite area. The other important feature of the model is that angles are correct. The circles in the interior of the space *are* circles (*i.e.* boundaries of metric balls in $\mathbf{H}^2$) although their centers are not always in the center of the circles as they are drawn here. The straight lines and pieces of circles that meeting the boundary of the disk at 90 degrees are geodesics. Geodesics that look like pieces of circles have an additional special name – *hypocircles*; we extend this term to all geodesics, and in the future will use these two terms more or less synonymously. If we changed coordinates, so that we were looking down straight onto what is currently the apex of a curved hypocircle, it would look straight – so the only time a distiction can be made is when we work with a given representation of the plane. Circles tangent to the boundary of the disk in the figure are not actually circles in the hyperbolic plane but *horocircles*; they do not have well-defined radii. These can not be unstuck from infinity by a coordinate change. A related fact that sheds some light on the situation is that two geodesics that appear to meet at infinity never actually intersect, but become arbitrarily close to one another.

The picture is similar in three dimensions: things that look like pieces circles in this picture will become a pieces of spheres, the straight lines will become a flat disks, and the Poincaré model itself will become a solid ball. The new vocabulary that goes along with the construction just replaces 'sphere' everwhere 'circle' was found previously. Geodesics are again straight lines and pieces of circles that meet the boundary of the ball at 90 degrees. If we add more dimensions, the definitions and terminology develops along the same lines as it did here.

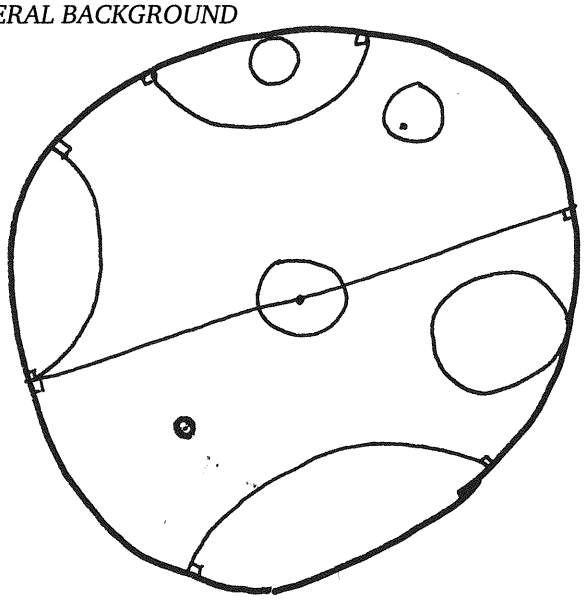


Figure 2.5: Horocircles, hypocircles, circles.

One definition that really needs to be understood clearly in a number of places later on in the thesis is that of a *hyperplane*. In $\mathbf{R}^n$, this is just an embedded isometric copy of $\mathbf{R}^{n-1}$. In Euclidean space, hyperplanes cut the space in two pieces: the line divides the plane, the plane divides space, space divides space-time (for example, everywhere-now is a hyperplane) and so on. Embedded isometric copies of Euclidean spaces with dimension less than n do not separate $\mathbf{R}^n$ – a line does not separate space, a plane does not separate space-time. This is seen to be true because a line will *never* seperate space and a *line-always* is a two-plane in space-time, whereas *space-always* is space-time. In other words, after rotating, at any given point in time a two-plane in space-time is just a line in space, so it doesn't ever space. Furthermore, by definition, a line-always goes on for all time, so it doesn't separate time either, and hence does not separate space-time. And similarly for higher dimensions. In spheres and in hyperbolic space, similar ideas apply - so, in the n-sphere, an equatorial $(n-1)$-sphere is a *hypersphere*; in hyperbolic space, a (hyperbolic) $(n-1)$-plane is a *hyper-hyperbolic plane*.). With tori, and other more tricky geometries, it is a little different, since here hyperplanes do not always cut the space in two. Hyperplanes are just one example of a broader category, namely *hypersurfaces*, which are embedded submanifolds of one dimension lower than the ambient manifold.

We will say one last thing about spheres. This is that $\mathbf{S}^n$ is in general not only a hypersurface in $\mathbf{R}^{n+1}$, but a *hypersurface of revolution*. That $\mathbf{S}^2$ is a surface of revolution probably does not come as a major revelation. The fact that $\mathbf{S}^3$ is a hypersurface of revolution may take a little explaining. First, lets walk through an illustrative example of rotation *in* $\mathbf{S}^3$. From this example we will get the ideas we need to define hypersurfaces of revolution.

There are two steps to the illustration. The first step should be familiar. Start with a half circle C with radius r in the upper half plane $\mathbf{R}^2_+$, sitting on the x-axis. We revolve C around the x-axis and we call the set of points it hits '$\mathbf{S}^2$'. Now the second step: we will take the upper half of $\mathbf{S}^2$ (let's call it C') and perform similar operations to generate $\mathbf{S}^3$. We will regard C' as sitting on $\mathbf{R}^2$ inside of $\mathbf{R}^3_+$ which is itself sitting inside of $\mathbf{R}^4$. But we also keep in mind the idea of C' as sitting inside $\mathbf{S}^3$ (specificall, we will think of C' as half of an equatorial two-sphere). We now rotate C' by interpolating from each of its points in the manner described previously. What we see in $\mathbf{R}^4$ are images of C', getting smaller and earlier, all centered on the same point p in $\mathbf{R}^2$ (*i.e.* the same point in space). Eventually these images shrink all the way down to p (the time then is exactly $-r$), reemerge *reflected across* $\mathbf{R}^2$, and begin to grow and get later again. When the the image in $\mathbf{R}^3$ is equal to the reflection of C' across $\mathbf{R}^2$, the time is zero again. Now the images shrink down to p again, getting later all the while, and when they hit p this time, the clock reads r. To make the final turn, the images grow again, and when they get back to C', the clock is back to zero; this completes the illustration.

At this point it would be natural to ask what we have actually done. This is a question perhaps as much due to the fact that spheres are so fearfully symmetric as it is to the fact that we are working in four dimensions. But we also note is that there are some big differences between the lower-dimensional case (the first step of the example) and the higher-dimensional case (the second step). The main difference that concerns us here is that none of the points on C' were fixed in the second step (whereas the two endpoints of C were fixed in the first step). Something that *is* fixed (in space-

time) is the location of the point p, since its distance from itself is zero (and distance-from-p was the parameter we used to determine how far out in the time direction we had to go). If we now examine a surface with less symmetry – the top half of a cylinder for example – we can define a rotation about the (spatial) axis of symmetry of the full cylinder by treating each half-circle segment of the half-cylinder in the same way we treated the meridinal segments of C'. Taking what we learn from this example, we can define the *rotation into the fourth dimension* for a rotationally symmetric surface S by treating each of the circles that make up S as if it was a meridian in the two-sphere sitting inside $\mathbf{S}^3$. In particular, we can take $\mathbf{S}^2$ and viewing it as a subset of $\mathbf{R}^3$, think of it as being a surface of revolution about some fixed axis ℓ. If we now rotate into the fourth dimension as described, we trace out a copy of the three sphere. As in the first step of the illustration, two points are fixed for the entire rotation, namely the points where ℓ intersects $\mathbf{S}^2$. Since the time at these points is always zero, it might be fitting to call them East and West poles of the three sphere, since they lie on line that is perpendicular to the time axis (which we usually say runs north and south).

Now, the hint for how to construct a standard 3-double bubble should make it very easy: start with a standard 2-double bubble and revolve it around its spatial axis of symmetry.

2.1.3 Three constant curvature space forms.

Given the intuitive definitions for Euclidean, spherical, and hyperbolic space we saw above, it may seem reasonable that they should have constant sectional curvature. (We're still thinking of this in very loose terms; the point being that we know intuitively which two dimensional subspaces of $\mathbf{R}^n$, $\mathbf{S}^n$, and $\mathbf{H}^n$ we will be looking at. They are the hyperplanes, equatorial spheres, or hyper-hyperplanes.)

A quick proof that the curvatures are constant and 'signed', so that Euclidean space has zero curvature, $\mathbf{S}^n$ has constant positive curvature, and $\mathbf{H}^n$ has constant negative curvature, will be given in this subsection. A simple corollary shows that by scaling the ambient $\mathbf{R}^n$ we can normalize the

curvature of $\mathbf{S}^{n-1}$ to be 1 and by scaling the ambient $\mathbf{R}^{n-1,1}$ we can normalize the curvature of $\mathbf{H}^{n-1}$ to be -1.

This suggests that there are in some sense only three constant curvature space forms. The exact sense in which this is true is that *up to scaling these are the only simply connected spaces with constant curvature, i.e.* that $\mathbf{R}^n$, $\mathbf{S}^n$, and $\mathbf{H}^n$ are the only constant curvature spaces in which all loops are contractible to a point, *i.e.* they are the only constant curvature space forms. (The fact that these spaces are all simply connected is easy to see, if one is good at visualizing things in many dimensions.) The *proof* that these are the only simply connected spaces with constant curvature is not particularly easy, and we omit it (see [12] or [59] for the details).

We henceforth give the constant curvature space forms a *new* collective name, $\mathcal{R}^n$, where the $\mathcal{R}$ refers to the (Riemann) curvature (tensor), which is all that we need to specify in order to fix a particular constant curvature space form.

The Manifold Curvatures. There is not much left to do but to plunge in and recall the official definitions. The various curvatures of Riemannian geometry are as incestuous and as plentiful could possibly be desired, so there could be a lot for us to recall if we liked; we will however try to be moderate. *It may be helpful to refer to the appendix on Geometry Preliminaries when reading this section.*

Definition 2.4. We define the *Riemann curvature operator* R of a Riemannian manifold $M = (X, \langle\cdot,\cdot\rangle)$ with Levi-Civita connection ∇^{LC} and commutator $[\cdot,\cdot]$ to be the map from $\mathrm{vect}(M) \times \mathrm{vect}(M) \times \mathrm{vect}(M)$ to $\mathrm{vect}(M)$ determined by the following action on vector fields:

$$R(\mathcal{U},\mathcal{V})\mathcal{W} = \nabla^{\mathrm{LC}}(\mathcal{V}, \nabla^{\mathrm{LC}}(\mathcal{U},\mathcal{W})) - \nabla^{\mathrm{LC}}(\mathcal{U}, \nabla^{\mathrm{LC}}(\mathcal{V},\mathcal{W})) - \nabla^{\mathrm{LC}}([\mathcal{V},\mathcal{U}],\mathcal{W}).$$

Remark 2.2. Note that R is local in the sense that $(R(\mathcal{U},\mathcal{V})\mathcal{W})|_p$ depends only on $\mathcal{U}|_p, \mathcal{V}|_p$ and $\mathcal{W}|_p$.

To be precise, in terms of local coordinates $\{u_i\}_{i=1}^n$, we have

$$R\left(\frac{d}{du_j},\frac{d}{du_k}\right)\frac{d}{du_i}=\sum_{l=1}^{n}R^l_{ijk}\frac{d}{du_l} \tag{2.3}$$

where the 'R^l_{ijk}'s are functions from the manifold to the reals defined by

$$R^l_{ijk}=\frac{d}{du_k}\Gamma^l_{ik}-\frac{d}{du_j}\Gamma^l_{ik}+\sum_{r=1}^{n}\Gamma^r_{ij}\Gamma^l_{rk}-\Gamma^r_{ik}\Gamma^l_{rj}. \tag{2.4}$$

The proof is just by expanding $R\left(\frac{d}{du_j},\frac{d}{du_k}\right)\frac{d}{du_i}$ in terms of its definition and the definition of a connection (DefinitionC.11). From the definition of the Riemann curvature operator, this is

$$\nabla^{\mathrm{LC}}\left(\frac{d}{du_k},\nabla^{\mathrm{LC}}\left(\frac{d}{du_j},\frac{d}{du_i}\right)\right)-\nabla^{\mathrm{LC}}\left(\frac{d}{du_j},\nabla^{\mathrm{LC}}\left(\frac{d}{du_k},\frac{d}{du_i}\right)\right)-\nabla^{\mathrm{LC}}\left(\left[\frac{d}{du_k},\frac{d}{du_j}\right],\frac{d}{du_i}\right)$$

(but $\left[\frac{d}{du_k},\frac{d}{du_j}\right]$ is zero; we apply the definiton of the connection coefficients (Equation C.5) to the first two terms)

$$=\nabla^{\mathrm{LC}}\left(\frac{d}{du_k},\sum_{m=1}^{n}\Gamma^m_{ij}\frac{d}{du_m}\right)-\nabla^{\mathrm{LC}}\left(\frac{d}{du_j},\sum_{m=1}^{n}\Gamma^m_{ij}\frac{d}{du_m}\right)$$

(by additivity in the second component for derivations (Equation C.2) this is)

$$=\sum_{m=1}^{n}\left(\nabla^{\mathrm{LC}}\left(\frac{d}{du_k},\Gamma^m_{ij}\frac{d}{du_m}\right)\right)-\sum_{m=1}^{n}\left(\nabla^{\mathrm{LC}}\left(\frac{d}{du_j},\Gamma^m_{ij}\frac{d}{du_m}\right)\right)$$

(combining the summations)

$$=\sum_{m=1}^{n}\left(\nabla^{\mathrm{LC}}\left(\frac{d}{du_k},\Gamma^m_{ij}\frac{d}{du_m}\right)-\nabla^{\mathrm{LC}}\left(\frac{d}{du_j},\Gamma^m_{ij}\frac{d}{du_m}\right)\right)$$

(and by the Leibnitz rule for connections (Equation C.3) and the definition of the connection coefficients again this is)

$$=\sum_{m=1}^{n}\left(\left(\left(\frac{d}{du_k}\Gamma^m_{ij}\right)\frac{d}{du_m}+\Gamma^m_{ij}\sum_{r=1}^{n}\Gamma^r_{mk}\frac{d}{du_r}\right)-\left(\frac{d}{du_k}\Gamma^m_{ik}\right)\frac{d}{du_m}+\Gamma^m_{ik}\sum_{r=1}^{n}\Gamma^r_{mj}\frac{d}{du_r}\right)$$

(which, upon collecting like terms, shows that the scalar coefficients are the 'R^l_{ijk}'s defined by Equation 2.4.)

Definition 2.5. With M as above, we define the *Riemann curvature tensor* to be the map

$$(\mathrm{vect}(M))^4\times M\ni(\mathcal{U},\mathcal{V},\mathcal{W},\mathcal{X},p)\mapsto\langle R(\mathcal{U},\mathcal{V})\mathcal{W},\mathcal{X}\rangle_p\in\mathbf{R}. \tag{2.5}$$

Remark 2.3. If we have local coordinates, the Riemann curvature tensor is locally computable in terms of the connection coefficients (and so, in terms of the metric). Using the local definition the Riemann curvature operator (Equations 2.3 and 2.4), we have

$$\left\langle R\left(\frac{d}{du_j}, \frac{d}{du_k}\right)\frac{d}{du_i}, \frac{d}{du_m}\right\rangle = R^l_{ijk} = \frac{d}{du_k}\Gamma^l_{ik} - \frac{d}{du_j}\Gamma^l_{ik} + \sum_{r=1}^{n}\Gamma^r_{ij}\Gamma^l_{rk} - \Gamma^r_{ik}\Gamma^l_{rj}. \tag{2.6}$$

In particular, a computation in the appendix (Corollary C.1) shows that the connection coefficients all vanish in the case of $\mathbf{R}^n$, which implies that the Riemann curvature tensor associated with $\mathbf{R}^n$ is identically zero.

Definition 2.6. With M as above, let $p \in M$ and let ξ, η be vectors in T_pM. Define

$$\kappa(\xi, \eta) = \langle R(\xi, \eta)\xi, \eta\rangle,$$

and define

$$\kappa_1(\xi, \eta) = \langle \xi, \xi\rangle\langle \eta, \eta\rangle - \langle \xi, \eta\rangle^2$$

Finally, let

$$K(\xi, \eta) = \frac{\kappa(\xi, \eta)}{\kappa_1(\xi, \eta)}.$$

K is the *sectional curvature at p of the plane determined by ξ and η.*

Remark 2.4. The idea is that in $\mathcal{R}^n$, K depends trivially on ξ and η, and p. For example, in Remark 2.6 we saw that the Riemann curvature tensor on $\mathbf{R}^n$ is identically zero; this implies that the sectional curvature of $\mathbf{R}^n$ is identically zero.

$\mathbf{R}^n$, $\mathbf{S}^n$, and $\mathbf{H}^n$ have the right curvatures. We have already done the case of Euclidean space; as summarized in Remark 2.4, the sectional curvature of $\mathbf{R}^n$ is identically zero. We say that Euclidean space is *flat*.

Now, spheres. Using stereographic projection from the north pole,

$$\mathbf{S}^n(\rho) \setminus \text{south pole} \ni y = (y_1, \ldots, y_{n+1}) \mapsto u = (u_1, \ldots, u_n) \in \mathbf{R}^n \subset \mathbf{R}^{n+1}.$$

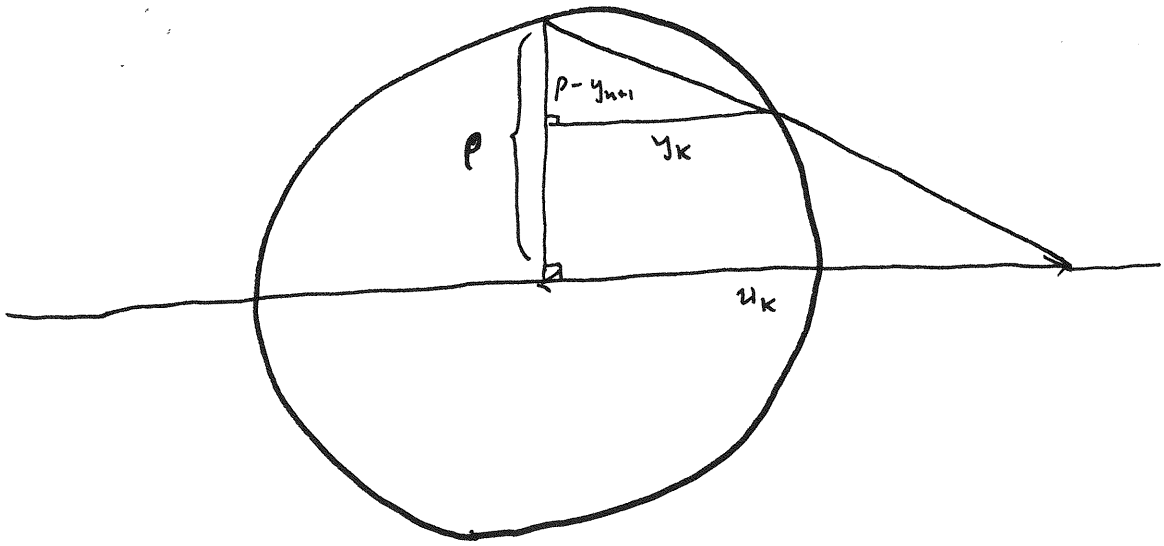


Figure 2.6: Stereographic projection

By similar triangles in Figure 2.6, we have

$$\frac{y_k}{\rho - y_{n+1}} = \frac{u_k}{\rho},$$

i.e.

$$y_k = \frac{u_k(\rho - y_{n+1})}{\rho},$$

hence

$$y_{n+1} = \frac{|u|^2 - \rho^2}{|u|^2 + \rho^2}\rho,$$

and

$$(y_1, \ldots, y_n) = \frac{2\rho^2}{\rho^2 + |u|^2}u.$$

We put $\bar{y} = (y_1, \ldots, y_n)$ and compute:

$$\begin{aligned} d\bar{y} &= 2\rho^2 \frac{(\rho^2 + |u|^2)du - 2u(u \cdot du)}{(\rho^2 + |u|^2)^2} \\ dy_{n+1} &= 4\rho^3 \frac{u \cdot du}{(\rho^2 + |u|^2)^2} \end{aligned}$$

and so

$$\begin{aligned} |d\bar{y}|^2 &= 4\rho^4 \frac{(\rho^2 + |u|^2)^2|du|^2 - 4\rho^2(u \cdot du)^2}{(\rho^2 + |u|^2)^4} \\ dy_{n+1}^2 &= 16\rho^6 \frac{(u \cdot du)^2}{(\rho^2 + |u|^2)^4}. \end{aligned}$$

Summing, we have

$$ds = \sum_{k=1}^{n+1} dy_k^2 = 4\rho^4 \frac{|du|^2}{(\rho^2 + |u^2|)^2},$$

i.e.

$$g_{jk} = \left\langle \frac{d}{du_j}, \frac{d}{du_k} \right\rangle = \frac{4\delta_{jk}}{(1 + |u|^2/\rho^2)^2}.$$

Using the formula for connection coefficients found in the appendix (Equation C.15), this implies that

$$\Gamma^i_{jk}(u) = \frac{-2}{\rho^2 + |u|^2}(\delta_{ik}u_j + \delta_{ji}u_k - \delta_{jk}u_i).$$

We now invite the reader to use the local formula for the Riemann curvature tensor (Equation 2.6) to show (by a long and nasty computation) that the sectional curvature of $\mathbf{S}^n(\rho)$ is ρ.

Finally, hyperbolic space. There are a couple of nice models for hyperbolic space: the disk model, and the upper half space model. It is sometimes helpful to be able to switch between the models.

If we start with a disk with the metric

$$ds = \frac{4|dz|^2}{(1 - |z|^2/\rho^2)^2},$$

then the sectional curvature will turn out to be $-1/\rho$. We show that in the case $\rho = 1$, the sectional curvature is -1. It is preferential to work in $\mathbf{R}^n_+$ with the metric

$$ds = \frac{|du|^2}{(u_n)^2},$$

i.e.,

$$g_{ij} = \frac{\delta_{ij}}{(u_n)^2}$$

which implies that

$$\Gamma^i_{ij} = -(u_n)^{-1}(\delta_{jn}\delta_{ik} + \delta_{kn}\delta_{ji} - \delta_{in}\delta_{jk}),$$

and so for $j, k \in \{1, \ldots, n-1\}$ we have

$$\nabla^{\mathrm{LC}}_{\mathbf{H}^n}\left(\frac{d}{du_j}, \frac{d}{du_k}\right) = -(u_n)^{-1}\delta_{jk}\frac{d}{du_n}$$

$$\begin{aligned}
\nabla^{\mathrm{LC}}_{\mathbf{H}^n}\left(\frac{d}{du_n},\frac{d}{du_k}\right) &= -(u_n)^{-1}\frac{d}{du_k}\\
\nabla^{\mathrm{LC}}_{\mathbf{H}^n}\left(\frac{d}{du_n},\frac{d}{du_n}\right) &= -(u_n)^{-1}\frac{d}{du_n}\\
\nabla^{\mathrm{LC}}_{\mathbf{H}^n}\left(\frac{d}{du_k},\nabla^{\mathrm{LC}}_{\mathbf{H}^n}\left(\frac{d}{du_j},\frac{d}{du_n}\right)\right) &= -(u_n)^{-1}\delta_{jk}\frac{d}{du_k}\\
\nabla^{\mathrm{LC}}_{\mathbf{H}^n}\left(\frac{d}{du_j},\nabla^{\mathrm{LC}}_{\mathbf{H}^n}\left(\frac{d}{du_k},\frac{d}{du_k}\right)\right) &= -(u_n)^{-1}\frac{d}{du_j}.
\end{aligned}$$

Therefore, when $j \neq k$ we have

$$R\left(\frac{d}{du_k},\frac{d}{du_j}\right)\frac{d}{du_k} = -(u_n)^{-2}\frac{d}{du_j}$$

which implies by the definition of sectional curvature that $K(\frac{d}{du_j},\frac{d}{du_j}) = -1$ and similarly

$$R\left(\frac{d}{du_n},\frac{d}{du_j}\right)\frac{d}{du_n} = -(u_n)^{-2}\frac{d}{du_j}$$

so that $K(\frac{d}{du_j},\frac{d}{du_n}) = -1$.

Geodesics and Metric Balls. This paragraph mainly reviews and specializes material found in the appendix. We also define spherical coordinates for constant curvature space forms.

The shortest paths between points are called geodesics. One can go through the derivations in some detail, but to summarize: in Euclidean space $\mathbf{R}^n$, geodesics are straight lines; in spherical space $\mathbf{S}^n$, geodesics are '$\mathbf{S}^1$'s with the same curvature as $\mathbf{S}^n$; and in hyperbolic space, geodesics are '$\mathbf{H}^1$'s with the same curvature as $\mathbf{H}^n$.

More generally, we have the following:

Definition 2.7. A submanifold $T \subset M$ is *totally geodesic* if every geodesic in M that has an open segment in T never leaves T.

Because the geodesics in $\mathcal{R}^n$ are easy, it is easy to see what the totally geodesic submanifolds are; they are the k-dimensional planes or spheres with the same curvature as $\mathcal{R}^n$.

For the sake of understanding the next several section(s) we need to know what metric balls (also interchangably called *disks*) in $\mathcal{R}^n$ look like. They are for the most part like balls or disks in Euclidean space. Disks on spheres are sometimes also called *spherical caps*, and they do not always look a lot like a disk in Euclidean space. Every round circle in $\mathbf{S}^2$ cuts off two spherical caps, so hemispheres as well as things both bigger and smaller are examples.

Round balls are associated with a very convenient coordinate system, namely, spherical coordinates. These are very easy to write down in $\mathcal{R}^n$ [7, page 39]. If $K > 0$, we take $(t, \xi) \in [0, \pi/\sqrt{K}) \times \mathbf{S}^{n-1}$, and if $K \leq 0$ we take $(t, \xi) \in [0, \infty) \times \mathbf{S}^{n-1}$; then there is some coordinate system with respect to which the metric can be written as

$$ds^2 = (dt)^2 + S_K^2(t)|d\xi|^2,$$

where $S_K^2(t)$ is the solution to the differential equation

$$y'' + Ky = 0,$$

with the initial conditions

$$S_K(0) = 0,\ S_K'(0) = 1.$$

If $K > 0$, the solution is $S_K(t) = (1/\sqrt{K}) \sin \sqrt{K}t$, if $K < 0$, the solution is $S_K(t) = (1/\sqrt{-K}) \sinh \sqrt{-K}t$ and if $K = 0$, the solution is t.

This is a nice way to summarize facts that can be determined by direct calculation, using the coordinates $x = p + t\xi$ in $\mathbf{R}^n$, $x = (\cos t/\rho)p + (\sin t/\rho)\xi$ in $\mathbf{S}^n(\rho)$, or (here we center on 0) $x = \rho \tanh(t/2\rho)\xi$ in $\mathbf{H}^n(\rho)$ and then finding the differentials.

2.2 Isoperimetric problem in constant curvature space forms

Charter We should understand the know results on the 'single bubble' or isoperimetric problem, which states that of regions with a fixed volume in constant curvature space forms, the ones with the least surface area are round balls.

In order to eliminate any potential confusion at such an early stage, and because it is at least somewhat relevant in the arguments which follow, the first note we'll make is that 'isoperimetric' (same-perimeter) is a fine, if somewhat obscure, name for the problem of finding the *least* perimeter needed to enclose a *fixed* area. The key idea to the simple argument that proves this is that maximal area increases monotonically with perimeter. This is obvious for Euclidean and hyperbolic space; for spheres this is only true for volumes that take up up to half of the volume of the space, but after that, we switch to think of the other smaller volume as the one being enclosed, so the same argument applies.

Here is the argument. Each perimeter P has a set of maximal-area objects $S(P)$ associated to it. Obviously all the elements of $S(P)$ must have the same area, which we'll call A. Change perspectives, and each area A has a set of perimeter-minimizing obects associated to it. We claim that this set is exactly $S(P)$. The proof is that if $p < P$ and there is some enclosure E with perimeter p and area A, so perimeter does not increase monotonically; this is a contradiction.

There are in fact a number of other equivalent formulations of the isoperimetric problem, but the one we want to keep in the forefront of our minds is the cannonical 'least perimeter for a given area' version. More generally, we will talk about the domains with the least surface area for a given volume, where we think in terms of the following formal definitions.

Definition 2.8. A *domain* in a manifold M is an open set whose boundary fits the needs at hand;

usually our proofs only require that the boundaries be embedded C^1 or C^2 submanifolds.

In defining domains, we need to pick something that works as a model for bubbles. In the later sections, when we work with more than one enclosed region, we will be allowing a lot of singular behavior on the boundaries, but at this stage we want regularity up front. This means, for example, that unlike Zenodorus and Pappus, we won't be thinking about how competative the Euclidean solids are with the sphere. In the much more general category we will be working with later, it is possible to show that the boundaries of competative single bubbles really are smooth, so our restrictive-seeming differentiability assumption will be vindicated.

Definition 2.9. The *volume* of a domain Ω in $\mathbf{R}^n$ is the Lebesgue measure of Ω. We write $\mu_n(\Omega)$ for the volume of a domain Ω. In $\mathcal{R}^n$ volume is defined using the analogue of Lebesgue measure (see the appendix on Analysis).

Remark 2.5. This is one of the reasons we require differentiability; we want to make the set of domains to be in the σ-algebra of Lebesgue measurable subsets of $\mathbf{R}^n$. (Or, for $(R)^n$, in the analogous σ-algebra.)

Definition 2.10. We define the *surface area* of Ω to be the derivative at zero with respect to ϵ of the volume of an ϵ-tube around Ω (an *ϵ-tube* Ω_ϵ is the set of points $p \in \mathcal{R}^n$ such that $d(p, \Omega) \leq \epsilon$). We write this in symbols as

$$\mu_{n-1}(\partial\Omega) = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mu_n(\Omega_\epsilon).$$

Note that μ_{n-1} as surface area is the same as μ_{n-1} as $(n-1)$-dimensional measure; Definition 2.10 gives a way to restrict n-dimensional measure to $(n-1)$-dimensional sets. However, as with the definition of volume, the definition of surface area relies on the assumption that our domains are bounded by differentiable curves. (We could have defined surface area with a slightly relaxed hypothesis; for example, *piecewise* differentiable would be ok.) Definition 2.10 is only the beginning of a sequence of very interesting things that can be done with tubes. See Gray [17] for some of the

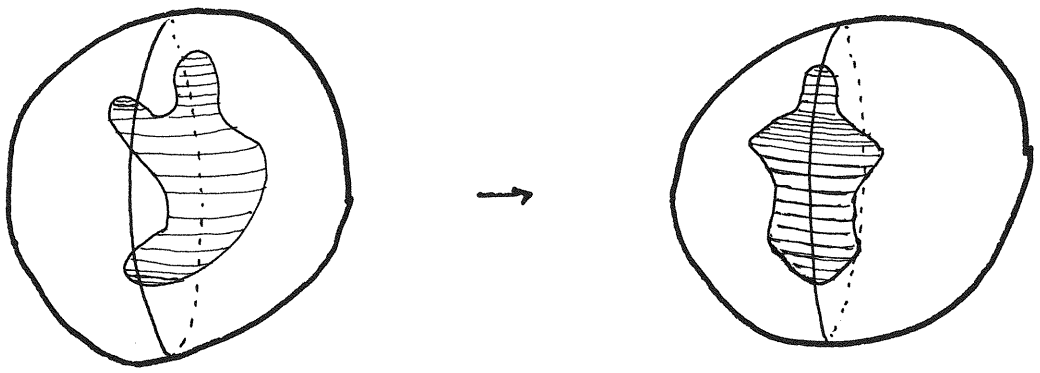


Figure 2.7: The Steiner symmetrization of a domain on the sphere.

most important details.

Theorem 2.1 (Isoperimetric property of metric balls). In constant curvature space forms, the domains with least surface area that enclose a given volume are the metric balls.

2.2.1 Proof of the isoperimetric property of round balls

We present a proof with a fairly geometric flavor based on an argument due to Figiel, Lindenstrauss, and Milman [52] which covers the spherical case. This proof was first encountered by the author in the text of Chavel [8] where it is modified to treat the Euclidean and hyperbolic cases. This restatement of Theorem 2.1 is based on that found in Chavel (*op. cit.*, page 283, Theorem 6.7) although I prefer in the notation of Figiel *et al.* and I reproduce their argument for the spherical case.

Restatement of Theorem 2.1. Let A be the closure of a domain in $\mathcal{R}^n$. For a given A, let B be a closed metric ball in $\mathcal{R}^n$ with the same volume as A. For every $\epsilon > 0$ let A_ϵ be the ϵ-tube in $\mathcal{R}^n$ about A. Then

$$\mu_n(B_\epsilon) \leq \mu_n(A_\epsilon) \tag{2.7}$$

and consequentially,

$$\mu_{n-1}(\partial B) \leq \mu_{n-1}(\partial A). \tag{2.8}$$

Proof. The fact that Inequality 2.7 is sufficient to prove Inequality 2.8 is an immediate consequence of our definition of surface area. We need some addititional definititions. For any bounded subset $C \subset \mathcal{R}^n$ we let

$$r(C) = \min\{\rho; \exists x \in \mathcal{R}^n : C \subset \overline{B(x,\rho)}\}. \tag{2.9}$$

The function r is continuous on the set of compact non-empty subsets of $\mathcal{R}^n$ with respect to the metric topology induced by the metric

$$\delta(C,D) = \min\{\epsilon; C \subset D_\epsilon, D \subset C_\epsilon\}. \tag{2.10}$$

Each bounded subset C of $\mathcal{R}^n$ then has a circumdisk

$$\text{circ}(C) = \text{the disk with radius } r(C) \textit{ containing } C. \tag{2.11}$$

We define

$$\text{cent}(C) = \text{center of circ}(C). \tag{2.12}$$

Note that with respect to δ, the volume functional μ_n is upper semi-continuous, ie.

$$\delta(C^k, C) \to 0 \Rightarrow \limsup_{k\to\infty} \mu_n(C_k) \leq \mu_n(C). \tag{2.13}$$

Let Γ be a geodesic in $\mathcal{R}^n$. For each $x \in \Gamma$ let

$$H^x = \text{the totally geodesic submanifold perpendicular to } \Gamma \text{ at } x. \tag{2.14}$$

Note that the map $y \mapsto \mu_{n-1}(H^y)$ is upper semi-continuous.

For a compact subset C of $\mathcal{R}^n$, for $x \in \Gamma$ we let

$$C^x = C \cap H^x \tag{2.15}$$

and let

$$\mathbf{D}^x(C) = \text{the disk in } H^x \text{centered at } x \text{ with } \mu_{n-1}(\mathbf{D}^x(C)) = \mu_{n-1}(C^x). \tag{2.16}$$

The set

$$\sigma_\Gamma C = \cup_{x\in\Gamma}\mathbf{D}^x_C \tag{2.17}$$

is called the (Steiner) symmetrization of C with respect to Γ (see Figure 2.7). It follows from the definition of Riemannian measure and the fact that the cut locus of a point is a set of measure zero (see Chavel, pp. 113, 134), or just Fubini's theorem, that $\mu_n(\sigma_\Gamma C) = \mu_n(C)$. A basic but central fact in the proof is that $\text{circ}(C)$ is invariant under symmetrization about geodesics that pass through $\text{cent}(C)$.

For the particular compact set A given in the statement of the theorem, we put

$$M(A) = \{C \in \mathcal{R}^n; C \text{ closed}, \mu_n(C) = \mu_n(A) \text{ and } \forall\epsilon > 0 : \mu_n(C_\epsilon) \leq \mu_n(A_\epsilon)\}. \tag{2.18}$$

The assertion of the theorem is that disks with the same volume as A are in $M(A)$ (see Inequality 2.7). The proof is broken into three main steps.

1. For all compact $A \in \mathcal{R}^n$, there is an element of $M(A)$ that is minimal for r.

2. Given any compact A in $\mathcal{R}^n$ and geodesic Γ, $\sigma_\Gamma A \subset M(A)$.

3. For any compact $A \in \mathcal{R}^n$ which is not a disk, there exists a finite set $\{\Gamma_1, \ldots, \Gamma_k\}$ such that

$$r(\sigma_{\Gamma_k}\sigma_{\Gamma_{k-1}}\ldots\sigma_{\Gamma_1}A) < r(A).$$

By steps 2 and 3, an r-minimal element of $M(A)$ must be a disk. By step 1, an r-minimal element (which we may now fittingly call 'D') exists.

These steps are carried out in the following paragraphs.

1. Existence of r-minimal element. It suffices to consider elements $C \in M(A)$ with $r(C) < r(A)$, and by an isometry we can restrict further to those elements C with $\text{cent}(C) = \text{cent}(A)$; we will call the set of lower-radius, concentric† sets $m(A)$. With respect to the metric δ, the set $m(A)$ is compact in the set of compact sets of $\mathcal{R}^n$. We consider a sequence of such sets C^k with $r(C^k)$ converging to the infimum of the 'r's associated with sets in $m(A)$. There is an associated subsequence which converges with respect to δ to a limit which we'll call B. We want to show that $B \in M(A)$.

For all $\eta > 0$, we trivially have $B \subset B_\eta$ and $B_\epsilon \subset B_{\eta+\epsilon}$, and hence for all k large enough, if $\epsilon > 0$, we have $B_\epsilon \subset B^k_{\eta+\epsilon}$. Therefore since $B^k \in m(A)$ we have that

$$\mu_n(B_\epsilon) \leq \mu_n(B^k_{\eta+\epsilon}) \leq \mu_n(A_{\eta+\epsilon}). \tag{2.19}$$

Hence

$$\begin{aligned} \mu_n(B_\epsilon) &\leq \inf_\eta \mu_n(A_{\eta+\epsilon}) && (2.20)\\ &= \mu_n(\cap_{\eta>0} A_{\eta+\epsilon}) && (2.21)\\ &= \mu_n(A_\epsilon) && (2.22) \end{aligned}$$

where 2.22 is true because A is compact. This shows that $\mu_n(B) \leq \mu_n(A)$. The upper semi-continuity of δ implies that

$$\mu_n \geq \limsup_{k\to\infty} \mu_n(B^k) = \mu_n(A), \tag{2.23}$$

and so we are done with the first step.

2. Symmetrization reduces measure of tubular neighborhood (non-compact case). We want to show that the symmetrization of A is in $M(A)$. We need only show that

$$\mu_n((\sigma_\Gamma A)_\epsilon) \leq \mu_n(A_\epsilon).$$

†*I.e.* concircumcentric.

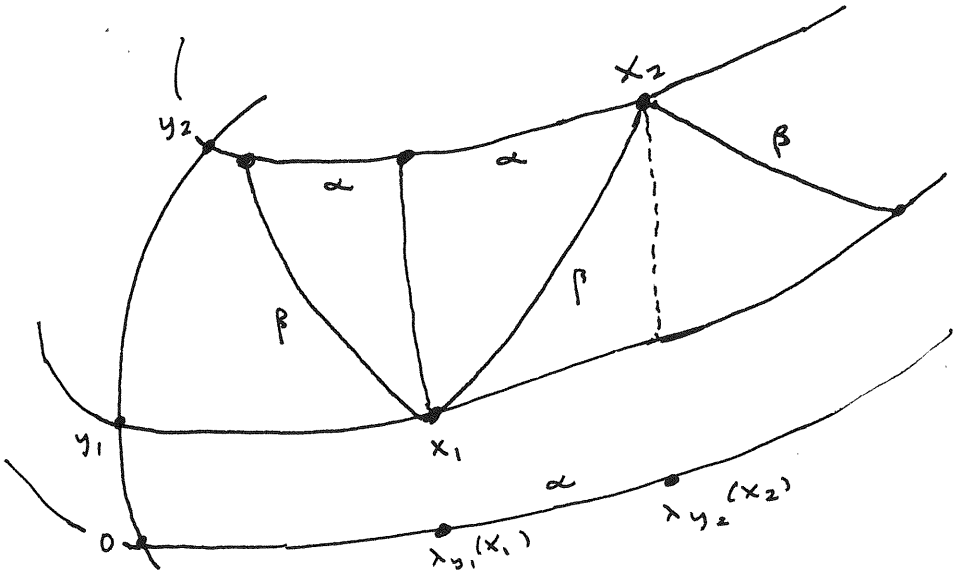


Figure 2.8: $\Psi(y_1, y_2, d(\lambda_{y_1}(x_1), \lambda_{y_2}(x_2)))$ gives the distance between x_1 and x_2.

Fix a point $o \in \Gamma$, and for any other point $y \in \Gamma$, let τ_y denote parallel translation along Γ from y to o. Recall that for every point $x \in \Gamma$ we are letting H^x denote the totally geodesic subspace perpendicular to Γ at x (see Equation 2.14). We now define an isometry

$$\lambda_y : H^y \to H^o$$

by

$$\lambda_y = \exp_o t\cdot|_{\left(\frac{d\Gamma}{dt}|_{(\Gamma^{-1}(o))}\right)^\perp} \circ \tau_y \circ \left(\exp_y t\cdot|_{\left(\frac{d\Gamma}{dt}|_{(\Gamma^{-1}(y))}\right)^\perp}\right)^{-1}. \tag{2.24}$$

(This map makes sense because the inverse of $\exp_y$ maps points in the submanifold to tangent vectors; we then move these tangent vectors using τ_y and project them back to the manifold using $\exp_o$; it is clearly an isometry because all of the $(n-1)$-dimensional totally geodesic subspaces of $\mathcal{R}^n$ have the same metric properties.)

As illustrated in Figure 2.8, there is some function $\Psi : \Gamma \times \Gamma \times [0, \infty) \to \mathbf{R}$ such that for $x_1 \in H^{y_1}$, $x_2 \in H^{y_2}$ we have

$$d(x_1, x_2) = \Psi(y_1, y_2, d(\lambda_{y_1}(x_1), \lambda_{y_2}(x_2))). \tag{2.25}$$

Recall that for a general compact set $C \subset \mathcal{R}^n$ we are letting C^x denote $C \cap H^x$. Equation 2.25 implies that if $y_1, y_2 \in \Gamma$ and $\epsilon > 0$ such that $d(y_1, y_2) \leq \epsilon$ then there is some number η, depending

on y_1, y_2 and ϵ such that if $Z \subset H^{y_1}$ then $(Z_\epsilon)^{y_2} = (\lambda_{y_2}^{-1} \circ \lambda_{y_1})((Z_\eta)^{y_1})$. (That is, we start with an $(n-1)$-dimensional subset Z of H^{y_1}, thicken it in $\mathcal{R}^n$ by ϵ and then slice by H^{y_2}. The result is the same as thickening Z by η in H^{y_1} and then projecting to H^{y_2}.)

Note that $d(y_1, y_2) > \epsilon$ implies that for any $(n-1)$-dimensional subset Z of H^{y_1} that $(Z_\epsilon)^{y_2} = \emptyset$. Therefore, given $\epsilon > 0$, and $y \in \Gamma$ we have

$$(A_\epsilon)^y = \bigcup_{\substack{z \in \Gamma \\ d(z,y)<\epsilon}} \lambda_y^{-1} \circ \lambda_z((A^z)_{\eta(z,y,\epsilon)}), \tag{2.26}$$

and of course we also have

$$((\sigma_\Gamma A)_\epsilon)^y = \bigcup_{\substack{z \in \Gamma \\ d(z,y)<\epsilon}} \lambda_y^{-1} \circ \lambda_z(((\sigma_\Gamma A)^z)_{\eta(z,y,\epsilon)}). \tag{2.27}$$

We have

$$\mu_{n-1}((A_\epsilon)^y) = \mu_{n-1}\left(\bigcup_{\substack{z \in \Gamma \\ d(z,y)<\epsilon}} \lambda_y^{-1} \circ \lambda_z((A^z)_{\eta(z,y,\epsilon)})\right) \tag{2.28}$$

$$\geq \sup_{\substack{z \in \Gamma \\ d(z,y)<\epsilon}} \mu_{n-1}(\lambda_z(A^z)_{\eta(z,y,\epsilon)}). \tag{2.29}$$

Replacing A by $\sigma_\Gamma A$ in 2.28 and 2.29 yeilds equality, since $\lambda_y^{-1} \circ \lambda_z(((\sigma_\Gamma A)^z)_{\eta(z,y,\epsilon)})$ is an $(n-1)$-dimensional disk centered at y. Assuming that Theorem 2.1 is true in dimension $n-1$ means that

$$\mu_{n-1}(((\sigma_\Gamma A)^z)_{\eta(z,y,\epsilon)}) \leq \mu_{n-1}((A^z)_{\eta(z,y,\epsilon)})$$

and so we have

$$\mu_{n-1}(\lambda_z((\sigma_\Gamma A)^z)_{\eta(z,y,\epsilon)}) \leq \mu_{n-1}(\lambda_z((A^z)_{\eta(z,y,\epsilon)})) \tag{2.30}$$

for all $z \in \Gamma$, which implies

$$\mu_{n-1}(((\sigma_\Gamma A)_\epsilon)^y) = \sup_{z \in \Gamma : d(z,y)<\epsilon} \mu_{n-1}(\lambda_z((\sigma_\Gamma A)^z)_{\eta(z,y,\epsilon)}) \tag{2.31}$$

$$\leq \sup_{z \in \Gamma : d(z,y)<\epsilon} \mu_{n-1}(\lambda_z(A^z)_{\eta(z,y,\epsilon)}) \tag{2.32}$$

$$\leq \mu_{n-1}(A_\epsilon^y). \tag{2.33}$$

Therefore, all $y \in \Gamma$ are such that

$$\mu_{n-1}(((\sigma_\Gamma A)_\epsilon)^y) \leq \mu_{n-1}(A_\epsilon^y), \tag{2.34}$$

and so by further properties of the cut-locus (see Chavel [8], Equation (3.51) page 134) we have $\mu_n((\sigma_\Gamma A)_\epsilon) \leq \mu_n(A_\epsilon)$.

2. Symmetrization reduces size of tubular neighborhood (Sphere case). Let A be a closed subset of $\mathbf{S}^{n-1}$ and let γ be a half-circle on $\mathbf{S}^{n-1}$ joining x_0 with $-x_0$. Let u be the midpoint of γ. Identify the sphere $\mathbf{S}_u^{n-2}$ (the equator) with $\mathbf{S}^{n-2}$ (this preserves measure and distances).

For every $y \in \gamma$ ($y \neq \pm x_0$) we define a mapping τ_y from $\mathbf{S}_y^{n-2}$ onto $\mathbf{S}_u^{n-2} = \mathbf{S}^{n-2}$ by projecting along the meridians (that is, for $x \in \mathbf{S}_y^{n-2}$, $\tau_y(x)$ is the point on $\mathbf{S}_u^{n-2}$ which belongs to the half circle joining x_0, x, and $-x_0$).

There is a function f such that if $x_1 \in \mathbf{S}_{y_1}^{n-2}$ and $x_2 \in \mathbf{S}_{y_2}^{n-2}$ then

$$d(x_1, x_2) = f(y_1, y_2, d(\tau_{y_1}(x_1), \tau_{y_2}(x_2))). \tag{2.35}$$

Hence for every $y_1, y_2 \in \gamma$ and $\epsilon > 0$ such that $d(y_1, y_1) \leq \epsilon$, there is an $\eta = \eta(y_1, y_2, \epsilon) \geq 0$ such that every subset $C \subset \mathbf{S}_{y_1}^{n-2}$ has the property that

$$C_\epsilon \cap \mathbf{S}_{y_2}^{n-2} = \tau_{y_2}^{-1}((\tau_{y_1} C)_{\eta(y_1, y_2, \epsilon)}).$$

(If $d(y_1, y_2) > \epsilon$ then $C_\epsilon \cap \mathbf{S}_{y_2}^{n-2} = \emptyset$.)

Hence for every $\epsilon > 0$ and $y \in \gamma, y \neq \pm x_0$,

$$\tau_y((A_\epsilon^y)) = \cup_{z \in \gamma, d(z,y) \leq \epsilon} (\tau_z A^z)_{\eta(z,y,\epsilon)} \tag{2.36}$$

Now let $B = \sigma_\gamma A$. We have automatically that

$$\tau_y((B_\epsilon^y)) = \cup_{z \in \gamma, d(z,y) \leq \epsilon} (\tau_z B^z)_{\eta(z,y,\epsilon)} \tag{2.37}$$

For every $z \in \gamma$ the set $\tau_z B^z$ is by definition a cap in $\mathbf{S}^{n-1}$ such that $\mu_{n-2}(\tau_z B^z) = \mu_{n-2}(\tau_z A^z)$.

By the induction hypothesis (the theorem for $n-1$) we get that

$$\mu_{n-2}((\tau_z B^z)_{\eta(z,y,\epsilon)}) \leq \mu_{n-2}((\tau_z A^z)_{\eta(z,y,\epsilon)}) \tag{2.38}$$

for all 'admissable' y, z, ϵ. Since all the sets appearing in the union on the right-hand side of Equation 2.37 are caps with the same center (*i.e.* u) we get by combining our equations that

$$\begin{aligned} \mu_{n-2}(\tau_y((B_\epsilon)^y)) &= \sup_{z\in\gamma, d(z,y)\leq\epsilon} \mu_{n-2}((\tau_z B^z)_{\eta(z,y,\epsilon)}) && (2.39)\\ &\leq \sup_{z\in\gamma, d(z,y)\leq\epsilon} \mu_{n-2}((\tau_z A^z)_{\eta(z,y,\epsilon)}) && (2.40)\\ &\leq \mu_{n-2}((\tau_y A^y)). && (2.41) \end{aligned}$$

So we have for every $y \neq \pm x_0$ in γ that

$$\mu_{n-2,y}((B_\epsilon^y)) \leq \mu_{n-2,y}((A_\epsilon)^y), \tag{2.42}$$

and by Fubini's theorem, this implies that $\mu_{n-1}(B_\epsilon) \leq \mu_{n-1}(A_\epsilon)$.

3. Repeated symmetrization reduces radius. We want to show that we can reduce r by repeated symmetrization. So, suppose A is not a disk. We note that symmetrization with respect to a geodesic Γ through $\mathrm{cent}(A)$ maps $\mathrm{circ}(A)$ isometrically onto itself. Since σ_Γ pulls A off of the boundary and towards Γ whenever there are gaps in $A \cap H^x$ $\sigma_\Gamma(A)$ (where we recenter[‡] $\sigma_\Gamma(A)$ on $\mathrm{cent}(A)$ if necessary) can not contain all of $\partial(\mathrm{circ}(A))$. We will show that by successive symmetrizations, we can reduce the intersection with the boundary of $\mathrm{circ}(A)$ to nothing, and so reduce r.

If $B \subset \mathrm{circ}(A)$, B closed, and $x \in \partial(\mathrm{circ}(A)) \setminus B$ then $x \notin \sigma_\Gamma(B)$.

If $B \subset \mathrm{circ}(A)$, B closed, and C is subset of $\partial(\mathrm{circ}(A))$ open with respect to the metric induced by d such that $C \cap B = \emptyset$, then since C is open it contains a spherical cap subtended by some

[‡]*I.e.* recircumcenter

angle $\alpha > 0$. If $\alpha > \pi$ then C contains a pair of antipodal points. Symmetrizing with respect to the geodesic determined by this pair of points finishes the argument If $\alpha < \pi$ then there is some Γ such that there exists some relatively open subset C' of $\partial(\mathrm{circ}(A))$ which is disjoint from $\sigma_\Gamma(A)$ and subtends an angle $3\alpha/2$. This then is enough. QEF

2.3 Existence and regularity

Charter We should understand the known results on the existence and regularity of solutions to area minimization problems with multiple volume constraints.

2.3.1 Generalized manifolds

Geometric measure theory works with classes of geometric sets much more general than the class of manifolds. These generalizations were developed, it seems, for two main reasons: (1) Physics, since many phenomena exhibit behavior that is not smooth eg. bubble clusters, black holes, crystals, and so on; (2) Mathematics, since generalized manifolds arise naturally, in much the same way as generalized functions (*i.e.*, distributions). Insofar as (1) and (2) are related (and it happily turns out that they are) we do a disservice to both if we do not note that the main place these extended ideas of manifolds are likely to be interesting is in mathematical physics, where they appear as solutions to differential equations (eg. the variational problems studied in this thesis).

The idea is to use quite general geometrical objects at first and then show that when further natural constraints are imposed (eg. physically motivated ones), the only objects that satisfy the constraints are actually the nice (eg. smooth almost everywhere) things that we were expecting to see. This allows us to do away with potentially very dissatisfying hypotheses, such as 'if A has smooth boundary' (cf. Theorem 2.1 and Definition 2.8). This does not come cheaply. *The sequence of definitions that follows assumes a basic familiarity with vector space duals, differential forms, and*

measures, as might be gained by reading the appendices.

Definition 2.11. We will denote by D^m the space of C^∞ differential m-forms on $\mathbf{R}^n$ with compact support.

Definition 2.12. We let D_m denote the space of linear functionals on D^m endowed with the weak topology, *i.e.* open sets are defined so that if $\{T_j\}$ is a sequence in D_m, it converges to $T \in D_m$ if and only if $T_j\varphi \to T\varphi$ for all $\varphi \in D^m$. An element of D_m is called an *m-dimensional current*.

Definition 2.13. We define the boundary of current by duality, that is, if T is an m-dimensional current, ∂T is an $(m-1)$-dimensional current defined by $(\partial T)(\omega) = T(d\omega)$ for all $(m-1)$-forms ω.

Definition 2.14. We also define the pushforward of a rectifiable current using duality, so that that if T is an m-dimensional current and $\varphi : \mathbf{R}^n \to \mathbf{R}^n$ then $\varphi_* T$ is an m-dimensional current defined by $\varphi_* T(\omega) = T(\varphi^* \omega)$ for all m-forms ω.

Remark 2.6. We might have suspectated that φ_* and φ^* would turn out to be dual operations from the beginning. Definition 2.14 makes this precise.

Definition 2.15. For $m \leq n$, we define the *m-dimensional Hausdorff measure of $A \subset \mathbf{R}^n$*, denoted $\mathcal{H}^m(A)$, in terms of covering classes $\{\mathcal{C}_\delta\}_{\delta>0}$ with the property that $S_\delta \in \mathcal{C}_\delta$ implies S_δ is countable and if $S_\delta = \{S_j \subset \mathbf{R}^n\}_{j=1}^n$ then $A \subset \cup_{j=1}^\infty S_j$ and every S_j has $r(S_j) < \delta$, where r is the circumradius of S_j, defined by Equation 2.9. Letting α_m denote the Lebesgue measure of the closed unit ball in $\mathbf{R}^m$, the Hausdorff measure of A is then given by

$$\mathcal{H}^m(A) = \lim_{\delta \to 0^+} \inf_{S_\delta \in \mathcal{C}_\delta} \sum_{j=1}^{\infty} \alpha_m (r(S_j))^m \tag{2.43}$$

where the limit is seen to exist (in $\mathbf{R} \cup \infty$) because the infima do not decrease as δ decreases.

Definition 2.16. We say that a set $A \subset \mathbf{R}^n$ is an *m-dimensional rectifiable set* if A is $\mathcal{H}^m$-measurable, $\mathcal{H}^m(A) < \infty$, and $\mathcal{H}^m$-almost all of A is contained in the image of some countable collection of Lipschitz maps from $\mathbf{R}^m$ to $\mathbf{R}^n$.

Remark 2.7. The set of smooth m-manifolds is clearly inside the set of rectifiable sets since by paracompactness (and the rest of the definition) a smooth manifold is contained it the image of some countable collection of smooth maps from $\mathbf{R}^m$.

Tangent planes to a rectifiable set are defined $\mathcal{H}^m$-almost everywhere. We can choose an orientation for each tangent plane, and we then write its (now, ordered) basis of tangent vectors as the m-vector $\Xi = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$.

Definition 2.17. An *m-dimensional rectifiable current* is an element $\mathcal{S}$ of D_m that has compact support in D^m and that is associated with a rectifiable set S with ordered basis of tangent vectors Ξ in the sense that the action of $\mathcal{S}$ on $\omega \in D^m$ is given by

$$\mathcal{S}\omega = \int_{\{x \in S\}} \langle \omega(x), \Xi(x) \rangle \, d\mathcal{H}^m.$$

If $\mu : S \to \mathbb{Z}$, we may define

$$\mathcal{S}\omega = \int_{\{x \in S\}} \langle \omega(x), \Xi(x) \rangle \mu(x) \, d\mathcal{H}^m;$$

we then say that $\mathcal{S}$ has *multiplicity* μ.

Remark 2.8. The general idea is to integrate some m-section of the m-tangent bundle over the underlying set. It almost goes without saying that these ideas work very naturally for the set of smooth orientable manifolds, so that we find that set sitting inside of the set of rectifiable currents. In general, if the underlying set we are working with does not admit some kind of global orientation, there could be trouble. Note that by Stokes' Theorem, the definition of boundary given above corresponds to the usual definition of boundary in the case of a rectifiable current that is associated to a smooth oriented manifold with boundary.

It does not necessarily have to be the case that the boundary of a rectifiable current (Definition 2.13) is again a rectifiable current. When this is the case give the current a special name.

Definition 2.18. An *integral current* T is a rectifiable current whose boundary ∂T is also a rectifiable current.

Definition 2.19. A *locally integral current* T is an element of D_m such that for all $x \in \mathbf{R}^n$ there is some integral current S such that $x \notin \text{supp}(T - S)$.

Definition 2.20. We define the *mass norm* (actually semi-norm) on D_m in terms of the comass norm on m-forms (Definition C.24) by

$$\mathbf{M}(T) = \sup_{\omega \in D^m} \{T\omega; \sup_{x \in \mathbf{R}^n} \{\text{comass}(\omega(x))\} \leq 1\}.$$

We can compute the mass norm, for example, of the boundaries of bubble cluster components.

Definition 2.21. A subset S of $\mathbf{R}^n$ is an $(\mathbf{M}, 0, \delta)$-*minimal set (with respect to the null set)* if it is bounded and any Lipschitz deformation Φ of $\mathbf{R}^n$ that is equal to the identity outside of some δ-ball has the property that $\mathcal{H}^m(S) \leq \mathcal{H}^m(\Phi(S))$.

Definition 2.22. We may relax the previous definition and talk about $(\mathbf{M}, \epsilon, \delta)$-*minimal sets (with respect to the null set)*, where ϵ is some function of the form $\epsilon(r) = Cr^\alpha$ and the condition of the previous definition becomes $\mathcal{H}^m(S) \leq (1 + \epsilon(r))\mathcal{H}^m(\Phi(S))$ for all $r < \delta$.

Remark 2.9. $(\mathbf{M}, \epsilon, \delta)$-minimal sets are *rectifiable* [25] and have other nice properties.

2.3.2 Soap bubble clusters

The statements of the theorems are as follows:

Theorem 2.2 ('Existence of compact minimizing bubble clusters'). In $\mathcal{R}^n$, given volumes $v_1, ..., v_m > 0$, there is a boundary mass-minimizing rectifable current comprised of bounded regions V^i with volumes v_i.

This theorem was proved in the more general context of $(\mathbf{M}, \epsilon, \delta)$-minimal sets (which are also rectifiable sets, and so rectifiable currents) by Almgren [25]. The theorem motivates the following definition:

Definition 2.23. A *bubble cluster* is a disjoint union of bounded n-dimensional locally integral currents of multiplicty 1.

If we denote the complement of the cluster $C = \sqcup_{i=1}^{m} V^i$ by V^0, there is a convenient expression for the surface area in terms of the mass norm,

$$\text{Area}(C) = \frac{1}{2}\sum_{i=0}^{m} \mathbf{M}(\partial V^i). \tag{2.44}$$

In [25], Almgren also proved a basic regularity result.

Theorem 2.3 ('Almgren Regularity'). A boundary-mass minimizing bubble cluster is bounded by $C^{1,\alpha}$ surfaces almost everywhere.

Note that where the surfaces are regular, a classical variational argument yeilds that they have constant mean curvature. Almgren's result been improved upon to give a more geometrical description of minimizers, with the first and best results in three dimensions by Taylor.

Theorem 2.4 ('Taylor Regularity'). A soap bubble cluster in $\mathcal{R}^3$ consists of real analytic constant-mean curvature surfaces meeting smoothly in threes at one hundred and twenty degrees along smooth curves, which in turn meet in fours at angles of $\cos^{-1}(-1/3) \approx 109°$.

The curves described in Taylor's regularity theorem are called *singular curves*. The points at which these curves meet are *singular points*.

Taylors' proof is in two pieces; the Euclidean case in [53] and a generalization in [54]. Both are increadibly difficult to read. The singular curves were only proved to be $C^{1,\alpha}$ by Taylor; the analyticity came later; see Morgan [34] for a more complete overview of the history and proof. Morgan points out that this result has been partially extended to higher dimensions.

Theorem 2.5 ('White's extension', unpublished). Soap bubble clusters in $\mathbf{R}^n$ consist of smooth constant mean curvature hypersurfaces meeting in threes at one hundred and twenty degrees along smooth $(n-2)$-dimensional hypersurfaces, which in turn meet in an $(n-4)$-dimensional set.

Remark 2.10. Theorem 2.2 holds in any smooth Riemanian manifold with compact quotient by the isometry group. (As far as I know, this statement is a 'folk theorem'; the idea behind it it that we can pull all the pieces of a minimizer into some compact set where the general ideas of the theorems above can be applied.) It need not hold otherwise – for example, if there is a sequence of bumps with ever-increasing curvature headed off to infinity, any candidate minimizer could be improved upon by moving it to a different bump. The exact statement by Taylor is that the same results hold in a manifold with Hölder continuously differentiable metric is; a discussion of the relationship of these two criteria would take us outside of the scope of this thesis.

Pressure and curvature. One of the most important tools that we can use to study bubble clusters is a physically motivated condition, namely that pressures have to equilibrate. We use the phrase 'pressures add' as short-hand for the longer phrase: the sum of the pressure differences around a singular curve or vertex is zero. This, combined with the fact that pressure difference across walls is proportional to curvature, gives us a useful geometrical tool. A simple proof that pressure is proportional to curvature follows, based on a letter from Miguel Carrión Álvarez.

The basic principle of statics is that energy is minimized at equilibrium. In our case, surface tension is assumed to be constant throughout the whole system, and we have

$$\text{energy} = (\text{surface tension}) * (\text{total area}).$$

Work is the change in energy produced by motion. For example,

$$\text{work} = \text{torque} * (\text{angle rotated})$$

$$\text{work} = \text{force} * \text{displacement}.$$

By definition,

$$\text{pressure} = \text{force}/\text{area}.$$

Since we are in equilibrium the pressure *difference* across a wall

$$\text{pressure}_1 - \text{pressure}_2 = (\text{force}_1 - \text{force}_2)/\text{area}$$

has to be equal to

$$(\text{surface tension})/\text{area}.$$

Hence, if a bubble cluster changes its shape slightly, we have

$$\begin{aligned} \text{work} &= (\text{surface tension}) * (\text{change in area}) \\ &= (\text{pressure difference}) * (\text{change in volume}) \end{aligned}$$

and so

$$(\text{pressure difference}) = (\text{surface tension}) * (\text{area change})/(\text{volume change}).$$

From the appendix, we have

$$(\text{area change})/(\text{volume change}) = (\text{mean curvature})$$

whence,

$$(\text{pressure difference}) = (\text{surface tension}) * (\text{mean curvature}).$$

The fact that pressure is higher for smaller volumes (and so smaller bubbles will curve into larger neighbors) is an experimental observation – made, for example, by Robert Boyle in the 1600's. The above argument provides geometrical intuition for this (smaller things should have higher curvature) but not quite a proof.

Chapter 3

dbs in ccsfs

Charter The known results on double bubbles in constant curvature space forms are to be presented here. (This chapter is based on the results of Michael Hutchings [22] and his collaborators [23], and others who have worked subsequently with these results ([10], [41].)

In the previous two chapters, we became familiar with the geometry of $\mathcal{R}^n$. We studied single bubbles and proved that metric balls give the least perimeter enclosure of a given volume. For the double bubble problem in $\mathcal{R}^n$, the conjectured minimizer is more complicated; it is especially complicated in the negative curvature case. Figure 2.5 and the surrounding discussion motivates the following definition:

Definition 3.1. A *hypersphere-like submanifold* of $\mathcal{R}^n$ is a sphere in the positive or zero curvature cases, and a sphere, a horosphere, or a hyposphere in the negative curvature case.

Lemma 3.1. For prescribed volumes v and w, there is a unique double bubble in $\mathcal{R}^n$, up to an isometry, that consists of three pieces of hypersphere-like submanifolds meeting at 120°. The outer caps are always pieces of spheres. This unique double bubble is called a *standard double bubble.*

To prove this lemma, one has to consider hypersphere-like submanifolds with all of the possible

curvatures and show that sticking them together at 120° yields all the possible volumes exactly once.

Conjecture 3.1 (double bubble conjecture for $\mathcal{R}^n$). The least-surface area way to enclose and separate two given volumes in $\mathcal{R}^n$ is the standard double bubble.

The results of this chapter grew up mainly in the context of the double bubble problem it $\mathbf{R}^n$, where the double bubble conjecture is known to be true for dimensions two [24], three [23], and four, and for some volumes in high dimensions [41]. However the conjecture has been thought of in the context of $\mathcal{R}^n$ for a long time.

Michael Hutchings [22] produced two key simplications of the problem in this general setting. The first simplification applies ideas of symmetrization to show that a non-standard minimizer must be bounded by a surface of revolution. This parallels the symmetrization techniques used in the proof of the single bubble problem in Chapter 2. Once we have this rotational symmetry the geometry becomes simple enough that we might hope to characterize non-standard competitors. The primary question then is, exactly how many regions can a minimizing bubble have? The second key simplification is an explicit bound on the number of components of each region in terms of the volumes. This makes a complete characterization possible. Indeed if the bound tells us that the largest of the two regions is connected, the a variational argument developed in the $\mathbf{R}^3$ by Hutchings, Morgan, Ritoré and Ros, successfully applied in the $\mathbf{R}^4$ version of the problem by Reichardt *et al.* and revised to work in $\mathbf{S}^n$ and $\mathbf{H}^n$ by Cotton and Freeman [10] can be applied to shows the instability of all remaining non-standard competitors. The two key simplifications of Hutchings and the instability argument for constant curvature space forms will be described in this chapter.

3.1 Hutchings structure theory

Hutchings proved many results for m prescribed volumes. We need only two, but will do things in the general language when possible, because usually when it is possible it is also usually no harder.

Definition 3.2. We define the *least-area function* or *isoperimetric profile* of a given manifold M to be the map that takes a given volume to the surface area of a least-area enclosure in the space that contains the given volume. We denote this function by I_M.

More generally, we define the *m-volume least-area function* to be the map that takes m given volumes to the surface area of a least-area enclosure in the space that contains these volumes in separate chambers. We denote this function by I_M^m.

We are especially interested in the *two-volume least-area function*, which we denote by the special symbol $\mathit{\Pi}_M$.

We frequently will drop the subscripts of these maps when there is no chance of this making things confusing.

Remark 3.1. The isoperimetric profile is a primative of mean curvature for minimizers, *i.e.*,

$$dI(x)/dx = dA_{\text{minimizer}}(x)/dx = H_{\text{minimizer}}(x).$$

This is in fact a proof if you accept the second equality. The reason for that the equality is a simple variational argument (see [33] p. 15).

3.1.1 Symmetry

We saw in Chapter 2 how important symmetry was for solving the isoperimetric problem in constant curvature space forms. Symmetry is just as important for the double bubble problem, even though it is weaker (minimizing double bubbles are not spherically symmetric) and does not finish the job. *For double bubbles, the central symmetry result is that if $n \geq 3$ any area-minimizing double bubble in $\mathbf{R}^n$, $\mathbf{S}^n$ or $\mathbf{H}^n$ is symmetric about some line.* This is a corollary to the broader-reaching

Theorem 3.1 (Symmetry Theorem). Let B be a minimal enclosure of m volumes in $\mathbf{R}^n$. If m is less than or equal to $n-1$ then B is symmetric about some $m-1$ plane.

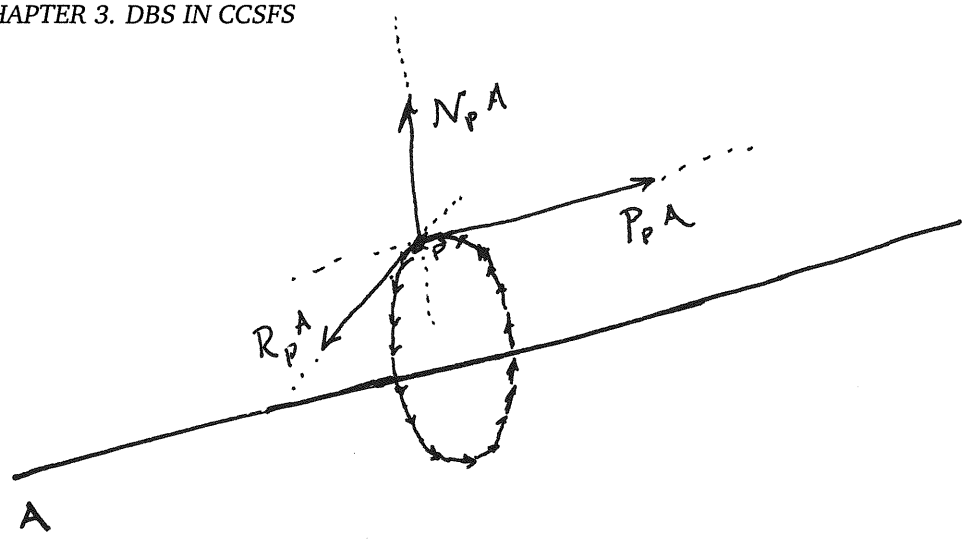


Figure 3.1: Hutchings' decomposition of $T_p(\mathbf{R}^3 \setminus A)$.

Hutchings proves Theorem 3.1 in a series of lemmas. We will do the same, after introducing some notation.

In the statement of the theorem and in all that follows, B is taken to be an n-dimensional rectifiable current made up of a disjoint collection of n-dimensional regions V^i, such that V^i has volume v_i. Until we say otherwise, these regions may be broken into many disjoint components.

Since it is area-minimizing its boundary is mainly made up of smooth manifold points. We let B_{reg} denote the set of smooth manifold points in the boundary of B.

Definition 3.3. Let A be an affine subspace of $\mathbf{R}^n$ with $\dim A \leq n-2$. We decompose the tangent bundle of $\mathbf{R}^n \setminus A$ as follows:

$$T(\mathbf{R}^n \setminus A) = \mathcal{P}A \oplus \mathcal{N}A \oplus \mathcal{R}A, \tag{3.1}$$

where $\mathcal{P}A$ are vectors parallel to A, $\mathcal{N}A$ are vectors normal to A, and $\mathcal{R}A$ are vectors that are perpendicular to both $\mathcal{P}A$ and $\mathcal{N}A$. It is important that we exclude A, because the decomposition degenerates at points in A (see Figure 3.1). At a given point $p \in \mathbf{R}^n \setminus A$ we write the associated decomposition of the tangent space in the same way, so $T_p(\mathbf{R}^n \setminus A) = \mathcal{P}_pA \oplus \mathcal{N}_pA \oplus \mathcal{R}_pA$.

Remark 3.2. The case in which A is a line in $\mathbf{R}^3$ is illustrated in Figure 3.1. Here it is clear that $\mathcal{R}A$

are vectors taken from vectorfields that determine a rotation about A. It works the same way in higher dimensions. For example, if A is a $2-$plane in $\mathbf{R}^4$ then $\mathcal{P}_p A$ is two dimensional, $\mathcal{N}_p A$ is one dimensional, and so there is an extra dimension left over for $\mathcal{R}_p A$.

Lemma 3.2 ('Infinitesimal Symmetry Implies Global Symmetry'). Let B be a minimizing enclosure of m volumes in $\mathbf{R}^n$, and let A be an affine subspace of $\mathbf{R}^n$ with $\dim(A) \leq n-2$. Suppose that for almost all $p \in B_{\text{reg}} \setminus A$ that we have $\mathcal{R}_p A \subset T_p B$. Then B is symmetric about A.

Proof. Since $\dim A \leq n-2$, an object mapped to itself invariantly under rotations about A will also be mapped to itself invariantly under reflections about A.

Let φ be a rotation about A. Our goal is to show that $\varphi_* \partial V^i = \partial V^i$ for all i. By the definition of the pushforward of a current (Definition 2.14), this means that the boundary of V^i is fixed under rotation about A.

Associated to φ is a one parameter group $\{\varphi_t\}$ of rotations about A such that $\varphi = \varphi_1$. These induce a vector field $\mathcal{V}$ on $\mathbf{R}^n$. If ω is an $(n-1)$-form on $\mathbf{R}^n$,

$$\begin{aligned} \frac{d}{dt}\Big|_{t=0} \langle \partial V^i, \varphi_t^* \omega \rangle &= \langle \partial V^i, (d\iota(\mathcal{V}) + \iota(\mathcal{V})d)\omega \rangle && (3.2) \\ &= \langle \partial V^i, d\iota(\mathcal{V})\omega \rangle + \langle \partial V^i, \iota(\mathcal{V}) d\omega \rangle && (3.3) \\ &= \langle \partial V^i, \iota(\mathcal{V}) d\omega \rangle, && (3.4) \end{aligned}$$

where the equality between 3.3 and 3.4 is due to Stokes' theorem. If the members of this chain of equalities are equal to zero, then the value of the pairing is constant to first order at $t = 0$, and this implies that the pairing is constant for all t. If $\mathcal{V}(p) \in T_p B$, Equation 3.4 implies that for almost all $p \in B_{\text{reg}} \cap \partial V^i$ we have

$$\langle \iota_{\mathcal{V}} d\omega(p), \Xi(p) \rangle = 0, \tag{3.5}$$

where $\Xi(p)$ denotes the ordered basis of $T_p(\partial B)$. This shows that the members of 3.4 are equal to zero. Since $\mathcal{V}(p) \in \mathcal{R}_p A$ for all $p \notin A$ and $\mathcal{R}_p(A) \subset T_p B$ for almost all $p \in B_{\text{reg}}$ such that $p \notin A$, we

have

$$\langle \partial V^i, \omega \rangle = \langle \partial V^i, \varphi^* \omega \rangle \tag{3.6}$$

Since $\langle \partial V^i, \varphi^* \omega \rangle = \langle \varphi_* \partial V^i, \omega \rangle$ by definition, this completes the proof. QEF

Lemma 3.3 (Bisectors Orthogonal). Let B be a minimizing enclosure of m volumes in $\mathbf{R}^n$, and let H be a hyperplane that bisects all of the V^i. Then H and B_{reg} are orthogonal wherever they intersect.

Proof. Let B_+ and B_- be the halves of B on either side of H. Since H bisects all of the regions, we can make a new cluster $\widehat{B_+}$, equal to the union of B_+ the reflection of B_+ across H. The regions $\widehat{V^i}$ that make up $\widehat{B_+}$ have the same volume as the V^i. Since B is minimizing, $\text{area}(\widehat{B_+}) \geq \text{area}(B)$, so $\text{area}(B_+) \geq \text{area}(B_-)$. Similarly, $\text{area}(B_-) \geq \text{area}(B_+)$, hence they must be equal, and so we must have $\text{area}(\widehat{B_+}) = \text{area}(\widehat{B_-}) = \text{area}(B)$.

Hence, $\widehat{B_+}$ and $\text{area}(\widehat{B_-})$ are minimizing too. By regularity, the tangent spaces to the regular points of B contained in H are either orthogonal to or parallel to H.

The final step is to show that if B is tangent to H at a regular point then either $\widehat{B_+}$ or $\widehat{B_-}$ can be modified to decrease area. (This is enough because the tangent spaces will only be parallel to H at points where $\widehat{B_+}$ is tangent to H or at points where $\widehat{B_+}$ and H do not intersect.)

Note that $\mathbf{R}^n$ can be parameterized by $H \times \mathbf{R}$. If $\mathbf{D}_p(r)$ is a closed ball of radius r about $p \in H$ then B is regular in $\mathbf{D}_p(r) \times [-\epsilon, \epsilon] \subset H \times \mathbf{R}$. Since B is tangent to H at p, chosing r small enough to have

$$B \cap \partial(\mathbf{D}_p(r) \times [-\epsilon, \epsilon]) \subset (\partial \mathbf{D}_p(r) \times [-\epsilon, \epsilon])$$

for small ϵ, and so that $\mathbf{D}_p(r) \times \{\pm\epsilon\}$ are in the same component of $\widehat{B_+}$, we remove $\widehat{B_+}$ from its intersection with $\mathbf{D}_p(r) \times [-\epsilon, \epsilon]$ and add back $\partial \mathbf{D}_p(r) \times [-\epsilon, \epsilon]$. This decreases area to the order of r^{n-1}, and changes the areas of $\widehat{V^i}$ of at most order ϵr^{n-1}.

By a standard first variation argument, we can restore the volumes by adjusting the cluster

elsewhere, with an area increase of at most order ϵr^{n-1}. This is smaller than the initial increase for r sufficiently small. QEF

Lemma 3.4 (Symmetry about intersections). Let B be a minimizing enclosure of m volumes in $\mathbf{R}^n$. Let if $k > 1$ and $H_1, \ldots, H_k$ be mutually orthogonal hyperplanes. If B is symmetric about each H_i then B is symmetric about $A = \cap_{i=1}^{k} H_i$.

Proof. B maps to itself invariantly under reflection across each H_i, so also maps to itself invariantly under reflection across A. So every hyperplane containing A bisects $V^1, \ldots, V^m$. Since B_{reg} meets its bisectors orthogonally wherever they intersect, and the directions perpendicular to all of the H_i are the directions of rotation about A, B is symmetric about A, because infinitesimal symmetry implies global symmetry. QEF

Lemma 3.5 (Linear Algebra). If A_1 and A_2 are affine subspaces of $\mathbf{R}^n$ with non-empty intersection, and $p \notin \text{span}(A_1, A_2)$, then $\mathcal{R}_p(A_1 \cap A_2) = \mathcal{R}_p A_1 \oplus \mathcal{R}_p A_2$

Proof. Any rotation fixing A_1 or A_2 fixes $A_1 \cap A_2$ so

$$\mathcal{R}_p(A_1 \cap A_2) \supset \mathcal{R}_p A_1 \oplus \mathcal{R}_p A_2.$$

If

$$\mathcal{R}_p(A_1 \cap A_2) \not\subseteq \mathcal{R}_p A_1 \oplus \mathcal{R}_p A_2,$$

then since

$$\dim \mathcal{R}_p A_1 \cap \mathcal{R}_p A_2 > n - \dim \text{span}(A_1, A_2) - 1,$$

and since

$$\mathcal{R}_p A_1 \cap \mathcal{R}_p A_2 = (\mathcal{P}_p A_1 \oplus \mathcal{P}_p A_2 \oplus \mathcal{N}_p A_1 \oplus \mathcal{N}_p A_2)^{\perp}$$

and, lastly, since

$$\dim(\mathcal{P}_p A_1 \oplus \mathcal{P}_p A_2) = \dim \text{span}(A_1, A_2),$$

we must have

$$\mathcal{N}_p A_1 \subset \mathcal{P}_p A_1 \oplus \mathcal{P}_p A_2.$$

Since $\mathcal{N}_p A_1$ is perpendicular to $\mathcal{P}_p A_1$ and $\mathcal{N}_p A_1 \subset \mathcal{P}_p A_2$, we must have $p \in \text{span}(A_1, A_2)$. This contradicts our assumption, so we have the claim. QEF

Lemma 3.6 (Assembly, Euclidean case). Let B be a minimizing enclosure of m volumes in $\mathbf{R}^n$. Let H be a hyperplane, and let B_+ and B_- be the two symmetrizations of B across H. Suppose that both B_+ and B_- minimize area for the volumes they enclose. Let A_+ and A_- be non-empty affine subspaces of H of dimension at most $n-2$. If $B_\pm$ is symmetric about $A_\pm$ then $A_+ \cap A_- \neq \emptyset$ and B is symmetric about $A_+ \cap A_-$.

Proof. First note thet $B \cap H$ is symmetric about A_1 and A_2 in H. We immediately rule out the possibility that $A_1 \cap A_2 = \emptyset$ since that would imply that B is not compact, which contradicts its minimality. To prove symmetry it is enough to show that for $\mathcal{H}^{n-2}$-almost every $p \in B_{\text{reg}}$ such that p is not in $A_1 \cap A_2$ that

$$\mathcal{R}_p(A_1 \cap A_2) \subset T_p B.$$

We first show that for almost every $p \in B \cap H$ that is not in $A_1 \cap A_2$ that each of B, B_1 and B_2 is regular at p and for these 'p's we have $\mathcal{R}_p(A_1 \cap A_2) \subset T_p B$.

Since B_i is regular almost everywhere and is symmetric about A_i we know that almost every point in $(B_i \cap H) \setminus A_i$ is regular. Hence almost every point of $B \cap H$ that is not in either A_1 or A_2 is a regular point of both B_1 and B_2. Hence, by the Bisectors Orthogonal lemma (Lemma 3.3), we have that B_1 and B_2 are orthogonal to H at each $p \in (B \cap H) \setminus (A_1 \cup A_2)$, and so a small ball $\mathbf{D}_\epsilon(p)$ meets only two of the 'V^i's that make up B, so B is regular in $\mathbf{D}_\epsilon(p)$ except on a set of Hausdorff dimension at most $n-8$. From this it follows that almost every point in $(B \cap H) \setminus (A_1 \cup A_2)$ is a regular point for B, B_1 and B_2, as we wanted.

If U is a regular neighborhood of such a point, then by uniqueness of analytic continuation we

have

$$U \cap B = U \cap B_1 = U \cap B_2.$$

Hence we have

$$\mathcal{R}_p(A_1), \mathcal{R}_p(A_2) \subset T_pB$$

for all $p \in U \setminus (A_1 \cup A_2)$. By the Linear Algebra lemma (Lemma 3.5), for all $p \in U \setminus H$ we have

$$\mathcal{R}_p(A_1 \cap A_2) \subset T_pB.$$

By continuity this holds for all $p \in U$.

Let W_i be the closed half-space in which B_i lies. Let O_1 be the union of orbits under rotation about A_2 whose intersection with H is contained in A_1. Note that O_1 is half of a k-plane where $k \leq \dim(A_1) + 1$. Define O_2 similarly. By our work in the previous paragraph, for $\mathcal{H}^{n-1}$-almost every $p \in (B \cap W_1) \setminus O_2$ there is a rotation φ about A_1 such that $\varphi(p) \in H$ and $\varphi(p)$ is a regular point for both B and B_1, and furthermore, $\mathcal{R}_p(A_1 \cap A_2) \subset T_pB$. Since B_1 is invariant under φ and φ is a rotation about $A_1 \cap A_2$ it must be the case that B is regular at p and that $\mathcal{R}_p(A_1 \cap A_2) \subset T_pB$. Similar statements can be made for almost every $p \in (B \cap W_2) \setminus O_1$.

Now suppose that $p \in B_{\text{reg}} \cap O_i$. Either there is some regular neighborhood of p contained in O_i or p is a limit of regular points of B not in O_i whose tangent spaces contain $\mathcal{R}_p(A_1 \cap A_2)$. In the latter case, continuity implies that $\mathcal{R}_p(A_1 \cap A_2) \subset T_pB$. In the former case, using the rotation invariance of both B_1 and B_2, we must have that the point in $\mathbf{R}^n$ on either side of U lie in the same V^i so U can be removed to decrease area, contradicting the minimality of B. QEF

We can now prove our big theorem.

Proof of Symmetry Theorem 3.1. Since B is compact, we can apply the ham sandwich theorem to produce a hyperplane H_1 that bisects all of the regions. By Assembly (Lemma 3.6) it is enough to show that each of the symmetrizations of B across H_1 are symmetric about an $(m-1)$-plane in H_1. Let B_1 be one of the two possible symmetrizations. By the Borsuk-Ulam theorem, there is a

hyperplane H_2 orthogonal to H_1 that bisects all of the regions. By Assembly again, it is enough to show that each of the symmetrizations of B_1 across H_2 is symmetric about an $(m-1)$-plane in $H_1 \cap H_2$. We let B_2 be one of the two possible symmetrizations, and continue in this fashion until we have obtained hyperplanes $H_1, \ldots, H_{n-m+1}$ and a minimal cluster B_{n-m+1}. By Lemma 3.4 this cluster is symmetric about the $(m-1)$-plane $A = \cap H_i$. QEF

Corollary 3.1. Let B be a minimal double bubble in $\mathbf{R}^n$ for $n \geq 3$ and let $H \subset \mathbf{R}^n$ be a hyperplane. If each symmetrization of B across H is area-minimizing for the volumes it encloses, then B is symmetric about a line in H.

Proof. Use the symmetry theorem, starting with $H_1 = H$. QEF

3.1.2 Concavity of the least-area function

In this subsection, I^m means the m-volume least-area function associated with a constant curvature space form of arbitrary dimension, and $\mathit{\Pi}$ means the 2-volume least-area function associated with a constant curvature space form of arbitrary dimension. The proof of the main theorem in this section (Theorem 3.2) relies on the symmetry we have in constant curvature space forms, and does not apply when there are more than two prescribed volumes. We begin with a simple lemma that does hold for any number of prescribed volumes.

Lemma 3.7 ('Continuity'). For all m, I^m is a continuous function.

Proof. Let C be a least-area bubble cluster for volumes $v_1, \ldots, v_m$. Add a small round ball of radius ϵ somewhere away from the cluster, and say that it is now part of V^1. This increases the volume of V^1 by less than ϵ and increases surface area by less than ϵ. This is an upper bound on the increase of surface area for a minimizing bubble cluster with volumes $v_1 + \epsilon, v_2, \ldots, v_m$, hence I^m is continuous in the first component, and so by the same argument, in all.

QEF

Theorem 3.2 ('Concavity'). Π is a concave function in the sense that if $\mathbf{v}$ and $\mathbf{w}$ are two pairs of volumes, then for $t \in (0,1)$

$$\Pi(t\mathbf{v} + (1-t)\mathbf{w}) > t\Pi(\mathbf{v}) + (1-t)\Pi(\mathbf{w}).$$

Proof. We must prove that

$$\Pi(tv_1 + (1-t)w_1, tv_2 + (1-t)w_2) > t\Pi_(v_1, v_2) + (1-t)\Pi(w_1, w_2) \tag{3.7}$$

for all $t \in (0,1)$. For a contradiction, suppose not.

By continuity, the function f defined by

$$f(t) = \Pi(tv_1 + (1-t)w_1, tv_2 + (1-t)w_2) - t\Pi(v_1, v_2) - (1-t)\Pi(w_1, w_2) \tag{3.8}$$

assumes a minimum on the interval [0,1] at some t_0.

Let B be a minimizing cluster of two regions V and W, with volumes

$$(v, w) = t_0(v_1, v_2) + (1-t_0)(w_1, w_2) = (t_0v_1 + (1-t_0)w_1, t_0v_2 + (1-t_0)w_2).$$

By the Symmetry Theorem (Theorem 3.1), B is symmetric about some line ℓ.

We can parameterize the set of oriented hyperplanes in $\mathbf{R}^n$ by $\mathbf{S}^{n-1} \times \mathbf{R}$, since each oriented hyperplane is determined by a normal direction and a distance from the origin in that direction.

We define a map h_+ from the set of oriented hyperplanes, so that given $H \in \mathbf{S}^{n-1} \times \mathbf{R}$ we let H^+ denote the upper half space determined by H, and we put

$$h_+(H) = (\mathrm{vol}(V \cap H^+), \mathrm{vol}(W \cap H^+)) \in \mathbf{R}^2$$

Since regularity implies B is compact, h_+ is continuous.

By an isometry, we may assume 0 is in the line of symmetry ℓ. We then choose $x \in \mathbf{S}^{n-1}$ perpendicular to ℓ. The hyperplane determined by $(x, r) \in \mathbf{S}^{n-1} \times \mathbf{R}$ contains ℓ if and only if $r = 0$, and is parallel to ℓ otherwise. For all $r \in \mathbf{R}$ we have that

$$h_+(x, r) + h_+(x, -r) = (v, w), \tag{3.9}$$

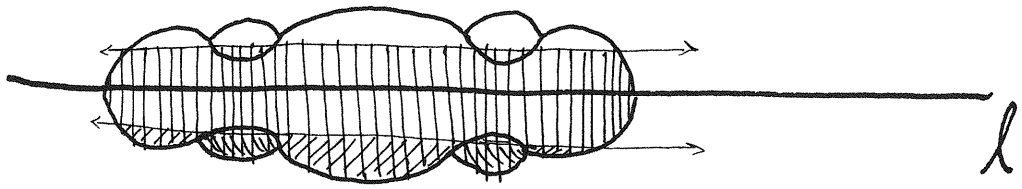


Figure 3.2: Cutting the bubble by oriented planes equidistant from ℓ gives all the volume.

since by symmetry the volumes that lie above (x, r) are equal the volumes that lie below $(x, -r)$, hence the total ammount of volume collected by the two maps is equal to the total volume of the bubble (see Figure 3.2).

The line segment

$$\mathcal{K} = \left\{ \frac{t(v_1, v_2) + (1-t)(w_1, w_2)}{2}; t, 2t_0 - t \in (0,1) \right\} \tag{3.10}$$

in the volume plane has the property that either $h_+(x, r) \in \mathcal{K}$ for some $r \neq 0$ or $h_+(y \times [0, \infty)) \in \mathcal{K}$ whenever y is close to x.

In both cases, there is a hyperplane $H_{x,t} \not\supseteq \ell$ such that

$$h_+ H_{x,t} = \frac{t(v_1, v_2) + (1-t)(w_1, w_2)}{2} \in \mathcal{K}.$$

Let

$$\begin{aligned} a_0 &= \text{area}(B \cap H_{x,t}) \\ a_+ &= \text{area}(B \cap H^+_{x,t}) \\ a_- &= \text{area}(B \cap H^-_{x,t}). \end{aligned}$$

Replacing $B \cap H^-_{x,t}$ with the reflection of $B \cap H^+_{x,t}$ across H, we find that

$$\begin{aligned} a_0 + 2a_+ &\geq \Pi(t(v_1, v_2) + (1-t)(w_1, w_2)) \\ &= f(t) + t\Pi(v_1, v_2) + (1-t)\Pi(w_1, w_2) \\ &\geq f(t_0) + t\Pi(v_1, v_2) + (1-t)\Pi(w_1, w_2) \\ &= \Pi(t_0(v_1, v_2) + (1-t_0)(w_1, w_2)) \\ &\quad + (t-t_0)\Pi(v_1, v_2) + (t_0 - t)\Pi(w_1, w_2). \end{aligned}$$

If we symmetrize in the other direction, we find that

$$\begin{aligned} a_0 + 2a_- &\geq \Pi(t_0(v_1, v_2) + (1-t_0)w) + ((2t_0 - t) - t_0)\Pi(v_1, v_2) \\ &\quad + (t_0 - (2t_0 - t))\Pi(w_1, w_2). \end{aligned}$$

Adding these, we get

$$2(a_0 + a_+ + a_-) \geq 2\Pi(t_0(v_1, v_2) + (1-t_0)(w_1, w_2)).$$

By the definition of the 'a's, this must be an equality, so each of the non-strict inequalities in the chain above must be equalities.

We can therefore conclude that each symmetrization of B across $H_{x,t}$ is minimizing for the enclosed volumes.

By Corollary 3.1, B is symmetric about some line $\ell' \subset H_{x,t}$. Since B is compact, ℓ and ℓ' must intersect. But $\ell \neq \ell'$, so by applying the Assembly Lemma 3.6 to a hyperplane containing ℓ and ℓ' we get that B is a union of concentric spheres. Then B must contain only one sphere by regularity. *So B encloses only one volume, and (v_1, w_1) both lie on the same coordinate axis in $\mathbf{R}^2$.*

But Π is strictly concave along any line through the origin, since (in general) by scaling

$$I^m_n(\lambda v_1, \ldots, \lambda v_m) = \lambda^{(n-1)/n} I^m_n(v_1, \ldots, v_m) \tag{3.11}$$

But this contradicts our assumption that Π is not concave along this line.

This completes the proof in the case of $\mathbf{R}^n$. For the other spaces, we just have some comments: the proof for $\mathbf{H}^n$ is essentially the same, the only significant thing that one must check is that $I_{\mathbf{H}^n}$ is concave and this is very easy (see Lemma 3.8 in the next section); the theorem for $\mathbf{S}^n$ is a straightforward rewording of the theorem for $\mathbf{R}^n$– some things to note for the *proof* are (1) oriented hyperplanes in $\mathbf{S}^n$ are parameterized by $\mathbf{S}^n$, (2) instead of looking at paths of the form $h_+(x \times [0, \infty))$, we look at images of h_+ under geodesics, and (3) we have strict concavity along the edge of the simplex $\{v_1 + v_2 + v_3 = \text{vol}(\mathbf{S}^n)\}$ because in a minimal sphere dividing $\mathbf{S}^n$ into two equal volumes, the pressure difference between the two regions decreases as two of the volumes become.

QEF

3.1.3 Connectivity (component bound)

Lemma 3.8 ('Ratio of area to volume is decreasing for spheres'). In $\mathbf{R}^n$, $\mathbf{S}^n$, and $\mathbf{H}^n$, $\frac{I(x)}{x}$ is decreasing in x.

Proof. We use the following standard formulas for area and volume as functions of radius (α_n is again the Lebesgue measure of the unit ball in $\mathbf{R}^n$):

	$\mathbf{R}^n$	$\mathbf{S}^n$	$\mathbf{H}^n$
A_{sphere}	$n\alpha_n r^n$	$n\alpha_n \sin^{n-1} r$	$n\alpha_n \sinh^{n-1} r$

Working from the appendix on variational calculus, we have

$$H_{\text{sphere}} = \frac{dA_{\text{sphere}}}{dV_{\text{sphere}}} = \frac{dA_{\text{sphere}}(r)/dr}{dV_{\text{sphere}}(r)/dr} = \frac{A'_{\text{sphere}}(r)}{A_{\text{sphere}}(r)},$$

By a very easy computation we get

	$\mathbf{R}^n$	$\mathbf{S}^n$	$\mathbf{H}^n$
H_{sphere}	n/r	$(n-1)\cot(r)$	$(n-1)\coth(r)$

A glance at the graphs of these functions (on the appropriate domains) proves that mean curvature of spheres is a decreasing function of radius in each space. Since $I(x)$ is a primative of mean curvature,

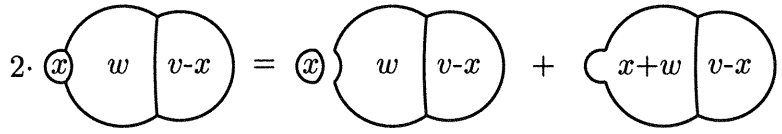


Figure 3.3: Decomposition of a double bubble in which V has a component of volume x

this proves that I is concave. From this we see directly that the ratio $I(x)/x$ is decreasing, since if I is is not decreasing, its slope is. QEF

Lemma 3.9 ('Basic Estimate'). Consider a minimal enclosure of volumes v and w in $\mathcal{R}^n$. Suppose that the region V has a connected component with volume x. Then

$$2\Pi(v,w) \geq I(w) + \frac{v}{x}I(x) + I(v+w).$$

Proof. From Figure 3.3,

$$2\Pi(v,w) \geq \Pi(v-x,w) + I(x) + \Pi(v-x,w+x). \tag{3.12}$$

From Concavity (Theorem 3.2), with $v_1 = 0, v_2 = v+w, w_1 = v, w_2 = w$ and $t = x/v$ in Inequality 3.7, we get

$$\Pi(v-x,w+x) \geq \frac{v-x}{v}\Pi(v,w) + \frac{x}{v}I(v+w),$$

and again with $v_1 = 0, v_2 = w, w_1 = v, w_2 = w$ and $t = x/v$ we get

$$\Pi(v-x,w) \geq \frac{v-x}{v}\Pi(v,w) + \frac{x}{v}I(w).$$

Substitution of these bounds into Inequality 3.12 gives our result, upon making the following rearrangements:

$$\begin{aligned} 2\Pi(v,w) &\geq \frac{v-x}{v}\Pi(v,w) + \frac{x}{v}I(w) + I(w)\frac{v-x}{v}\Pi(v,w) + \frac{x}{v}I(v+w) \\ &= 2\Pi(v,w) - 2\frac{x}{v}\Pi(v,w) + \frac{x}{v}I(w) + I(x) + \frac{x}{v}I(v+w) \\ \Rightarrow 2\frac{x}{v}\Pi(v,w) &\geq \frac{x}{v}I(w) + I(x) + \frac{x}{v}I(v+w). \end{aligned}$$

Now divide through by x/v. QEF

Remark 3.3. The theorem we just proved and the following corollaries work in any space where $I\!I$ is concave.

Corollary 3.2. In a double bubble enclosing and separating the pair of volumes (v, w) where (v, w) satisfies

$$F(v,w) = 2I\left(\frac{v}{2}\right) + I(w) + I(v+w) - 2I\!I(v,w) > 0, \tag{3.13}$$

V is connected.

Proof. Suppose Inequality 3.13 holds. Let the *smallest* component of the region V be V_1. By concavity of $I\!I$, (Theorem 3.2) we can apply the Basic Estimate (Lemma 3.9). This yields

$$\frac{v}{v_1} I(v_1) \leq 2I\!I(v,w) - I(w) - I(v+w). \tag{3.14}$$

Since the left hand side of this inequality is decreasing in $v_1 \leq v$ by Lemma 3.8, if $v_1 \leq v/2$, we have

$$\frac{v}{v/2} I(v_1) \leq 2I\!I(v,w) - I(w) - I(v+w) \tag{3.15}$$

i.e.

$$2I(v_1) \leq 2I\!I(v,w) - I(w) - I(v+w) \tag{3.16}$$

i.e.

$$2I(v_1) - 2I\!I(v,w) + I(w) + I(v+w) \leq 0, \tag{3.17}$$

which implies

$$2I(\frac{v}{2}) - 2I\!I(v,w) + I(w) + I(v+w) \leq 0, \tag{3.18}$$

which is a direct contradiction of Inequality 3.13. Therefore we must have $v_1 > v/2$ and therefore the *smallest* component of V must have volume *greater* than half of the volume of V. Any other component would have to have still greater volume, which is impossible, so V must be connected.

QEF

Essentially the same proof works to show

Corollary 3.3 ('Component Bound'). If

$$F_k(v,w) = kI\left(\frac{v}{k}\right) + I(w) + I(v+w) - 2\Pi(v,w) > 0,$$

V has less than k components.

3.2 Instability arguments

Hutchings' result on the rotational symmetry of area minimizing double bubbles (Theorem 3.1 in the previous section) underlies a series of arguments that showed that area minimizing double bubbles are standard in $\mathbf{R}^3$, $\mathbf{R}^4$, and for a range of volumes, in higher dimensions, as well as for equal volumes in $\mathbf{S}^3$ and $\mathbf{H}^3$. The symmetry, plus the component bound that follows from symmetry by way of Concavity (Theorem 3.2), give key simplifications to the geometry of the problem. However, these results do not in and of themselves finish the task.

Hutchings, Morgan, Ritoré, and Ros [23] developed an instability argument that rules out all the non-standard double bubbles left after these simplifications. (One of the most important steps was to precisely describe which non-standard double bubbles were still possible.) Reichardt *et al.* [41] adapted the instability argument to $\mathbf{R}^4$, and Cotton and Freeman [10] generalized the Hutchings-Morgan-Ritoré-Ros-Reichardt instability argument to $\mathbf{S}^3$ and $\mathbf{H}^3$, to use in their proof for the equal volumes cases of the double bubble problem in these spaces. Not a lot has to change, but there are lots of details to check; we review the statements of the main theorems from Cotton and Freeman's REU report.

3.2.1 Structure and Stability

Lemma 3.10 ('Fundamental Instability Argument'). Let Σ be a stable double bubble in $\mathbf{R}^n$, $\mathbf{S}^n$, or $\mathbf{H}^n$. Let $\mathcal{V}$ be a vectorfield corresponding to some isometric motion of $\mathbf{R}^n$, $\mathbf{S}^n$, or $\mathbf{H}^n$. Suppose that

the points where $\mathcal{V}$ is tangent to Σ separate the bubble into at least four pieces. Then the normal component of $\mathcal{V}$ vanishes on any smooth component of Σ.

Lemma 3.11. Consider a stable double bubble of revolution $\Sigma \subset \mathbf{S}^n, n \geq 3$ with a piecewise smooth generating curve $\Gamma = \cup\bar{\Gamma}_i$ such that Γ_i end either on the axis or meet in threes at vertices v_{ijk}. Let the axis of symmetry be denoted by ℓ. At each $p \in \cup\Gamma_i$ let $N(p)$ denote the geodesic normal to Γ at p. Consider the map

$$f : \cup\Gamma_i \ni p \mapsto N(p) \cap \ell \subset \ell \tag{3.19}$$

(that is, f takes p to the pair of antipodal points on ℓ that are determined by intersecting ℓ with the geodesic normal to Γ at p). Let $\{p_i\}_{i=1,\ldots,k}$ be the smallest set of points in $\cup\Gamma_i$ that separates Γ with the additional property that f maps all of the 'p_i's to the same pair of antipodal points, which we call (x, x'). Then every connected component of Σ which contains one of the points p_i is part of a sphere centered at x (and x').

Definition 3.4. Two geodesics in $\mathbf{H}^n$ are *parallel* if they do not intersect but for any positive ϵ there is a point on each line such that the distance between the two points is less than ϵ.

Definition 3.5. Two geodesics in $\mathbf{H}^n$ are *disjoint* if they do not intersect and are not parallel.

Lemma 3.12. Consider a stable double bubble of revolution $\Sigma \subset \mathbf{H}^n, n \geq 3$ with generating curve Γ and axis of symmetry ℓ as in the previous lemma. We let $\bar{\ell}$ denote $\ell \cup \{-\infty, \infty\}$. At each $p \in \cup\Gamma_i$ we again let $N(p)$ denote the geodesic normal to Γ at p, and we consider the map

$$f : \cup\Gamma_i \to \{0, 1\} \times \bar{\ell},$$

defined so that

- if $N(p)$ intersects ℓ then f maps p to $(0, N(p) \cap \ell)$,
- if $N(p)$ is parallel ℓ then f maps p to $\pm\infty$, depending on which end of ℓ comes arbitrarily close to $N(p)$,

- if $N(p)$ and ℓ are disjoint, there is a unique line ℓ' which intersects both ℓ and $N(p)$ orthogonally, and f maps p to $(1, \ell' \cap \ell)$.

Let $\{p_i\}_{i=1,\ldots,k}$ be the smallest set of points in $\cup\Gamma_i$ that separates Γ with the additional property that f maps all of the 'p_i's to the same point, which we call x. Then every connected component of Σ which contains one of the points p_i is part of

- a sphere centered at the second coordinate of x when x is of the form $(0, q)$ with $q \in \ell$,
- part of a horosphere centered at one end of ℓ if $x = \pm\infty$
- part of a hyposphere intersecting ℓ orthogonally if x is of the form $(1, q)$ with $q \in \ell$.

In the hypospherical case, the hyposphere in question is generated by revolving a hypocircle that is a constant distance from the perpendicular to ℓ through q about ℓ.

3.2.2 Criteria for standardness

Given enough connectedness, the characterization of stable bubbles appearing above provides the geometrical engine that proves the double bubble theorem in constant curvature space forms. The idea is that the propositions above limit the possibilities for a non-standard double bubble with enough connectivity to be among some set of double bubbles that we can apply some volume and perimeter preserving variation to that brings about a violation of regularity. (It is the role of the Fundamental Instability Argument to carrying out this final critical step.)

Corollary 3.4 (Structure Summary). In $\mathbf{R}^n$, $\mathbf{S}^n$, or $\mathbf{H}^n$, any Γ_i (as above) which can be disconnected by removing one point in its interior is a constant curvature arc which, if completed, would hit the axis of rotation ℓ orthogonally. In $\mathbf{R}^n$ it is a piece of a circle; similarly in $\mathbf{S}^n$; and in $\mathbf{H}^n$ it is a piece of a circle, horocycle, or hypercircle.

Adapting an argument from the William's honors thesis of Joel Foisy [**?**] about on double bubbles in $\mathbf{R}^3$, Cotton and Freeman showed that an area minimizing double bubble in $\mathbf{H}^n$ intersects its axis

of symmetry. With the addition of this final piece, Cotton and Freeman then had the right tools to prove:

Criterion for standardness in $\mathbf{H}^n$. An area minimizing double bubble in $\mathbf{H}^n$ in which both regions are connected must be the standard double bubble.

Criterion for standardness in $\mathbf{S}^n$. An area minimizing double bubble in $\mathbf{S}^n$ in which both regions and the exterior are connected must be the standard double bubble.

Chapter 4

s3h3

The culmination of the previous chapter was that

- If both enclosed regions and the exterior are connected, an area minimizing double bubble in $\mathbf{S}^n$ is standard.

- If both enclosed regions are connected, an area minimizing double bubble in $\mathbf{H}^n$ must be the standard double bubble.

The *other* culmination of the previous chapter (which took place in an earlier section) was that we have a criterion for connectivity, Proposition 3.2. That proposition said that if

$$2I(\frac{v}{2}) + I(w) + I(v+w) > 2\mathit{\Pi}(v,w) \tag{4.1}$$

then the region containing volume v is connected. We can check this criterion by plotting

$$2I\left(\frac{v}{2}\right) + I(w) + I(v+w) \tag{4.2}$$

and

$$2\tilde{\mathit{\Pi}}(v,w), \tag{4.3}$$

(where $\tilde{\Pi}(v, w)$ gives the surface area of a standard double bubble containing volumes v and w, and so an upper bound on $\Pi(v, w)$) and comparing the plots to see which function is bigger. In fact, both 4.2 and 4.3 are increasing functions in both variables (in $\mathbf{H}^3$; in $\mathbf{S}^3$, they are only increasing on a certain range). Where the functions are increasing, there won't be 'jumps' between plotted points. Hence to plot these functions is to prove the double bubble theorem in $\mathbf{S}^3$ and $\mathbf{H}^3$, modulo asymptotic analysis to show that the graphs of 4.2 and 4.3 continue to bear the correct relationship to each other outside of the range of the plot.

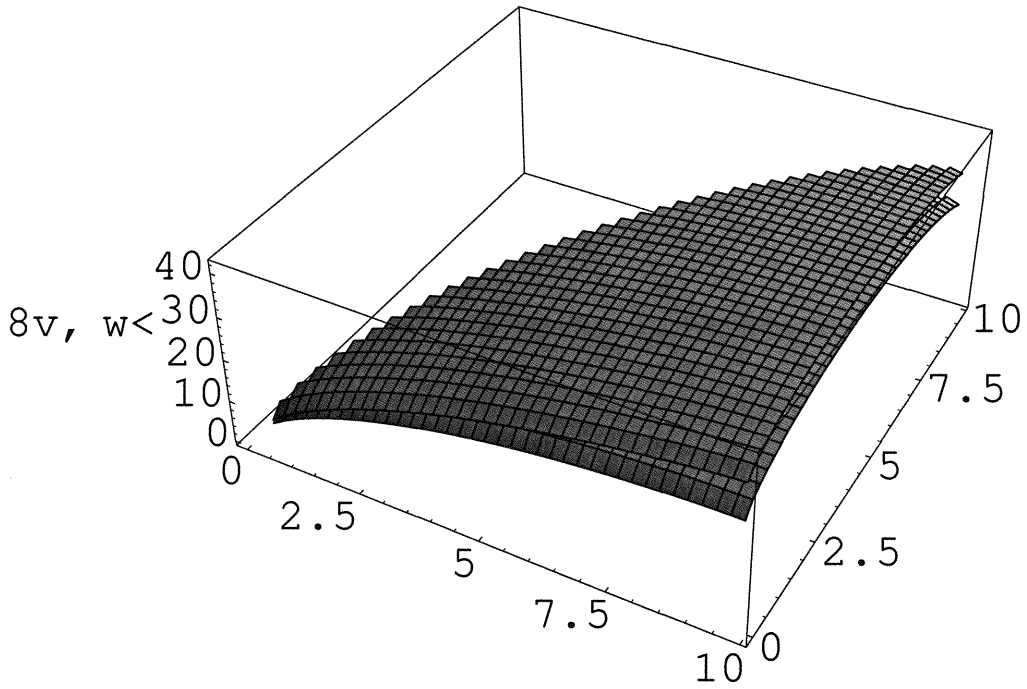


Figure 4.1: This plot proves connectivity in $\mathbf{S}^3$.

Theorem 4.1. The standard double bubble is the least-area way to partition $\mathbf{S}^3$ into three regions.

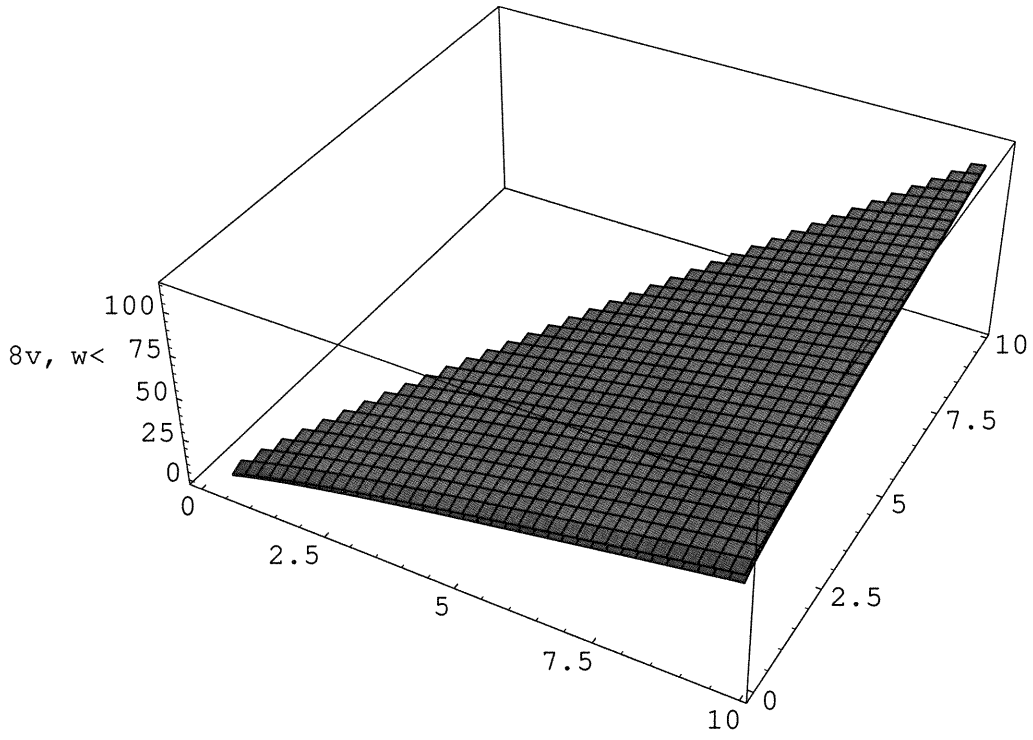


Figure 4.2: This plot is a little harder to read, but nonetheless, exhibits the bound we were looking for, proving connectivity in $\mathbf{H}^3$ for a finite range of volumes.

"Proof". Pending asymptotic analysis at zero, where the functions plotted are equal, the fact that the plots show that $2I(\frac{v}{2}) + I(w) + I(v+w) > 2\tilde{I}(v,w)$ on the domain where $v < |\mathbf{S}^3|/2$ and $w \leq v$ proves the claim, because the functions are increasing an this domain so the top function can not drop below the bottom function in between the mesh of the plot, and this domain is a general one in the sense that all we require is that one of the volumes takes up less than half of the total volume of the space and be larger than one of the other volumes; hence, by switching labels the properties of the given domain are representative of the properties of all volume pairs. QEF

Theorem 4.2. The standard double bubble is the least-area way to enclose and separate two volumes v and w in $\mathbf{H}^3$.

"Proof". The logic of this proof is the same as the proof of Theorem 4.1. Note that here, to finish the proof, not only must we study asymptotics at zero, but also through an entire quarter circle of directions heading out towards infinity. QEF

We don't have time to do the asymptotic analysis in this thesis. We satisfy ourselves with showing that the plotted formulas are correct.

4.1 Area and Volume Formulas

We can get a simple upper bound on $\mathit{\Pi}$ by using $\tilde{\mathit{\Pi}}$, the area of a standard double bubble, which will of course be greater than or equal to the area of the least-area double bubble. In this section we find parameterizations of the area and volumes of the standard double bubble and spheres in terms of the relevant radii (or curvatures, in the hyperbolic case). Using these parameterizations, we found I and $\tilde{\mathit{\Pi}}$ numerically. In order to create Figures 4.1 and 4.2 it was necessary to take this roundabout path, because no simple closed form for the relevant functions in terms of volumes is known (or likely to be found).

The formulas is this section are derived using spherical and hyperbolic trigonometry as found without explaination in [40]; for the explaination see Thurston [55, Section 2.4].

4.1.1 Spheres

Here we present the standard results on volume and surface area of spheres in $\mathbf{S}^3$ and $\mathbf{H}^3$.

Remark 4.1. The formulas for surface area of a sphere of radius r in $\mathbf{S}^3$ and $\mathbf{H}^3$ are

$$A_{\mathbf{S}^3} = 4\pi \sin^2 r$$

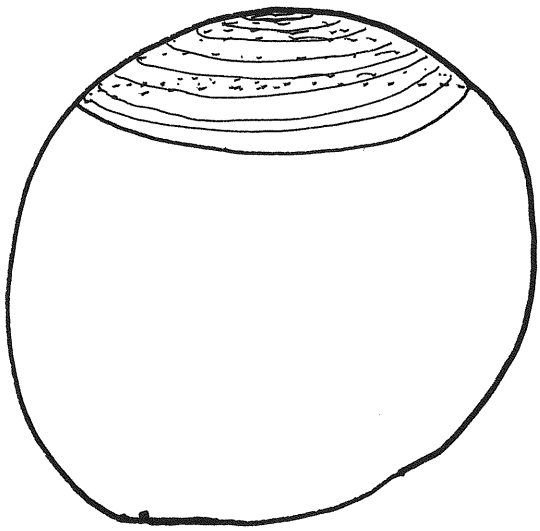


Figure 4.3: Spherical cap on sphere

$$A_{\mathbf{H}^3} = 4\pi \sinh^2 r$$

The volume formulas for a ball of radius r in $\mathbf{S}^3$ and $\mathbf{H}^3$ are

$$V_{\mathbf{S}^3} = \pi(2r - \sin 2r)$$

$$V_{\mathbf{H}^3} = \pi(\sinh 2r - 2r)$$

Lemma 4.1. The mean curvatures, $\frac{dA}{dV}$, of spheres of radius r in $\mathbf{S}^3$ and $\mathbf{H}^3$ are

$$(\mathbf{S}^3) \qquad 2\cot r \tag{4.4}$$

$$(\mathbf{H}^3) \qquad 2\coth r. \tag{4.5}$$

Proof. This is a special case of the computation appearing in Lemma 3.8. QEF

4.1.2 Spherical Caps

Even though it seems out of line with the name of the chapter, in this section we will prove the results we want in the Euclidean setting and then state the $\mathbf{S}^3$ and $\mathbf{H}^3$ versions as corollaries.

Proposition 4.1. In Euclidian space the volume of a spherical cap belonging to a sphere of radius r

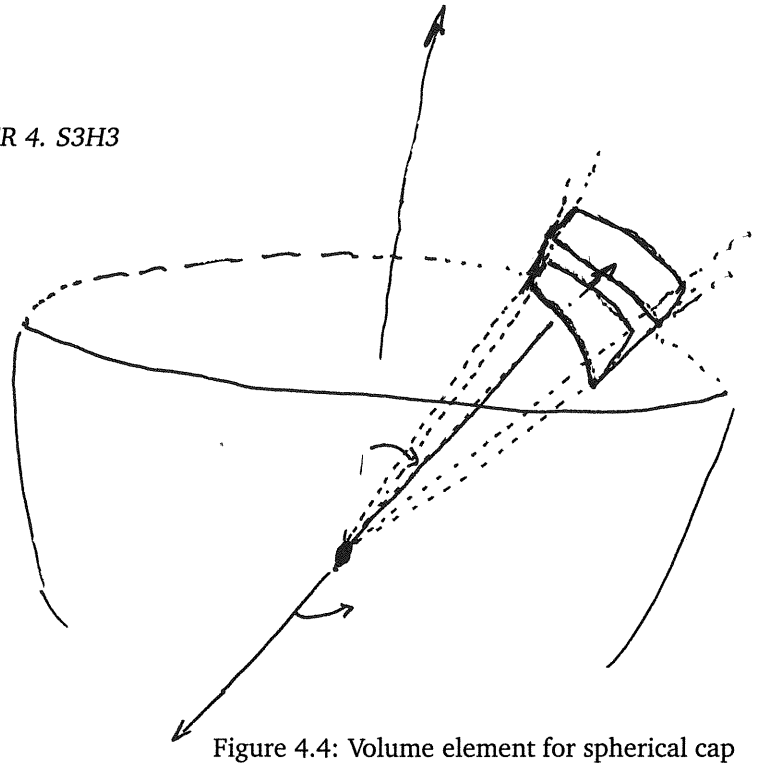


Figure 4.4: Volume element for spherical cap

and subtended by an angle of $2\varphi_0$ is given by

$$\int_0^{2\pi} d\theta \int_0^{\varphi_0} d\varphi \int_{\frac{r\cos\varphi_0}{\cos(\varphi)}}^{r} d\rho\, \rho^2 \sin\varphi$$

Proof. Spherical coordinates in $\mathbf{R}^3$

$$\begin{aligned} x &= \rho\cos\theta\sin\varphi \\ y &= \rho\sin\theta\sin\varphi \\ z &= \rho\sin\varphi \end{aligned} \tag{4.6}$$

have an associated Jacobian of $\rho^2\sin\varphi$. We do the integral in the standard way, *around* the sphere, *through* all the azimuthal angles, and *from* the disk that bounds the cap to the surface of the sphere (Figure 4.4). For this parameterization, we choose to have φ run *down* from the north pole of the sphere, since this makes the formulas come out simpler.

A little bit of trig gives the correct bounds for the integral. Actually, given the strategy above, the only question is what the lower bound is on the 'ρ' integral. (See Figure 4.5.)

We want to solve for ρ_0. The distance from the diameter d to the point p is $\rho_0\sin(\varphi)$. The distance from the center of the sphere to the base of the cap is $r\cos\varphi$. By the pythagorean theorem, we have

$$\rho_0^2\sin^2(\varphi) + r^2\cos^2\varphi_0 = \varphi_0^2.$$

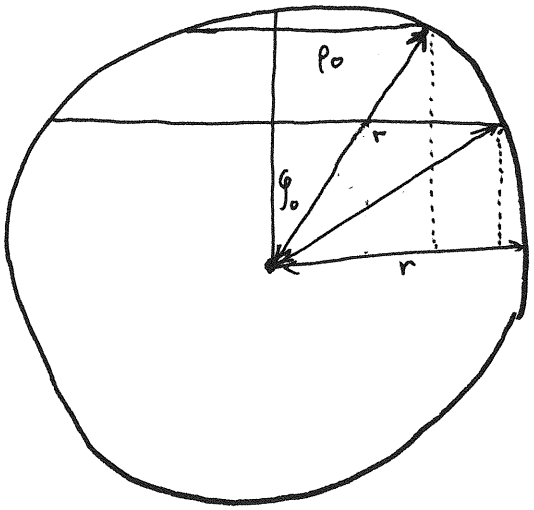


Figure 4.5: Find Lower Bound For ρ Using Trig

Hence,

$$\rho_0^2(1-\sin^2(\varphi)) = r^2\cos^2\varphi_0,$$

so

$$\rho_0 = \sqrt{\frac{r^2\cos^2\varphi_0}{1-\sin^2(\varphi)}} = \frac{r\cos\varphi_0}{\cos(\varphi)}$$

QEF

Corollary 4.1. The corresponding formulas in $\mathbf{S}^3$ and $\mathbf{H}^3$ are

$$(\mathbf{S}^3) \qquad \int_0^{2\pi} d\theta \int_0^{\varphi_0} d\varphi \int_{\tan^{-1}(\frac{\tan r\cos\varphi_0}{\cos(\varphi)})}^{r} d\rho\ \sin^2(\rho)\sin\varphi \tag{4.7}$$

$$(\mathbf{H}^3) \qquad \int_0^{2\pi} d\theta \int_0^{\varphi_0} d\varphi \int_{\tanh^{-1}(\frac{\tanh r\cos\varphi_0}{\cos(\varphi)})}^{r} d\rho\ \sinh^2(\rho)\sin\varphi. \tag{4.8}$$

Proof. For spherical coordinates in these spaces replace ρ in the parameterization of Equations 4.6 by $\sin\rho$ or $\sinh\rho$ respectively, and use spherical or hyperbolic trigonometry in Figure 4.5. QEF

Corollary 4.2. (Formulas for Volume of Spherical Caps) The volume of a spherical cap of radius r that is subtended by an angle $2\varphi_0$ is

$$(\mathbf{S}^3) \qquad -\pi(\tan^{-1}(\cos\varphi_0\tan r) - r\cos\varphi_0 + (-1+\cos\varphi_0)(r-\cos r\sin r)) \tag{4.9}$$

$$(\mathbf{H}^3) \qquad \pi(-r+\tanh^{-1}(\cos\varphi_0\tanh r) - \cos\varphi_0\cosh r\sinh r + \cosh r\sinh r))). \tag{4.10}$$

Proof. We did the integrals symbolically in *Mathematica*. QEF

Proposition 4.2. In Euclidian space the surface area of a spherical cap belonging to a sphere of radius r and subtended by an angle of $2\varphi_0$ is given by

$$\int_0^{2\pi} d\theta \int_0^{\varphi_0} d\varphi \, r^2 \sin\varphi$$

Proof. The equation for a hemisphere of radius r is $f(x,y) = \sqrt{r^2 - x^2 - y^2}$. The area integral is

$$\int_{\text{projection}} \sqrt{1 + \left(\frac{df}{dx}\right)^2 + \left(\frac{df}{dx}\right)^2}\, dxdy.$$

Calculating the derivatives and subsitituting in polar coordinates

$$\begin{aligned} x &= \rho\cos\theta \\ y &= \rho\sin\theta, \end{aligned} \tag{4.11}$$

this becomes

$$\int_0^{2\pi} d\theta \int_0^{\rho_0} d\rho \, \rho \sqrt{1 + \frac{\rho^2}{r^2 - \rho^2}},$$

From Figure 4.5, ρ_0 is seen to be $r\sin\varphi_0$. The following sequence of transformations then gives the stated form:

$$\begin{aligned} \int_0^{2\pi} d\theta \int_0^{r\sin\varphi_0} d\rho \, \rho \sqrt{1 + \frac{\rho^2}{r^2 - \rho^2}} &= \int_0^{2\pi} d\theta \int_0^{r\sin\varphi_0} d\rho \, \rho \sqrt{\frac{r^2}{r^2 - \rho^2}} \\ &= \int_0^{2\pi} d\theta \int_0^{r\sin\varphi_0} d\rho \, \rho \frac{1}{\sqrt{1 - (\frac{\rho}{r})^2}} \end{aligned} \tag{4.12}$$

(putting $\bar\rho = \rho/r$, we note that $d\bar\rho = \frac{1}{r}d\rho$, and that $\rho = r\sin\varphi_0$ implies $\bar\rho = \sin\varphi_0$, so 4.12 can be written:)

$$= \int_0^{2\pi} d\theta \int_0^{\sin\varphi_0} d\bar\rho \, r^2 \bar\rho \frac{1}{1 - \bar\rho^2} \tag{4.13}$$

(4.14)

(in general, $\bar\rho = \sin\varphi$, so $d\varphi = d(\sin^{-1}\bar\rho) = \frac{1}{\sqrt{1-\bar\rho^2}}d\bar\rho$, and we note that if $\rho = \rho_0 = r\sin\varphi_0$, then $\varphi = \varphi_0$, so we end with this change of variables:)

$$= \int_0^{2\pi} d\theta \int_0^{\varphi_0} d\varphi \, r^2 \sin\varphi. \tag{4.15}$$

QEF

Corollary 4.3. The corresponding formulas in $\mathbf{S}^3$ and $\mathbf{H}^3$ are

$$(\mathbf{S}^3) \qquad \int_0^{2\pi} d\theta \int_0^{\varphi_0} d\varphi \ \sin^2 r \sin\varphi \tag{4.16}$$

$$(\mathbf{H}^3) \qquad \int_0^{2\pi} d\theta \int_0^{\varphi_0} d\varphi \ \sinh^2 r \sin\varphi \tag{4.17}$$

Proof. For spherical coordinates in these spaces replace ρ in the parameterization of Equations 4.11 by $\sin\rho$ or $\sinh\rho$ respectively, and use spherical or hyperbolic trigonometry in Figure 2. QEF

Corollary 4.4. (Formulas for Area of Spherical Caps) The area of a spherical cap of radius r subtended by an angle $2\varphi_0$ is

$$(\mathbf{S}^3) \qquad 2\pi \sin^2 r(1-\cos\varphi_0) \tag{4.18}$$

$$(\mathbf{H}^3) \qquad 2\pi \sinh^2 r(1-\cos\varphi_0). \tag{4.19}$$

Proof. These integrals can be done in ones head. QEF

4.1.3 Standard Double Bubbles

Since each standard double bubble in $\mathbf{S}^3$ and $\mathbf{H}^3$ is symmetric about a geodesic line, it is convenient to compute area and volume of a standard double bubble by considering the revolution of some generating curve consisting of 3 circular arcs meeting at 120°. Figure 4.6 represents an arbitrary generating curve (and its reflection about the line of symmetry) labelled with relevant measurements.

Spherical Space From Figure 4.6 and repeated application of an identity of spherical trigonometry (see Figure 4.7) we have

$$\sin\varphi_1 \quad = \quad \frac{\sin x}{\sin r_1};$$

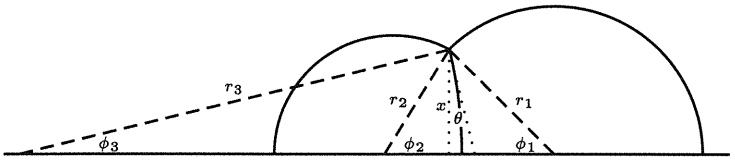


Figure 4.6: Some important geometrical features of the generating curve for a standard double bubble.

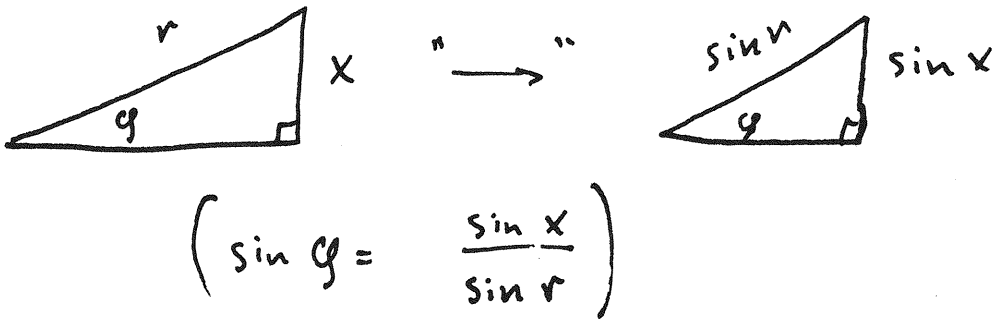


Figure 4.7: Sphere sine relation

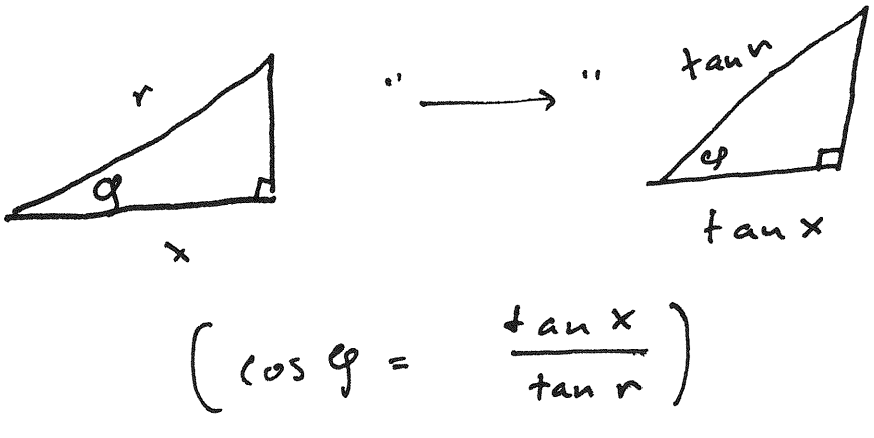


Figure 4.8: Sphere tantancos

$$\begin{aligned} \sin\varphi_2 &= \frac{\sin x}{\sin r_2}; \\ \sin\varphi_3 &= \frac{\sin x}{\sin r_3}. \end{aligned}$$

Another identity (see Figure 4.8) yields

$$\begin{aligned} \tan x &= \tan r_1 \cos(\frac{\pi}{6} - \theta); \\ \tan x &= \tan r_2 \cos(\frac{\pi}{6} + \theta); \end{aligned}$$

or

$$\begin{aligned} \tan x &= \tan r_1 (\frac{\sqrt{3}}{2}\cos\theta + \frac{1}{2}\sin\theta); \\ \tan x &= \tan r_2 (\frac{\sqrt{3}}{2}\cos\theta - \frac{1}{2}\sin\theta). \end{aligned}$$

Solving for θ, we get

$$\theta = \tan^{-1}(\frac{\sqrt{3}(\tan r_2 - \tan r_1)}{\tan r_1 + \tan r_2}).$$

We can plug this into the first of the two preceding equations above and rearrange terms to get

$$x = \tan^{-1}(\tan r_1 \cos(\frac{\pi}{6} - \tan^{-1}(\frac{\sqrt{3}(\tan r_2 - \tan r_1)}{\tan r_1 + \tan r_2}))),$$

which we can put into the first triple of equations to obtain a (complicated) expression for the 'φ's in terms of r_1 and r_2:

$$\varphi_1 = \sin^{-1}\left(\frac{\sin(\tan^{-1}(\tan r_1 \cos(\frac{\pi}{6} - \tan^{-1}(\frac{\sqrt{3}(\tan r_2 - \tan r_1)}{\tan r_1 + \tan r_2}))))}{\sin r_1}\right);$$

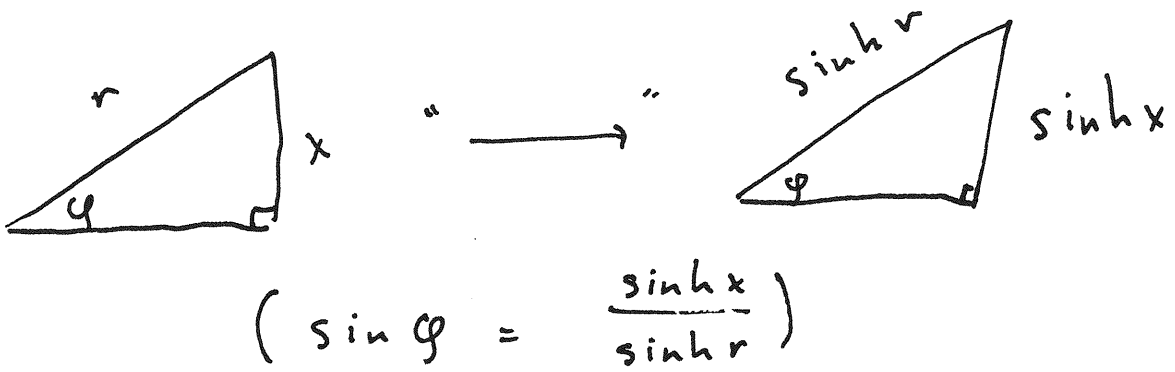


Figure 4.9: Hyperbolic sine relation

$$\varphi_2 = \sin^{-1}\left(\frac{\sin(\tan^{-1}(\tan r_1 \cos(\frac{\pi}{6} - \tan^{-1}(\frac{\sqrt{3}(\tan r_2 - \tan r_1)}{\tan r_1 + \tan r_2}))))}{\sin r_2}\right);$$

$$\varphi_3 = \sin^{-1}\left(\frac{\sin(\tan^{-1}(\tan r_1 \cos(\frac{\pi}{6} - \tan^{-1}(\frac{\sqrt{3}(\tan r_2 - \tan r_1)}{\tan r_1 + \tan r_2}))))}{\sin r_3}\right).$$

Note that there are only two relevant parameters here. Since pressures add, we have r_3 in terms of r_1 and r_2:

$$2\cot r_3 = 2\cot r_1 - 2\cot r_2,$$

or

$$r_3 = \cot^{-1}(\cot r_1 - \cot r_2).$$

Hyperbolic Space We derive similar formulas for the angles φ_i, but this time using trigonometric identities in hyperbolic space (see Figures 4.9 and 4.10):

$$\begin{aligned}
\sin\varphi_i &= \frac{\sinh x}{\sinh r_i}\ (i = 1, 2, 3); \\
\tanh x &= \tanh r_1 \cos(\frac{\pi}{6} - \theta); \\
\tanh x &= \tanh r_2 \cos(\frac{\pi}{6} + \theta).
\end{aligned}$$

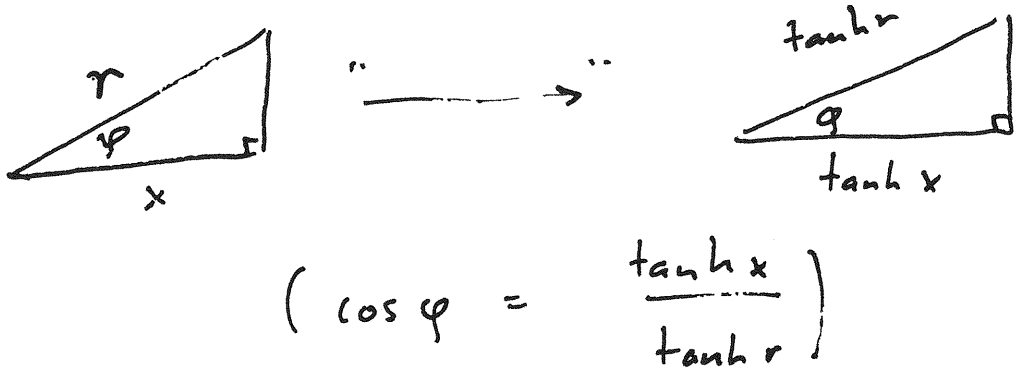


Figure 4.10: Hyperbolic tantancos

Solving for φ_i (i = 1, 2, 3) yields

$$\varphi_i = \sin^{-1}\left(\frac{\sinh(\tanh^{-1}(\tanh r_1 \cos(\frac{\pi}{6} - \tan^{-1}(\frac{\sqrt{3}(\tanh r_2 - \tanh r_1)}{\tanh r_1 + \tanh r_2}))))}{\sinh r_i}\right)$$

So far this derivation has been analogous to the work we did in $\mathbf{S}^3$. But in hyperbolic space we have already mentioned the fact that not all hypersphere-like submanifolds in hyperbolic space have well-defined radii. We work around this by computing area and volume of double bubbles in $\mathbf{H}^3$ in terms of curvature. To do this, we derive the area and volume formulas as before, and convert formulas using the relation:

$$k = \coth r \tag{4.20}$$

The wonderful fact is that analytic continuation implies that our new formulas yield the desired areas and volumes.

We can now compute φ_i in terms of k_1 and k_2 (the curvatures of the two outer caps). Here, the curvature constraint gives, $k_3 = k_1 - k_2$.

$$\varphi_1 = \sin^{-1}\left(\frac{\sqrt{1-\frac{1}{k_1^2}}\cos(\frac{\pi}{6}-\tan^{-1}(\frac{\sqrt{3}(k_1-k_2)}{k_1+k_2}))}{\sqrt{1-\frac{\cos(\frac{\pi}{6}-\tan^{-1}(\frac{\sqrt{3}(k_1-k_2)}{k_1+k_2}))}{k_1^2}}}\right)$$

$$\varphi_2 = \sin^{-1}\left(\frac{k_2\sqrt{1-\frac{1}{k_2^2}}\cos(\frac{\pi}{6}-\tan^{-1}(\frac{\sqrt{3}(k_1-k_2)}{k_1+k_2}))}{k_1\sqrt{1-\frac{\cos(\frac{\pi}{6}-\tan^{-1}(\frac{\sqrt{3}(k_1-k_2)}{k_1+k_2}))}{k_1^2}}}\right)$$

$$\varphi_3 = \sin^{-1}\left(\frac{(k_1-k_2)\sqrt{1-\frac{1}{(k_1-k_2)^2}}\cos(\frac{\pi}{6}-\tan^{-1}(\frac{\sqrt{3}(k_1-k_2)}{k_1+k_2}))}{k_1\sqrt{1-\frac{\cos(\frac{\pi}{6}-\tan^{-1}(\frac{\sqrt{3}(k_1-k_2)}{k_1+k_2}))}{k_1^2}}}\right)$$

Remark 4.2. Now we can compute the volumes and surface area of standard double bubbles, in terms of the radii or curvatures of the outer caps, by substituting the formulas we have just derived in Corollary 4.2 and Corollary 4.4. Refering to Figure 4.6, we observe that the angle φ_1 is either acute or obtuse depending on the relative size of the two bubbles; the transition between regimes is when φ_1 is 90°. Tracing back through the equations, we notice that this happens when

$$\frac{\tan(r_2)-\tan(r_1)}{\tan(r_2)+\tan(r_1)} = 1/3 \quad (\text{in } \mathbf{S}^3)$$

and when

$$\frac{k_1-k_2}{k_1+k_2} = 1/3 \quad (\text{in } \mathbf{H}^3)$$

We must make adjustments to φ_1 in our computations accordingly. Specifically, if $\varphi_1 < \frac{\pi}{2}$, the formulas return values corresponding to a cap that is larger than a hemisphere, which is what we desired, otherwise they return values corresponding to a cap smaller than a hemisphere, so we have to use $\pi - \varphi_1$ in place of φ_1.

Definition 4.1. In light of the last remark, we define $\widehat{\varphi}_1$ to be equal to φ_1 if $\varphi_1 < \frac{\pi}{2}$ and to be $\pi - \varphi_1$ otherwise.

Now, the following lemma summarizes how we compute area and volumes for the standard double bubble.

Lemma 4.2. The volume of the first bubble is the volume of the spherical cap in the sphere of radius r_1 subtended by an angle $2\widehat{\varphi}_1$, plus the volume of the spherical cap in the sphere of radius r_3 subtended by an angle $2\varphi_3$. The volume of the second bubble is the volume of the spherical cap in the sphere of radius r_2 subtended by an angle $2(\pi - \varphi_2)$, minus the volume of the spherical cap in the sphere of radius r_3 subtended by angle $2\varphi_3$.

Proof. For each bubble component, we compute the volume and area of the spherical cap which separates it from the exterior. For area of the double bubble, we simply add these areas plus the area of the separating cap. To compute the volume of each bubble component we have to include the volume contained within the separating cap in the smaller region, and exclude it from the larger region. QEF

4.2 The Curvature Conjecture

The inequality we have been studying in this chapter is

$$\hat{F}(v,w) = 2I(\frac{v}{2}) + I(w) + I(v+w) - 2\tilde{\Pi}(v,w) > 0 \tag{4.21}$$

$\tilde{\Pi}(v,w)$ denotes the surface area of the standard double bubble enclosing the volumes v and w.

We can perturb the definition of F_k of Corollary 6.2 in the same way to deduce that if

$$\hat{F}_k(v,w) = kI(\frac{v}{k}) + I(w) + I(v+w) - 2\tilde{\Pi}(v,w) > 0 \tag{4.22}$$

then V has less than k components.

Lemma 3.8 tells us that $kI(\frac{v}{k})$ is increasing in k. Therefore there is some largest k for which the Basic Estimate Lemma 3.9 (with $x = \frac{v}{k}$) applies *i.e.* for this k we have $2\Pi(v,w) = I(w) + \tilde{k}I(\frac{v}{\tilde{k}}) + I(v+w)$. A possibly even greater 'k', which we will call $\tilde{k}$ satisfies

$$2\tilde{\Pi}(v,w) = I(w) + \tilde{k}I(\frac{v}{\tilde{k}}) + I(v+w). \tag{4.23}$$

This defines $\tilde{k}$ implicitly as a function of v and w.

In light of the fact that mean curvature is proportional to the derivative of area with respect to volume, differentiating Equation 4.23 gives

$$\frac{d(\tilde{k}I(\frac{v}{k}))}{dw} = kH_{v/k} - H_w - H_{w+v} \tag{4.24}$$

$$\frac{d(\tilde{k}I(\frac{v}{k}))}{dv} = kH_{v/k} - H_{w+v}. \tag{4.25}$$

(*N. B.* Upon differentiating the surface area of the standard double bubble with respect to one of the volumes, we only pick up the curvature of the outer cap because (a) the standard double bubble is stable under small variations; and (b) therefore all variations must give the same dA/dv or some combination could reduce perimeter, contradicting (a); (c) a variation that raises a small bump on the outer cap picks up its curvature.)

The curvature conjecture says that each of 4.24 and 4.25 is signed. Specifically, in the most important case $k < 2$, we have:

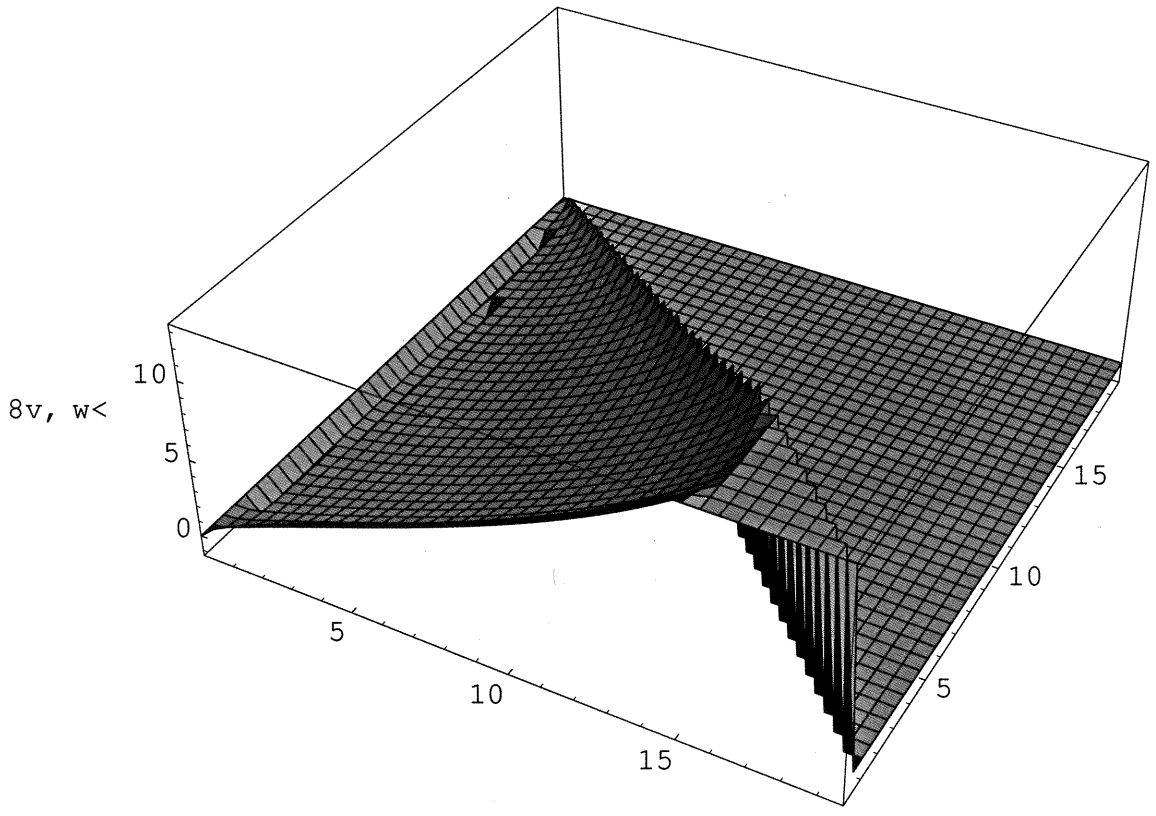
Conjecture 4.1. Curvature Conjecture In $\mathbf{R}^n$, $\mathbf{S}^n$, or $\mathbf{H}^n$, the various mean curvatures associated with standard double bubbles satisfy the following inequalities, where we H_x denote the curvature of a sphere of volume x and we let $H_{x,y}$ denote the curvature of the the outer cap of the 'x component' of a standard double bubble containing volumes x and y

$$H_{v,w} > \frac{H_{v/2} + H_{w+v}}{2}.$$

$$H_{w,v} < \frac{H_{v/2} + H_{w+v}}{2}.$$

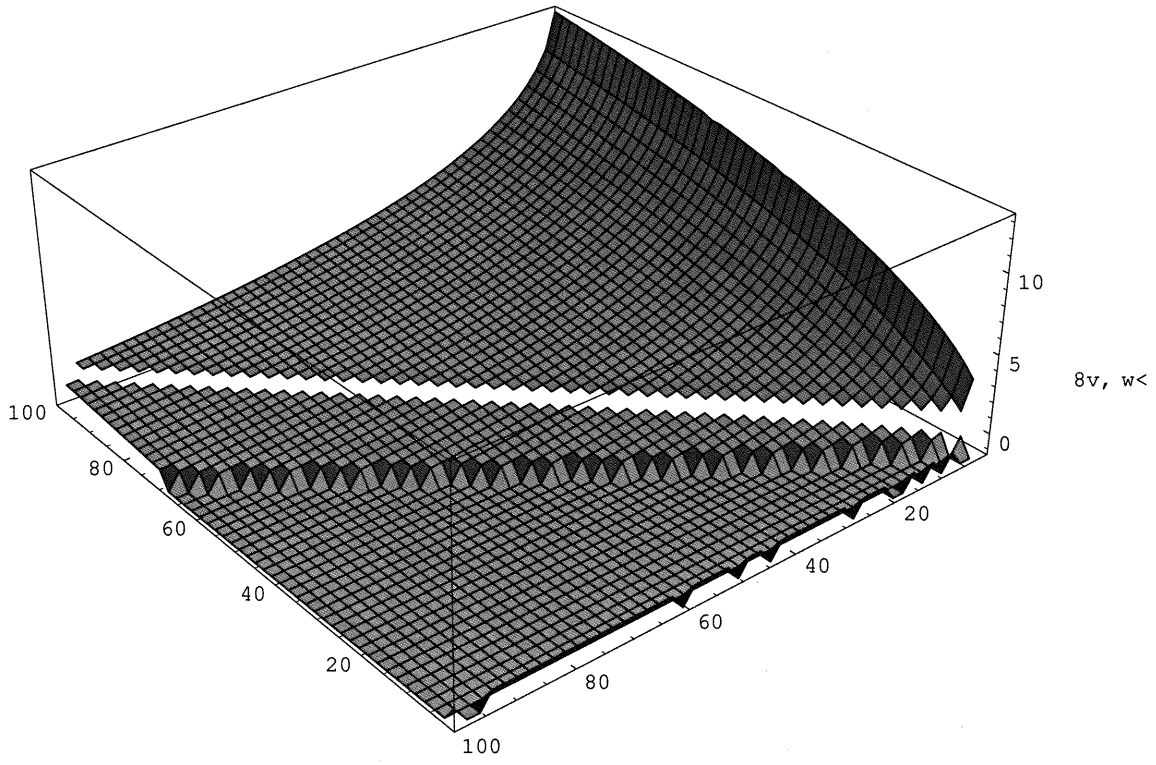
If this conjecture is true, then $\hat{F}$ is a decreasing function one of the variables and an increasing function in the other, and so plots of $\hat{F}$ will not have jumps in between the mesh of the plot. Plots and analysis suggest that the curvature conjecture holds in $\mathbf{R}^n$, although the analysis part is not complete. In $\mathbf{R}^n$, the volumes can be normalized so that one of the volumes is equal to one. David Futer then showed that in the case that the other volume is bigger than one. The proof uses the fact that the standard double bubble is invariant under inversion in the sphere that separating cap is a piece of.

Since we already know that the component functions of $\hat{F}$ are increasing, we are probably better off studying them. At present, plots of the function $\hat{F}$ are useful because they show (non-rigorously) what the gaps between the functions plotted in Figures 4.1 and 4.2 look like.



The function value is lowered artificially on region where graph is negative for ease of viewing.

Figure 4.11: The Hutchings function $\hat{F}(v, w)$ in $\mathbf{S}^3$.



The function value is lowered artificially on regions where graph is negative for ease of viewing; the gap on the diagonal is due to a singularity in our formulas that doesn't show up in the proof-plots for some reason. Cotton and Freeman already treated the diagonal in detail anyway.

Figure 4.12: The Hutchings function $\hat{F}(v, w)$ in $\mathbf{H}^3$.

Chapter 5

gn

Charter. We prove that the least-area division of n-dimensional Gauss space into three equal pieces is given by the standard configuration of three $(n-1)$-dimensional hyperplanes meeting along an $(n-2)$-dimensional hyperplane at 120°.

5.1 Introduction

Let $\mathbf{G}^n$ denote $\mathbf{R}^n$ with the volume form weighted by Gaussian measure and area given by restricting this measure to $(n-1)$-dimensional subsets in the standard fashion (the area of an $(n-1)$-dimensional subset A is given by finding the derivative with respect to ϵ at zero of the volume of ϵ-tubes about A, so, for example, a hyperplane through the origin has total $(n-1)$-dimensional measure of $1/2\pi$). Borell [4], Sudakov-Tsirel'son [51] and others (cf. Barthe and Maurey [2] and the references therein) showed that the solution to the isoperimetric problem in this space (that is, the least-area division of $\mathbf{G}^n$ into two regions of prescribed volume) is given by an $(n-1)$-dimensional hyperplane. Our Theorem 5.1 will show that the least-area division of Gauss space into three pieces of equal size is given by a 'Y', that is, by three half-hyperplanes meeting each other at 120°along an

$(n-2)$-dimensional plane through the origin.

Theorem 5.1 solves a 'double bubble problem', the name given to the problem of enclosing and separating two regions of prescibed volume in a give space with the least area. This problem has recently been solved in $\mathbf{R}^3$ by Hutchings, Morgan, Ritoré and Ros ([23], see [34]).

The proof. A classical observation ascribed to Poincaré (Proposition 5.1) says that $\mathbf{G}^n$ can be realized as a limit of projections from $\mathbf{S}^N$ onto $\mathbf{R}^n$ as N goes to infinity and the radius of $\mathbf{S}^N$ grows. This reduces the problem to studying double bubbles in $\mathbf{S}^N$.

Hutchings' theory of the structure of area minimizing double bubbles in spheres [22] shows how to bound the number of components of minimizing bubbles that divide the $\mathbf{S}^N$ into three equal volumes. Using this connectivity result, an argument based on the that used in $\mathbf{R}^3$ [23, Section 6] shows that minimizing double bubbles must be standard. Taking limits of projections onto $\mathbf{S}^n$ then gives the result.

Acknowledgements Theorem 5.1 came out of conversations with Frank Morgan about a lecture by Antonio Ros (and the lecture notes [44]). Morgan provided lots of useful ideas about how to *prove* the theorem. Pat McDonald had a number of very helpful conversations with me on this topic.

5.2 The Geometry of Gauss Space

Proposition 5.1, attributed to Poincaré shows that n-dimensional Gauss space is the limit of projections onto $\mathbf{R}^n$ of uniform measure on spheres whose size and dimension grow in a prescibed way. The proof of Proposition 5.1 is based on that found in McKean [30]. We also point out that it is very easy to compute areas of hyperplanes and pieces of hyperplanes in $\mathbf{G}^n$ (Proposition 5.2), reducing this to computations with a one-dimensional Gaussian.

Proposition 5.1 (attrib. Poincaré). Gauss space of n dimensions is realized as the limit as $N \to \infty$ of projections of $\mathbf{S}^N(\sqrt{N})$ onto some fixed $\mathbf{R}^n$.

Proof. Because of how we have defined volumes and areas in $\mathbf{G}^n$, it suffices to show that if $\mathbf{x} = (x_1, ...x_N)$ is uniformly distributed on $S^N(\sqrt{N})$, then for fixed $n < \infty$ the probability that $\mathbf{x}$ is in the box $(\mathbf{a}, \mathbf{b}) = (a_1, b_1) \times \ldots \times (a_n, b_n) \subset \mathbf{R}^n$ is equal to the integral of n-dimensional Gaussian measure over this box. In symbols, what we wish to prove is:

$$\lim_{N\to\infty} P(\cap_{i=1}^n (a_i \leq x_i \leq b_i)) = \prod_{i=1}^n \int_{a_i}^{b_i} \frac{e^{-x_i^2/2}}{\sqrt{2\pi}} dx_i.$$

The probablity that $\mathbf{x}$ is in $(\mathbf{a}, \mathbf{b})$ is given by the integral of the area form dA induced by the volume form on $\mathbf{R}^{N+1}$ over pieces of the sphere that lie 'above' or 'below' $(\mathbf{a}, \mathbf{b})$, divided by the total measure of $\mathbf{S}^N$, *i.e.* the integral of dA over $\mathbf{S}^N$.

The form of the integral is given in three dimensions by parameterizing by the azimuthal angle φ, so if $R = \sqrt{N}$, then $x_3 = R\cos\varphi$, and $dx_3 = R\sin\varphi\, d\varphi$, which by trigonometry is $\sqrt{R^2 - x_3^2}$. We also have $dA = 2\pi R \sin\varphi\, d\varphi = 2\pi\sqrt{R^2 - x_3^2}\, d\varphi$, which is therefore just $2\pi\, dx_3$. We rotate so that $x = x_3$, so the area form is proportional to $(1 - \frac{x^2}{N})^0$.

By an analogous procedure in higher dimensions the probability that $\mathbf{x}$ lies above or below an interval (a, b) (that is, the probability that one of the dimesions x of $\mathbf{x}$ is in the range of values for which $\mathbf{x}$ might be in the box $(\mathbf{a}, \mathbf{b})$ that we specified, not taking into account what is going on in the other dimensions) works out to be

$$\frac{\int_a^b (\sqrt{N}^2 - x^2)^{(N-3)/2}\, dx}{\int_{-\sqrt{N}}^{\sqrt{N}} (\sqrt{N}^2 - x^2)^{(N-3)/2}\, dx} = \frac{\int_a^b (1 - x^2/N)^{(N-3)/2}\, dx}{\int_{-\sqrt{N}}^{\sqrt{N}} (1 - x^2/N)^{(N-3)/2}\, dx}. \tag{5.1}$$

Recall that $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$, so $e^{-x^2} = \lim_{n\to\infty}(1 - \frac{x^2/2}{n})^n = \lim_{n\to\infty}(1 - \frac{x^2}{n})^{n/2}$. It is easy to check that this is $\lim_{n\to\infty}(1 - \frac{x^2}{n})^{(n-3)/2}$. Therefore as $n \to \infty$, the ratio in Equation 5.1 approximates $\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$. The linearity of the integral now gives the stated form for the integral over a box. QEF

Proposition 5.2. The $(n-1)$-dimensional measure of a hyperplane that divides $\mathbf{G}^n$ into two pieces of any given ratio is the same for all n.

Proof. To prove this, we need to use the basic fact that an $(n-1)$-hyperplane has an $(n-1)$-dimensional Gaussian on it, whose total measure depends only on its distance from the origin and not on the dimension. This is a consequence of the fact that up to normalization, $\mathbf{G}^n = \mathbf{G}^{n-1} \times \mathbf{G}^1$. For any given volume ratio and any $n > 1$, a hyperplane P whose distance from the origin is the same as the distance from the origin of a point p in $\mathbf{G}^1$ that cuts $\mathbf{G}^1$ into pieces with the given ratio will cut each perpendicular hyperplane H into two pieces of the given ratio, and so it cuts $\mathbf{G}^n$ into two pieces of the given ratio. Since the total measure of this P depends only on its distance from the origin, this proves the claim. QEF

Corollary 5.1. The $(n-1)$-dimensional measure of three half-hyperplanes that divides $\mathbf{G}^n$ into three pieces of equal volume is the same for all $n > 2$.

Proof. Three half-hyperplanes through the origin have total $(n-1)$-dimensional measure $\frac{1}{\sqrt{2\pi}}\frac{3}{2}$. QEF

5.3 Area-minimizing division of $\mathbf{G}^n$

We now introduce some notation. We let $I_{\mathbf{G}^n}$ denote the isoperimetric profile of $\mathbf{G}^n$. By Proposition 5.2, the surface area of a standard single bubble is given by the same function in $\mathbf{G}^n$ and $\mathbf{G}^1$, and can be written down (fairly) explicitly: $I_{\mathbf{G}^n}(v) = I_{\mathbf{G}^1}(v) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, where x is defined by $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt = v$. We let $I\!I_{\mathbf{G}^n}$ denote the two-volume least-area function of $\mathbf{G}^n$.

We will denote the isoperimetric profile and two-volume least-area function of $\mathbf{S}^m$ by $I_{\mathbf{S}^m}$ and $I\!I_{\mathbf{S}^m}$ respectively. These can be thought of classically or by using a Hausdorff measure construction. Viewed the latter way, we can the realize the functions $I_{\mathbf{G}^n}$ and $I\!I_{\mathbf{G}^n}$ as projections of limits of $I_{\mathbf{S}^m}$ and $I_{\mathbf{S}^m}$.

Hutchings theory [22] says that if

$$F(v,w) = 2I_{\mathbf{S}^m}(v/2) + I_{\mathbf{S}^m}(w) + I_{\mathbf{S}^m}(v+w) - 2\Pi_{\mathbf{S}^m}(v,w) > 0 \tag{5.2}$$

then the region containing the volume v in an area-minimizing double bubble in $\mathbf{S}^m$ that contains volumes v and w is connected. As discussed in the previous two chapters

$$\tilde{F}(v,w) = 2I_{\mathbf{S}^m}(v/2) + I_{\mathbf{S}^m}(w) + I_{\mathbf{S}^m}(v+w) - 2\widetilde{\Pi}_{\mathbf{S}^m}(v,w) > 0, \tag{5.3}$$

where $\widetilde{\Pi}_{\mathbf{S}^m}(v,w)$ denotes the area of a standard double bubble containing volumes v and w, is also a condition for connectivity.

Taken together we now have enough tools to prove our theorem.

Theorem 5.1. For $n > 1$, the least-area division of n-dimensional Gauss Space $\mathbf{G}^n$ into three equal pieces consists of three hyperplanes, meeting at one hundred and twenty degrees along an $(n-2)$-dimensional plane through the origin.

Proof. In the case of three equal volumes in $\mathbf{S}^m(\sqrt{m})$, with $v = |\mathbf{S}^m(\sqrt{m})|$, the Hutchings criterion 5.3 becomes

$$2I_{\mathbf{S}^m}\left(\frac{v}{6}\right) + I_{\mathbf{S}^m}\left(\frac{v}{3}\right) + I_{\mathbf{S}^m}\left(\frac{2v}{3}\right) - 2\widetilde{\Pi}_{\mathbf{S}^m}\left(\frac{v}{3},\frac{v}{3}\right) > 0. \tag{5.4}$$

By symmetry, if this inequality holds, all of the regions in are connected.

By the considerations of Proposition 5.1, there is some N such that if $m > N$ then

$$2I_{\mathbf{G}^n}\left(\frac{1}{6}\right) + I_{\mathbf{G}^n}\left(\frac{1}{3}\right) + I_{\mathbf{G}^n}\left(\frac{2}{3}\right) - 2\widetilde{\Pi}_{\mathbf{G}^n}\left(\frac{1}{3},\frac{1}{3}\right) > 0 \tag{5.5}$$

is a criterion for connectivity in $\mathbf{S}^m$.

By the considerations of Proposition 5.2, this last criterion is the same as

$$2I_{\mathbf{G}^1}\left(\frac{1}{6}\right) + I_{\mathbf{G}^1}\left(\frac{1}{3}\right) + I_{\mathbf{G}^1}\left(\frac{2}{3}\right) - 2\left(\frac{1}{\sqrt{2\pi}}\frac{3}{2}\right) > 0, \tag{5.6}$$

which proves that Inequality 5.5 establishes a lower bound that is uniform in n. By symmetry, Inequality 5.6 is equivalent to

$$2I_{\mathbf{G}^1}\left(\frac{1}{6}\right) + 2I_{\mathbf{G}^1}\left(\frac{1}{3}\right) - 2\left(\frac{1}{\sqrt{2\pi}}\frac{3}{2}\right) > 0. \tag{5.7}$$

Since

$$\int_{-100}^{-1} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \approx .15866 < .16666\ldots = 1/6, \tag{5.8}$$

we have that $\frac{1}{\sqrt{2\pi}} e^{-(-1)^2/2} = \sqrt{\frac{e}{2\pi}} \approx .657745$ is a lower bound on $I_{\mathbf{G}^1}(1/6)$, and similarly since

$$\int_{-100}^{-1/2} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \approx .30854 < .33333\ldots = 1/3, \tag{5.9}$$

we have that $\frac{1}{\sqrt{2\pi}} e^{-(-1/2)^2/2} \approx .352065$ is a lower bound on $I_{\mathbf{G}^1}(1/3)$.

Combining these approximations we get

$$2I_{\mathbf{G}^1}\left(\frac{1}{6}\right) + 2I_{\mathbf{G}^1}\left(\frac{1}{3}\right) > 2 \cdot .65775 + 2 \cdot .35207 \approx 2.01964 > .7978 \approx 2\left(\frac{1}{\sqrt{2\pi}}\frac{3}{2}\right), \tag{5.10}$$

which proves that the Hutchings connectivity criterion is satisfied for equal volumes for spheres close enough to the Poincaré limit.

An argument like that used in $\mathbf{R}^3$ by Hutchings *et al.* proves that the minimizing enclosure for three equal volumes in these high-dimensional spheres is the standard double bubble; specifically, Cotton and Freeman ([10], or see Chapter 3) prove that in $\mathbf{S}^n$, an area minimizing double bubble that has both regions and the exterior connected must be standard.

By rotating and translating in $\mathbf{R}^{m+1}$ as necessary, we can recenter so that the origin is contained in determined by the $(m-2)$-dimensional singular set in the standard double bubble contains the origin. In addition we require that the first n dimensions of $\mathbf{S}^m$ line up for all m. These are the dimensions which become $\mathbf{G}^n$ in the limit of projections. Since the standard double bubble in $\mathbf{S}^m$ is defined as a surface of revolution generated by the standard double bubble in $\mathbf{S}^{m-1}$, it is easy to see in the case $n = 2$ that the limit of the projections of the intersection of the standard double bubble with gthe first two dimensions of $\mathbf{S}^m$ the standard double bubble is the Y-shaped division of $\mathbf{G}^2$. This division is minimizing because it lifts to a minimizer in each of the spheres $\mathbf{S}^N$ as these spheres tend to the Poincaré limit. The same argument is in fact true when n is greater than two, but it is easier to just observe that Y-shaped division of $\mathbf{G}^2$ lifts to the Y-shaped division of $\mathbf{G}^n$ and therefor the Y-shaped division of $\mathbf{G}^n$ must be minimizing. QEF

Remark 5.1. The proof shows that for large N, the least area division of $\mathbf{S}^N$ into three pieces of equal volume is given by a standard double bubble (in this case, three half-hyperspheres meeting along an $N-2$ sphere). The conjecture is that this is in fact the optimal division for all N.

Chapter 6

t2

Charter Characterize the perimeter minimizing double bubbles in the flat two-torus $\mathbf{T}^2$.

6.1 Introduction

Our Main Theorem (Section 6.7) shows that on the flat two-torus $\mathbf{T}^2$, the least-perimeter way to enclose and separate two prescribed areas is a double bubble of one of the six types of Figure 6.1: the standard planar double bubble Θ, the band-lens Φ, the symmetric two-component chain Ω, the double band Ξ, the double hexagon H (on the hexagonal torus only), or an octagon-square tiling X. We conjecture that the last of these never occurs as a minimizer.

Conjecture 6.1. The octagon-square tiling X is never perimeter minimizing.

If Conjecture 6.1 is correct, Figure 6.2 shows which types prevail for which prescribed areas for four different tori.

The proof of the main theorem separates all candidates into five classes, and shows that the special double bubbles of Figure 6.1 are the only ones that are potentially perimeter minimizing: Propo-

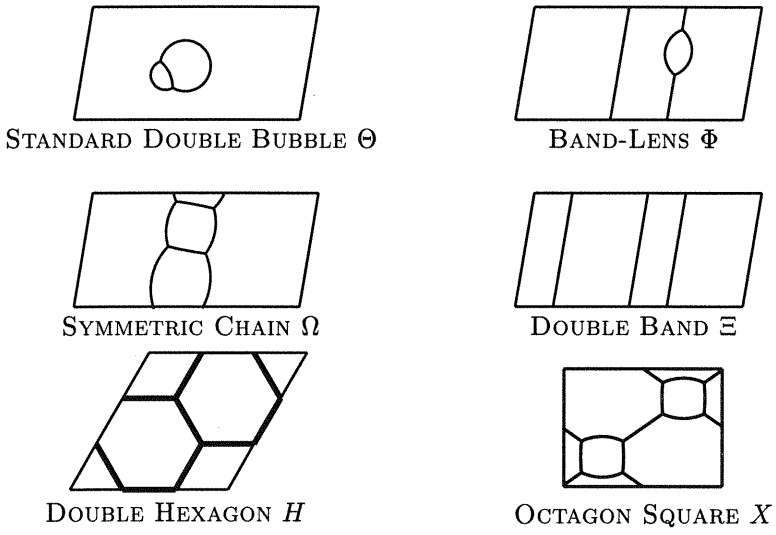


Figure 6.1: The six candidate perimeter- minimizing double bubbles on the flat two-torus.

sition 6.6 says that minimizers must be the double band or belong two one of four other topological classes; section 6.5.1 treats double bubbles in which all of the regions are contractible; section 6.5.2 treats double bubbles whose components taken together wrap around the torus, but whose components are all contractible; and section 6.5.3 treats double bubbles with one non-contractible region. The fourth class includes double bubbles that wrap around the torus in both directions. We call these candidates "tilings." Proposition 6.12 and 6.13 use the Euler characteristic of $\mathbf{T}^2$ and some basic facts about the regularity of minimizing bubbles to show that only two tilings are potential minimizers: the hexagon tiling H on the hexagonal torus, and (perhaps) an octagon-square tiling. When it occurs as a minimizer, H ties the double band Ξ for perimeter. Section 6.6 gives formulas for the perimeter and area of the four finalists. Figure 6.2 gives a computational plot of the minimizing type for given areas.

We have been unable to prove Conjecture 6.1, which would eliminate the octagon-square and thus complete our analysis of tilings. Some tests using *Surface Evolver* suggest that deflating one of the separated bubble and expanding the other will reduce perimeter.

6.2 Definitions

Here we rigorously define, among other things, the classes of potential minimizers mentioned above. Thus, at this point it is necessary to introduce a convention that we use throughout the paper. Any double bubble partitions the space into three regions, but in naming and in speaking about these configurations, we will usually only look at two of these regions, R_1 and R_2 and not the third, which we call the *exterior*. So for example, Ξ, which might be otherwise be named the "triple band", is the double band. We should remark that on the torus the exterior need not be the largest of the three regions.

6.2.1 Tori and the Cylinder

Definition 6.1. A *flat two-dimensional torus* can be represented by a planar parallelogram with opposite sides identified, for which the shortest side is a shortest closed geodesic, and no interior angle is less than sixty degrees. In this paper we normalize so that the shortest closed geodesic has unit length. This also implies that the shortest side of the torus has length 1. We call the length of the other side L. If $L = 1$, there are two short directions, and on the hexagonal torus (see below) there are three short directions, one corresponding to the geodesic with homology $(-1, 1)$.

Definition 6.2. The *hexagonal torus* can be represented either by a regular hexagon with opposite sides identified, or by a 60-degree rhombus with opposite sides identified.

Definition 6.3. The *flat infinite cylinder* can be represented as the surface contained between two identified infinite parallel lines in the plane. We take the distance between the two lines to be one.

6.2.2 Topological Definitions

Definition 6.4. A *double bubble* consists of two disjoint regions bounded by piecewise smooth curves, each of which may have multiple components.

Definition 6.5. A closed curve has *homology* (p, q) if it wraps around the torus p times in the short direction and q times in the long direction, where p and q are signed to signify the orientation of the curve.

Definition 6.6. A *tiling* is a double bubble in which every component is contractible, including components of the exterior.

Definition 6.7. A *band* Π is a non-contractible annulus. We often refer to the homology of a band, meaning the homology of one of its boundary components.

Definition 6.8. A *swath* Ψ is a set of contractible components with non-contractible and connected closure $\overline{\Psi}$ and with non-contractible complement.

Definition 6.9. A *chain* is a non-contractible cycle of contractible components. Note that every swath contains at least one chain and that every chain is itself a swath. We often refer to the homology of the closure of the union of a chain, meaning the homology of one of its boundary components.

Definition 6.10. A *hexagon tiling* Υ is a tiling in which every component is a curvilinear hexagon.

Definition 6.11. An *octagon-square tiling* is a four-component tiling consisting of two curvilinear octagons and two curvilinear quadrilaterals. The two quadrilaterals are components of the same region.

6.2.3 Geometric Definitions

Definition 6.12. The *standard double bubble* Θ is comprised of three circular arcs meeting in threes at 120°, such that the curvature of the separating arc is the difference of the curvatures of the outer caps. For prescribed areas, Θ is unique up to congruence [34, Proposition 14.1].

Definition 6.13. The *double band* Ξ is the union of two adjacent bands, each of which is bounded by two short geodesics.

Definition 6.14. A *symmetric chain* is a chain that is symmetric about a geodesic circle.

Definition 6.15. A *standard 2-component chain* Ω is comprised of six circular arcs meeting in threes at four common vertices at 120°, such that the curvature of each separating arc is the difference of the curvatures of the outer arcs (see Figure 6.1). Further, by Lemma 6.2 a minimizing standard 2-component symmetric chain has non-zero homology only in a shortest direction of the torus, and by Lemma 6.3 any such standard 2-component symmetric chain is unique up to isometries of the torus.

Definition 6.16. A *lens* is a pair of congruent circular arcs meeting at two vertices at 120°.

Definition 6.17. The *band lens* Φ is the union of a band and a lens, such that the boundary curves of the resulting configuration meet in threes at two vertices, at 120°. Further, by Proposition 6.6, a minimizing band-lens has non-zero homology only in a shortest direction of the torus and encloses the smallest area in the lens. By Lemma 6.19 any immersed band-lens is unique up to isometries of the torus.

Definition 6.18. The *hexagon tiling H* is a 3-component straight-edge hexagonal tiling on the hexagonal torus.

6.3 Existence, Regularity, and Basic Properties

This section provides existence and regularity (Proposition 6.1), a component bound (Proposition 6.2), and some easy consequences essential to all of our analysis.

Proposition 6.1 (Existence and Regularity Theorem [32, 2.3 and 2.4]). In a smooth Riemannian surface S with compact quotient S/Γ by the isometry group Γ, for any two areas A_1 and A_2 (whose sum is less than area(M)), there exists a least-perimeter enclosure of the two areas. This enclosure consists of smooth constant-curvature curves meeting in threes at 120° angles. All curves separating a specific pair of regions have the same curvature.

Remark 6.1. It follows that minimizing bubbles are connected, and that each component of either of the two regions or the exterior is adjacent to both others. Also, note that a double bubble satisfying regularity is in equilibrium, which means that the "first derivative" of its perimeter under any variation fixing areas is zero.

Proposition 6.2 (Wichiramala Component Bound [38, Proposition 3.1]). Unless all three regions have equal pressure, each region of highest pressure has at most two components.

Proof. The topology of the torus does not interfere with the Wichiramala proof, as all variations used in his argument are strictly local. QEF

Proposition 6.3. The perimeter of a perimeter minimizing double bubble is at most three.

Proof. Three geodesic circles can enclose any pair of prescribed areas A_1, A_2. QEF

Proposition 6.4. No perimeter-minimizing tiling can occur on a torus with unit shortest side length and longest side length greater than $\sqrt{3}$.

Proof. Consider a torus with a tiling on it. Now, at each coordinate in the long direction of the torus, consider the unique non-contractible shortest closed geodesic associated with that coordinate. This shortest closed geodesic must intersect at least one curve at some point, otherwise the configuration could not be a tiling. Further, it must intersect some other or potentially the same curve in some other point, otherwise two components of the same region would be bordering themselves and the configuration could not be a minimizer. Now, considering the sum over all such shortest closed geodesics, we see that the configuration must span the long direction of the torus at least twice. Using the co-area formula, the configuration must have perimeter at least $2 \cdot L \cdot \sin(\theta) \geq \sqrt{3} \cdot L$. But by Proposition 6.3, $P \leq 3$, so $L \leq \sqrt{3}$. QEF

6.4 Topological Classification of Minimizers

In this section we build up several lemmas to prove two important propositions. Proposition 6.5 says that if a perimeter minimizing double bubble is not a tiling, and both regions and the exterior all have equal pressure, then the configuration is the double band. This is important in later sections. Proposition 6.6 is the backbone for the entire proof. It shows that all minimizing double bubbles must either be a tiling, a contractible double bubble, a swath, a band with some contractible blobs, or the double band. The next section and each of its subsections deals with the topological classes developed in this proposition.

Lemma 6.1. In a partition of the torus into piecewise smooth non-contractible components, each component is a band.

Proof. Every closed curve in the boundary must be non-contractible. Hence the boundary must consist of disjoint closed curves, all in the same non-trivial homology class. It follows that each component is a band. QEF

Lemma 6.2. Any non-contractible component, A, in a minimizing double bubble on the torus must either be a band or have contractible complement.

Proof. If the complement of A is contractible, we are done. Suppose that the complement of A is not contractible. If there are any contractible components of the complement inside of A, merge them with A. Now we have a partition of the torus into two non-contractible regions, and by Lemma 6.1, A must be a band or a punctured band (since we filled the holes in A). But since A is part of a minimizing double bubble, by Remark 6.1 A must be a band. QEF

Lemma 6.3. Given a chain or band that is part of a minimizing double bubble, and with homology (p, q), $|p| \leq 1$ and $|q| \leq 1$.

Proof. Since a band and the closure of the union of a chain are simply annuli, each chain or band has

at least two disjoint, non-contractible boundary curves. Thus, if $|p| > 1$ or $|q| > 1$ then the perimeter is at least four. By Proposition 6.3 the double bubble cannot be a minimizer. QEF

Lemma 6.4. A perimeter minimizing double bubble on the torus, together with its exterior, has at most three disjoint bands or chains. Further, if there are three, then they must all be bands, and the double bubble must be the double band.

Proof. Since a band and the closure of the union of a chain are non-contractible annuli, each chain or band has at least two non-contractible boundary curves, each of length at least one. Since each such boundary curve borders at most two such annuli, if there are n annuli, there will be at least n such non-contractible boundary curves. Thus, by Proposition 6.3, $n \leq 3$. Further, if $n = 3$, the perimeter of the double bubble is greater than three unless each of the non-contractible boundary curves is a geodesic in the short direction of the torus and if there is no other boundary, so that the double bubble must be the double band.

QEF

Proposition 6.5. A perimeter minimizing double bubble of three equal pressures is either a tiling or the double band.

Proof. Assume that the double bubble is not a tiling, so that there is at least one non-contractible component. By Lemma 6.2, this component is either a band, or its complement is contractible. The latter is impossible, for it would be a planar polygon with interior angles of 240 degrees.

Thus, the configuration must contain at least one band, bounded by piecewise geodesic curves with interior angles of 120 degrees. Lifting to the plane shows that the boundary curves must be smooth closed geodesics.

Now, denote the rest of the boundary by X. If X is contractible, merely slide X up against the band to contradict regularity. If X is non-contractible, then the double bubble plus its exterior must have at least three bands. By Lemma 6.4, the double bubble must be the double band.

QEF

Proposition 6.6. All minimizing double bubbles on the torus must be either a contractible double bubble, a tiling, a swath, a band adjacent to a set of components whose union is contractible, or the double band (after perhaps relabelling the two regions and the exterior).

Proof. We may assume that every non-contractible component is a band, since otherwise by Lemma 6.2 it is the complement of a contractible double bubble.

If the double bubble plus its exterior has no bands, then all regions are contractible, and thus by definition the double bubble is a tiling.

If the double bubble plus its exterior has one band, then the complement is a set of contractible components. But the closure of the union of this set of components, being the complement of a band, is itself a band, and thus non-contractible. Thus, by definition the double bubble is a swath.

If the double bubble plus its exterior has two bands, consider the complement, C, of the closure of the union of the two bands. If C is non-contractible, then we have a partition of the torus into non-contractible regions, and thus, by Lemma 6.1 this complement is also a band. But since the two bands are the only non-contractible components, we must have a configuration with two bands and a swath. But the swath contains at least one chain, and so the double bubble plus its exterior contains one chain and two bands, which impossible by Proposition 6.4. So the complement of the closure of the union of the two bands must be contractible. And thus the double bubble must be a band adjacent to a set of components whose union is contractible

If the double bubble plus its exterior has three or more bands, then by Proposition 6.4, the double bubble must be the double band.

QEF

6.5 The Four Classes with Contractible Components

By Proposition 6.6, all minimizing double bubbles on the torus must be either a contractible double bubble, a tiling, a swath, a band adjacent to a set of components whose union is contractible, or the double band. While the double band describes a specific geometric configuration, each of the remaining possibilities is merely a topological class of configuration. In this section we identify the minimizer in each of these four remaining topological classes: the standard double bubble among contractible double bubbles (Proposition 6.1), a two-component symmetric chain among swaths (Proposition 6.10), and the band-lens among bubbles with a single band (Proposition 6.11) Proposition 6.12 and 6.13 use the Euler characteristic of $\mathbf{T}^2$ and some basic facts about the regularity of minimizing bubbles to show that only two tilings are potential minimizers: the hexagon tiling H on the hexagonal torus, and (perhaps) an octagon-square tiling.

6.5.1 Contractible Double Bubbles

Proposition 6.1 uses the result from $\mathbf{R}^2$ that perimeter-minimizing double bubbles are standard [24], to show that the only potential minimizer among contractible double bubbles is the standard double bubble.

Lemma 6.1. The perimeter of a standard double bubble is greater than Pi times its diameter: $P > \pi D$.

Proof. We obtain a formula for P/D as a function $f(\theta)$, where θ is the angle between the interior arc of the standard double bubble and the chord that connects its endpoints. Then we show $f(\theta) > \pi$ for the entire domain $0 < \theta \leq \pi/3$.

Let r_1, r_2, c_1, c_2 be the radii and centers of the longer and shorter exterior arcs of the standard double bubble, respectively. Then the diameter is the sum of r_1, r_2, and the distance between c_1 and

c_2:

$$D = r_1 + r_2 + r_1 \cos(\pi/3 - \theta) + r_2 \cos(\pi/3 + \theta).$$

Applying the formula for radii in Section 6.6.1, we obtain a expression that can be solved for C, the length of the separating chord:

$$D = C((1 + \cos(\pi/3 - \theta))/(2\sin(2\pi/3 + \theta)) + (1 + \cos(\pi/3 + \theta))/(2\sin(2\pi/3 - \theta))).$$

Substituting for C in the equation for perimeter in Section 6.6.1, we obtain the desired function in θ:

$$f(\theta) = P/D = (8\pi \sin\theta\cos\theta + 3\sqrt{3}\theta)/(6\sin\theta\cos\theta + 3\sin\theta)$$

.

To show $f(\theta) \geq \pi$ is equivalent to showing

$$8\pi \sin\theta\cos\theta + 3\sqrt{3}\theta \geq \pi(6\sin\theta\cos\theta + 3\sin\theta).$$

Moving the terms to the left hand side and using the identity $\sin(2\theta) = 2\sin\theta\cos\theta$, we obtain the equivalent inequality

$$\pi\sin(2\theta) + 3\sqrt{3}\theta - 3\pi\sin\theta > 0. \qquad (*)$$

We view the left hand side L as a function of θ; its second derivative is

$$-4\pi\sin(2\theta) + 3\pi\sin\theta = \pi\sin\theta(-8\cos\theta + 3).$$

For $\theta \in (0, \pi/3)$, we have $\sin\theta > 0$ while $\cos\theta > \cos(\pi/3) > 3/8$. Hence, this last expression is strictly negative, and the left hand side L of $(*)$ is strictly concave for $\theta \in [0, \pi/3]$. Because L equals 0 when θ is at the endpoints of this interval, it must be positive along the interior of this interval, as desired.

QEF

Proposition 6.1 (Generalization of [28]). The only possible minimizer among contractible double bubbles is the standard double bubble.

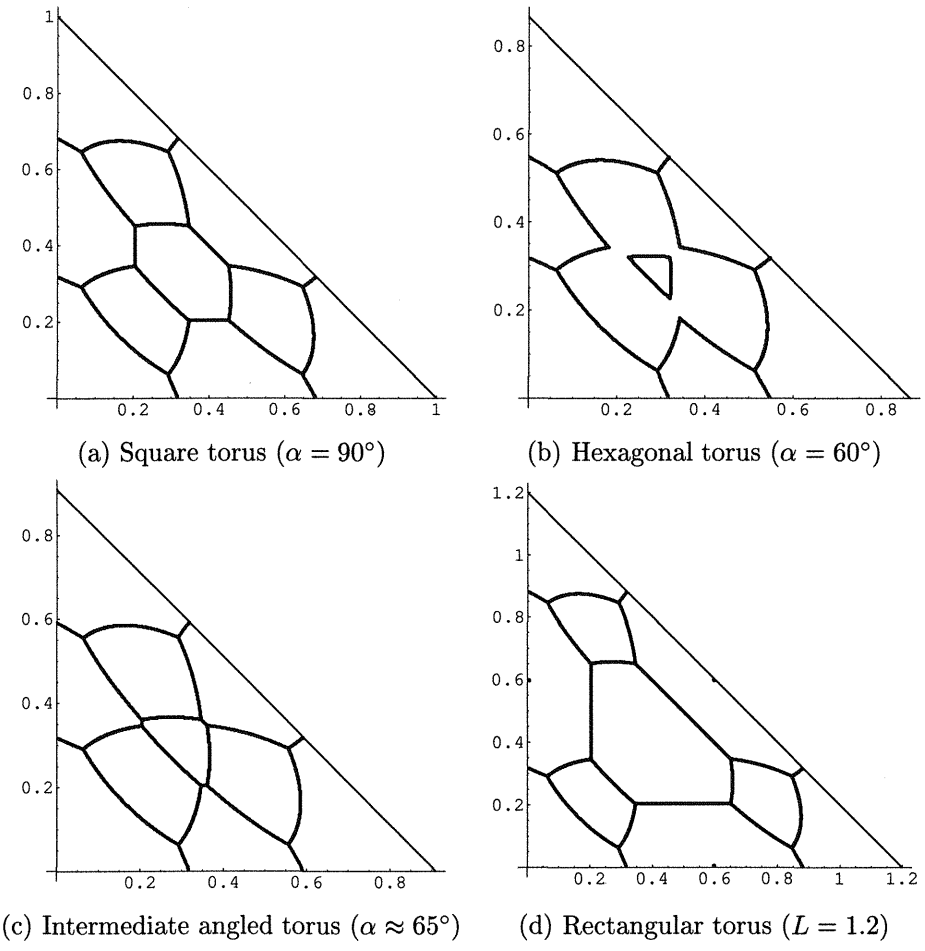
Proof. Assume that there is a perimeter minimizing contractible double bubble Σ enclosing areas A_1 and A_2 that is not the standard double bubble. By [24], the unique perimeter-minimizing solution in $\mathbf{R}^2$ for the same prescribed areas is a standard double bubble Θ. If Θ fits on the torus, it will beat Σ, since it beats Σ in the plane. Therefore Θ does not fit on the torus.

But since Θ does not fit on the torus, then its diameter must be greater than the shortest side length of the torus (otherwise, it would fit with the axis of symmetry in any direction). But, by Lemma 6.1, $P > \pi \cdot D > 3$, and thus by Proposition 6.3, Θ is not a minimizer.

QEF

6.5.2 Swaths

By Proposition 6.5, minimizing swaths must be chains consisting of two or four components. By Proposition 6.8, asymmetric chains cannot be minimizers since they have perimeter greater than three. Finally, by Proposition 6.9, minimizing symmetric chains can have only two components.



(a) Square torus ($\alpha = 90°$)

(b) Hexagonal torus ($\alpha = 60°$)

(c) Intermediate angled torus ($\alpha \approx 65°$)

(d) Rectangular torus ($L = 1.2$)

Square Torus. *Twenty-one smooth-looking curves divide the space into ten regions. The approximate locations of the four vertices in the lower left corner are* $\approx (0.34596, 0.20352)$, $\approx (0.20352, 0.34596)$, $\approx (0.06, 0.29)$, *and* $\approx (0.29, 0.06)$. *The two curves parallel to the axes are* A_1 *and* $A_2 = (\frac{8\pi}{3} - 2\sqrt{3})^{-1} \approx 0.2035$, *respectively. The curves that touch the axes in the lower left corner touch at* $\frac{1}{\pi} \approx 0.3183$.

Hexagonal Torus. *Eighteen smooth-looking curves divide the space into eight regions.*

Figure 6.2: How minimizing double bubble types depend on the prescribed areas for four flat tori

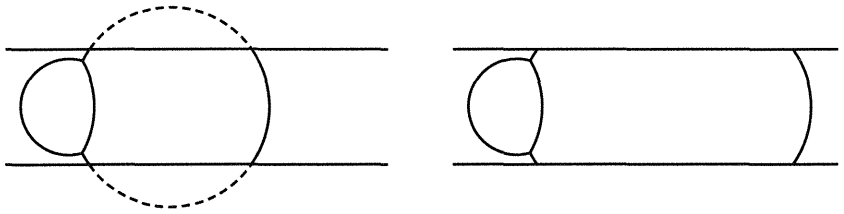


Figure 6.3: On the cylinder, we can make simple adjustments when the standard double bubble does not fit; the same argument does not work on the torus.

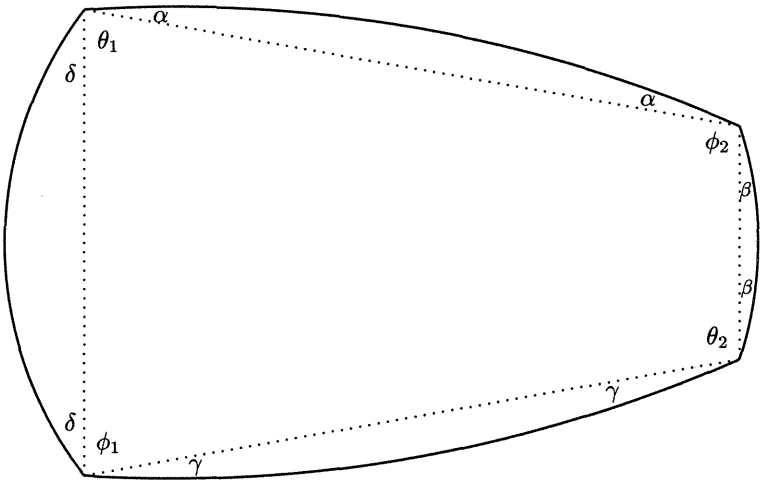


Figure 6.4: Important angles in a regularity satisfying quadrilateral

Lemma 6.2. If a component of a region is a curvilinear quadrilateral, then the quadrilateral is cyclic (i.e., the vertices lie on a common circle).

Proof. We must show that opposite angles sum to 180°. Consider the chords joining consecutive vertices as in Figure 6.4.

Since each boundary curve must meet its corresponding chord at equal angles $\alpha, \beta, \gamma, \delta$ and these curves meet each other at 120° angles, we have the following set of equations:

$$2(\alpha + \beta + \gamma + \delta) + \theta_1 + \theta_2 + \varphi_1 + \varphi_2 = 480°;$$

$$\theta_1 + \theta_2 + \varphi_1 + \varphi_2 = 360°.$$

From this we easily get $\alpha + \beta + \gamma + \delta = 60°$. Summing opposite angles of the component yields

$$\alpha + \beta + \gamma + \delta + \theta_1 + \theta_2 = 240°;$$

$$\alpha + \beta + \gamma + \delta + \varphi_1 + \varphi_2 = 240°.$$

Whence, we get $\theta_1 + \theta_2 = \varphi_1 + \varphi_2 = 180°$ QEF

Remark 6.2. The center of the circle is of course where the two bisectors of any two nonparallel sides intersect. The following Lemma shows that Figure 6.4 describes the general case.

Lemma 6.3. The vertices of a curvilinear quadrilateral component of a regularity satisfying planar double bubble form an isosceles trapezoid. Furthermore, if the trapezoid is not a rectangle, then the two boundary components corresponding to the parallel sides of the trapezoid lie on the same circle.

Proof. We reflect across the line of symmetry of each arc to see that the component must be an isosceles trapezoid.

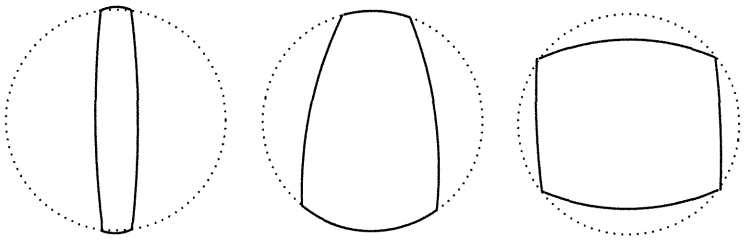


Figure 6.5: Quadrilateral components

Suppose it is not a rectangle. Since the arcs corresponding to the two non-parallel sides make a 120° angle with the longer of the two parallel sides, they will also make a 120° interior angle with the completion of this arc to a circle. Since the component is not rectangular, any arc other than these two with the correct curvature which makes a 120° angle with one of the non-parallel arcs will not make a 120° angle with the other. We therefore have the desired result.

QEF

Remark 6.3. In the previous two lemmas it is actually enough to have two parallel sides be circular arcs of the same curvature, the other two boundary curves can then be arbitrary smooth curves. This is relevant for example in Figure 2 of [41].

Proposition 6.2. In a regularity satisfying double bubble, then arcs λ_1 and λ_2 connecting to one of the two parallel sides of a quadrilateral component that is not a rectangle lie on the same circle.

Proof. Consider the digon obtained by replacing μ_2 with the completion of the arc containing σ_1 and σ_2. This gives us two boundary components of some standard double bubble. Since pressures add, we can complete this double bubble with the correct circular piece. Since, locally, the original quadrilateral component looks like the standard double bubble, the boundary curves connecting to it lie on the same circle. QEF

Proposition 6.3. Unless both regions and the exterior all have equal pressure, every component

of each lowest pressure region of a regularity satisfying double bubble that is not a tiling is non-contractible.

Proof. Suppose a component of the lowest pressure region is contractible. Then it is must be a curvilinear n-gon, where n is strictly greater than six, since each edge of a lowest pressure region must either be straight or bow into the region, and not all the edges are straight, leaving room for 120°angles only if there are more than six edges. By regularity, consecutive sides must border different regions, so n must be even.

By Proposition 6.2, a region of highest pressure, which we will call R_1, has at most two components. Since every component of a highest pressure region can have at most four vertices, R_1 can have at most eight vertices.

Therefore, there is exactly one lowest pressure region, and it has exactly eight edges, otherwise R_1 would have too many vertices.

Furthermore, we note that none of the components of R_1 can be digons, since this would leave six vertices for the other component of R_1, which as we have noted is too many. Thus, there are two quadrilateral components of R_1 each bounded by a pair of sides on the octagon.

By the results of the previous subsection, the double bubble can not be contractible. Therefore at least one of the quadrilateral components must wrap around the torus. If only one of the quadrilateral components wraps around the torus, we get the absurd configuration in Figure 6 (a), in which two pieces of the quadrilateral that does not wrap are disconnected from one another. If both of the components wrap around the torus, either they wrap around different sides, in which case the configuration is a tiling (the octagon square actually – but at any rate, not addressed in this proposition) or we have a configuration like that of Figure 6 (b).

As illustrated, each of the quadrilaterals must admit the inscription of a rectangle on their vertices, otherwise by Lemma 6.3 the sides which do not border the low pressure octagon would lie on the same circle, and by Proposition 6.2 the two adjacent edges of the low pressure octagon would

A. B.

Figure 6.6: Two absurd low pressure components

have to lie on a circle as well, which is ridiculous.

It now follows from a standard fact about parallel lines and transversals, that the edge of the low pressure octagon that joins the two edges shared with the quadrialteral must meet one of these edges at an acute angle. But this contradicts regularity, proving the proposition.

QEF

Proposition 6.4. In a perimeter minimizing swath containing a chain K, no component of K is a digon.

Proof. Suppose K contains a digon D. Since D is adjacent to only two components, if it occurs in K, it must be adjacent to components of the same region. This requires either that four arcs meet at a point, or that some region is separated from itself by arcs which could be removed to reduce perimeter. In either case, the original configuraion can not have been minimizing, proving the claim. QEF

Proposition 6.5. Minimizing swaths must be chains with a total of two or four components from two regions.

Proof. By Proposition 6.5 the three regions in a minimizing swath Ψ do not have the same pressure. Furthermore, by Proposition 6.3 a least pressure region does not appear in the swath. Also, the least pressure region is unique since a swath contains components from two regions. Consider a chain K in Ψ. The components in K must alternate between two higher pressure regions R_1 and R_2. By Proposition 6.2 the region of highest pressure, R_1, has at most two components, so K has exactly 2 or 4 total components.

Suppose K has four components. Then any additional components of the R_2 in the swath have to be adjacent to components of R_1 in the chain (since by definition, the union of swath components is connected). However, this is impossible since components of R_1 already have four vertices and a convex non-polygonal component has at most four vertices. Thus, a swath containing a four component chain must be a four-component chain.

Next, suppose K has two components. Further, suppose the swath contains additional components. Then there is some component adjacent to the chain (otherwise the closure of the swath is disconnected). As before, there can be no extra components adjacent to the component of R_1 in K. So, there is a component of R_1 adjacent to the component of R_2 in K. If this is the only other component in the swath, then it is a digon which can be eliminated by sliding it along ∂R_2

Figure 6.7: Extra digons do not attach to the symmetric chain

until it touches another component or a vertex creating an illegal singularity (see Figure 6.7). If, instead, there is another component of R_2 attached to the second component of R_1, then it is either a digon or it is attached to another component of R_1. In the case of the former, we can eliminate this configuration using the same sliding technique. The latter case is impossible since, by Proposition 6.2 it would have to be attached to an existing quadrilateral component of R_1. This contradicts our assumption that the swath contains additional components; therefore, a minimizing swath is a chain with two or four components. QEF

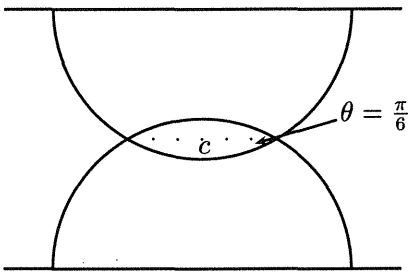


Figure 6.8: Chain violates regularity.

Proposition 6.6. In a symmetric 2-component chain where the arcs separating each component from the exterior lie on the same circle, the total perimeter of the chain is strictly greater than 3.

Proof. In a cylinder with geodesic circle of length 1, we can parameterize a symmetric chain composed of two components, by two parameters: the length C of a chord connecting vertices shared by both components of the chain, and the angle θ between this chord and the boundary curve con-

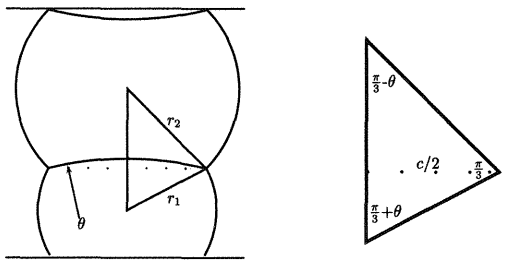


Figure 6.9: Chain constraints.

necting the same vertices. Here θ must range between 0 and $\frac{\pi}{6}$ radians, since at $\theta = \frac{\pi}{6}$, the chain violates regularity (see Figure 6.8).

At an endpoint of C we draw the radii r_1, r_2 of the two exterior arcs of the chain that meet at this point (See Figure 6.9). Connecting the opposite endpoints of these radii (the centers of the circular arcs), we have a set of triangles from which we can compute a value of C for a given θ.

Using the law of sines in Figure 6.9, we discover that $\frac{\sin\frac{\pi}{3}}{\frac{1}{2}} = \frac{\sin(\frac{\pi}{3}-\theta)}{r_1}$. Since, $\sin(\frac{\pi}{3}+\theta) = \frac{\frac{C}{2}}{r_1}$, we get

$$\frac{\sin(\frac{\pi}{3}-\theta)}{\sqrt{3}}\sin(\frac{\pi}{3}+\theta) = \frac{C}{2}$$

Further simplification yields:

$$C = \frac{\sqrt{3}}{2} - \frac{2}{\sqrt{3}}\sin^2\theta$$

We can obtain a lower bound on the perimeter of the chain by approximating the length of each arc by its corresponding chord. This length is exactly $2C + 2$. Since C is decreasing in θ and $C(\frac{\pi}{6}) = \frac{1}{\sqrt{3}} \approx .577$, we conclude that perimeter is greater than 3. By scaling, we get the result for arbitrary cylinders. QEF

Proposition 6.7. For any symmetric chain Σ consisting of four components, there exists a two-component chain with equal perimeter that is similar to an adjacent pair of components in Σ.

Proof. Since alternating components of a 4 component symmetric chain are congruent (this comes from symmetry about a line perpendicular to the geodesic circle), we construct a two-component chain by taking adjacent components of Σ and scaling up by a factor of 2. The perimeter remains constant since we doubled half of the original perimeter of Σ. Also, by construction, the chain is similar to Σ. QEF

Proposition 6.8. All asymmetric chains have perimeter greater than 3.

Proof. By Lemma 6.3, at least one component of an asymmetric chain must have vertices that form a non-rectangular isosceles trapezoid, and the arcs separating this component from the exterior lie on the same circle. By Proposition 6.2, the same is true for the other components in the chain.

We can expand the cylinder continuously, rotating components as needed, until the chain fits symmetrically along a geodesic circle (see Figure 6.10) . By Propositions 6.6 and 6.7, the perimeter of the chain is greater than three times the side length of the expanded cylinder. Hence the perimeter of the chain is greater than 3, and by Proposition 6.3, we have the desired result. QEF

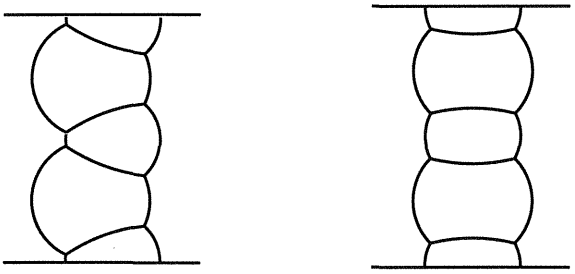


Figure 6.10: Rotate bubbles and stretch cylinder to make asymmetric chain symmetric.

Proposition 6.9. A four-component symmetric chain cannot be a minimizer.

Proof. Suppose that a four component chain Σ is a minimizer for some pair of prescribed volumes. Let R_1 be the region of highest pressure. Since each pair of vertices shared by adjacent components is separated by the same horizontal distance, we can rearrange the components so that those belonging to the same region are adjacent (see Figure 6.11). Upon careful relabelling of the regions, area is maintained and we have violated regularity, a contradiction. QEF

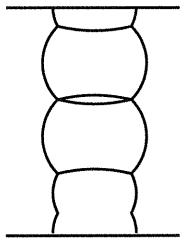


Figure 6.11: A regularity violating rearrangement of the components of a four component symmetric chain.

Proposition 6.10. The only potential minimizer among `sws` is a standard two-component symmetric chain.

Proof. By Proposition 6.5, minimizing swaths must be chains consisting of two or four components. By Proposition 6.8, asymmetric chains cannot be minimizers since they have perimeter greater than 3. Finally, by Proposition 6.9, minimizing symmetric chains can have only two components. QEF

Conjecture 6.2. A minimizing standard two-component symmetric has non-zero homology only in the short direction of the torus.

(This can be checked computationally.)

6.5.3 A Single Band Adjacent to a Set of Components with Contractible Union

Proposition 6.11 uses regularity and the results of Proposition 6.2, and Proposition 6.3 to show that the only minimizer in the class of double bubbles with a single band adjacent to a set of components with contractible union is the band-lens, Φ.

We will consider a proposed minimizer Σ with regions R_1, R_2 and the exterior, such that R_1 contains a band, Π.

In several of the proofs, we do some area maintaining deformation which reduces perimeter, or which causes the bubbles to bump up against one other, violating regularity. In either case, the original configuration was not a minimizer.

Lemma 6.4. The pressure difference between R_1 and the exterior must be zero, i.e. all curves separating R_1 from the exterior must be straight lines.

Proof. Consider a curve separating the band, Π, from the exterior. If it is a straight line segment, the proof is complete, since the all curves separating R_1 from the exterior have the same curvature.

If the curve to be considered has nonzero curvature, continuously straighten the curve to reduce perimeter, and slide one of the sides of $\partial\Pi$ in order to maintain area. Either we will be able to straighten all such pieces while reducing perimeter, or we will first cause a violation of regularity.

Note that this implies that all curves enclosing components of R_2 must have the same curvature.

QEF

Lemma 6.5. In the class of double bubbles with a single band, R_2 must have a higher pressure than both R_1 and the exterior.

Proof. Assume not. Then components of R_2 are either bounded by straight lines, or are concave with respect to both R_1 and the exterior.

Now because the blob is contractible, we can create a closed geodesic T which is not in Π and which is not adjacent to any region but the exterior. If we slide T towards one of the blobs, they

must eventually collide. Since all components are either concave or bounded by straight lines, the collision either occurs at a point which must be an illegal singularity (i.e. it cannot be 3 smooth curves meeting a 120 degree angles), or in a line. If it is in a line, the line with which T collided must be a complete closed geodesic, otherwise it would contain an illegal singularity. This configuration either is the meridional band solution Ξ, or Ξ beats it. QEF

Corollary 6.1. R_2 has at most two components.

Proof. This follows directly from the previous lemma and Proposition 6.2. QEF

Corollary 6.2. All contractible components of a double bubble with a single band consist of components of R_2.

Proof. By Lemma 6.5 and Lemma 6.4, R_1 and the exterior have equal, lowest, pressure. Thus, by Proposition 6.3, their components must be non-contractible. Hence only components of R_2 can be contractible. QEF

Proposition 6.11. The only potential minimizer among double bubbles with a single band is the band-lens Φ.

Proof. By Proposition 6.2, R_2 (the region without the band) has either one or two components.

If R_2 has a single component, by Corollary 6.2, this is the only contractible component in the double bubble. The side of $\partial\Pi$ which does not intersect ∂R_2 is, by Lemma 6.4, a straight line; and, by the definition of a band, closed. Now consider the side which intersects ∂R_2. Again by Lemma 6.4, the curve which separates Π from the exterior is a straight line. Now, rotate the component until this line lies on a geodesic circle, while sliding the boundary of Π to maintain area. This will always save perimeter or cause a violation of regularity. Now the only way in which curvature and regularity conditions can be satisfied is if we have the standard band-lens solution.

If R_2 has two components, by Corollary 6.2 these are the only contractible components that share boundary with Π. Further, either both components are on the same side of $\partial\Pi$ or they are on

opposite sides of $\partial\Pi$.

Assume they are on opposite sides of $\partial\Pi$ and consider one of these sides. By Lemma 6.4, any curve which separates Π from the exterior is a straight line. Now, rotate the component until this curve lies on a geodesic circle, while sliding the boundary of Π to maintain area. This will always save perimeter or cause a violation of regularity. Repeat for the other side. Move one of the components so that they both lie on the same side. Now we can slide them, while maintaining area, until there is a violation of regularity, showing the configuration to be bad. If the components are too large to fit on the same side, the configuration must have perimeter greater than 3, which by Lemma 6.3 is a contradiction.

If both components of R_2 are on the same side of $\partial\Pi$, we must consider their structure in a bit more detail. By Lemma 6.4, any curve which separates Π from the exterior is a straight line. Now, each component must meet the line of $\partial\Pi$ in two frames where three curves meet at 120 degree angles. Now rotate and shift the two components until all 4 frames and thus the lines connecting them lie along a shortest closed geodesic. We can maintain area by sliding the boundary of Π, and we will either have a violation of regularity, or save perimeter by reducing the original configuration to a configuration in which connecting lines are part of a shortest closed geodesic. If we have the latter, slide the two components, while maintaining area, until there is a violation of regularity showing the configuration to be bad.

Thus, the double bubble must be the band-lens, Φ. QEF

Lemma 6.6. In a minimizing band-lens, the band must have non-zero homology only in a short direction of the torus and the lens must enclose the smallest of the three areas.

Proof. First we show that any pair of areas that can be enclosed by a band-lens along the long direction can be enclosed by a band-lens in the shorter direction. Assume not. Then either the lens is too long or too wide to fit in the short direction. If it is too long to fit in the shorter direction, then the double bubble will have perimeter greater than three, and, by Proposition 6.3, it cannot be a

minimizer. If the lens is too wide, then the geodesic not adjacent to the lens cannot get close enough to make the band thin enough. But if this happens in the short direction, then it will certainly happen in the long direction where there is less leeway.

It is easy to see that the lens must enclose the smallest of the three possible areas. This is because the other two regions can exchange area for free, but the perimeter of the band-lens is strictly increasing with the area of the lens. QEF

6.5.4 Tilings

Propositions 6.12 and 6.13 show that the only minimizing tilings are the hexagon tiling H and possibly an octagon-square tiling. The proofs use pressure conditions, the Wichiramala component bound, and a variational argument, as well as elementary graph theory and plane geometry.

Lemma 6.7. A component C of a perimeter minimizing tiling Σ must be a curvilinear polygon with an even number of edges greater than or equal to four.

Proof. Of course, by regularity we know that if x and y are adjacent edges of C, then C is adjacent to different regions along each side. This implies that the regions adjacent to C must alternate with the edges of C, which is only possible if C has an even number of edges.

To complete the proof, we must show that a digon cannot occur in a tiling solution. To do this, note that by regularity a component that is a digon must be a lens. This lens either lies on a pair of straight lines, or on a pair of curves which lie on the same circle.

Either way, we can slide the lens along the lines or arcs without changing perimeter or area. Then either we reach a violation of regularity, or we can slide the lens around the torus without bumping into anything. If the latter is the case, either the lens lies on a free circle, which can be translated to cause a violation of regularity, or the lens lies on a free, closed geodesic. The latter implies that Σ contains a non-contractible component (actually, two), which contradicts our hypothesis. QEF

Lemma 6.8. If a tiling of a torus has vertices of degree 3, the average number of edges per component is exactly 6.

Proof. The Euler characteristic on the torus is $v - e + c = 0$, where v is the number of vertices, e is the number of edges, and c is the number of components. Now, let n be the average number of edges per component. Then $c = \frac{2}{n}e$, since each edge is adjacent to two components, and each component averages n edges. We also know that $v = \frac{2}{3}e$, since each edge is adjacent to two vertices, each vertex is adjacent to three edges. Substituting into the Euler formula, we obtain $(\frac{2}{3} - 1 + \frac{2}{n})e = 0$, from which we obtain $n = 6$. QEF

Proposition 6.12. A perimeter-minimizing tiling Σ is either a straight-line hexagon tiling, or an octagon square tiling (a tiling in which the region of highest pressure consists of two curvilinear quadrilaterals, and the other two regions are curvilinear octagons).

Proof. If the three regions of Σ have equal pressure, then all boundary curves are straight lines. Now, the only straight-edged polygon with 120° interior angles is the hexagon, so by regularity all components must be hexagons.

If the three regions of Σ do not have equal pressure, then there is a region R_1 with pressure at least as low as all other regions, and a region R_2 with pressure at least as high as all other regions. For the following vertex-counting argument, note that by regularity each vertex of Σ must be a boundary constituent in components of all three regions.

Now, every boundary curve of an R_1 component is either a straight line or bows in, so by regularity every R_1 component must have at least eight sides. Then R_1 components must each contain at least eight vertices in their boundaries, so Σ must contain at least eight vertices.

Similarly, every boundary curve of an R_2 component is either a straight line or bows out, so by regularity and by Lemma 6.7, every R_2 component must have exactly 4 sides and 4 vertices. Furthermore, we know by the Wichiramala Component Bound that R_2 has at most 2 components.

Then R_2 components are adjacent to at most $(2\ components)(4\frac{vertices}{component}) = 8\ vertices$, so Σ can contain at most 8 vertices.

Thus Σ must contain exactly eight vertices, and R_2 must consist of exactly two four-sided components. Now the components of the remaining region, R_3, contain a total of eight vertices in their all boundaries. are adjacent to eight vertices, so R_3 must have either one 8-sided component or two four-sided components. But by Lemma 6.8, the average number of sides per component must be exactly six. The only combination that satisfies this condition is for R_3 to have one eight-sided component. This implies by regularity that R_3 must have less pressure than R_2, since a region of highest pressure cannot have eight-sided components.

Thus if the three regions of Σ do not have equal pressure, there is a unique region of highest pressure consisting of two curvilinear quadrilaterals, while the other two regions each consist of a single curvilinear octagon. QEF

Lemma 6.9. The number of components in a minimizing hexagon tiling Υ must be a multiple of three, since each region must have the same number of components.

Proof. By regularity, every vertex in Υ must be adjacent to exactly 1 component from each of the three different regions. Then every region has exactly the same number of vertices lying in the boundaries of its components. By definition, every component in a hexagon tiling has the same number of vertices (6) in its boundary. Then every region must have the same number of components. Since there are three regions, the total number of components must be a multiple of three. QEF

Lemma 6.10. A hexagon tiling with 9 or more components is not a perimeter minimizer.

Proof. We first fix some notation. Given a tiling τ_0 with $t = 3n$ hexagons on the 2-torus, the claim is that the total perimeter of the tiling is greater than three.

Extend the tiling τ_0 of the torus to a tiling τ of the plane, coloring the plane in t colors in the

natural manner. Without loss of generality, we may assume that any segment in the perimeter of the tiling forms an angle of 0, $\pi/3$, or $2\pi/3$ with a horizontal line in the plane.

We can travel from any fixed hexagon H in the tiling to a neighboring hexagon in any of six *directions*, which we label `N`, `NE`, `SE`, `S`, `SW`, `NW` in clockwise order, where by travelling in the `N` direction, we travel to the hexagon which borders the top edge of H. Similarly, we can travel from any vertex in the tiling to a neighboring vertex in several possible directions, in three directions from any fixed vertex but in six directions total. We label these directions `e`, `se`, `sw`, `w`, `nw`, `ne`, where any rays in these directions make clockwise angles of 0, $\pi/3$, ..., $5\pi/3$ with the horizontal ray pointing to the right.

The first major step is to finding a 'gridlike' fundamental domain

Suppose that the two hexagons A_1 and A_2 have the same color, and that we can travel from A_1 to A_2 via some sequence $\mathcal{S}$ of directions $(d_1, \ldots, d_r)$. (A_1 and A_2 must be associated with different fundamental domains, since by choice each component of τ_0 has a distinct color.)

Let B_1 and B_2 be any two hexagons separated by the same sequence of directions. *A priori*, B_1 and B_2 need not have the same color – but the folowing inductive argument shows that, in fact, they do:

We travel from A_1 to B_1 along some sequence of directions $(e_1, \ldots, e_s)$; the same sequence of directions will take us from A_2 to B_2. Now, starting from A_1 and A_2, and travelling along each direction in $(e_1, \ldots, e_s)$ simultaneously from each of the two hexagons, we pass through regions of the same color at each point (either because all the fundamental domains are colored the same, or because they have been identified in a routine way); hence, B_1 and B_2 are of the same color.

We claim that there exists some $a > 0$ such that any two hexagons separated by a sequence of a 'S's (or equivalently a '`N`'s) are of the same color. Pick any hexagon in τ, and travel in the `S` direction. Among the first $t + 1$ hexagons reached during this journey, some two must have the same color – say, separated by a sequence of a_0 `S`s. From our work above, it follows that *any* two hexagons

separated by such a sequence are of the same color. So, we may set $a = a_0$. We can go even further, though, and choose $a > 0$ above to be minimal.

We now construct our desired fundamental domain as follows. Find a as above. Pick any hexagon $H_{1,1}$ in the tiling of the plane, and take the a hexagons $H_{1,1}, \ldots, H_{a,1}$ obtained by starting from $H_{1,1}$ and traveling in the S direction $a-1$ times. By the minimal definition of a, these have distinct colors. If there are already t hexagons here, we have our fundamental domain.

Otherwise, travel from $H_{1,1}$ in the SE direction to reach another hexagon $H_{1,2}$, and travel in the S direction $a-1$ times to again yield a hexagons $H_{1,2}, \ldots, H_{a,2}$ of distinct colors (by the minimal definition of a). (Note that by traveling from any $H_{r,1}$ in the SE direction, one can reach $H_{r,2}$.) Suppose, for sake of contradiction, that one of $H_{1,2}, \ldots, H_{a,2}$ were the same color as one of $H_{1,1}, \ldots, H_{a,1}$. We would then have that *each* of $H_{1,2}, \ldots, H_{a,2}$ is the same color as one of $H_{1,1}, \ldots, H_{a,1}$; and furthermore, that *all* hexagons in the tiling τ is the same color as one of $H_{1,1}, \ldots, H_{a,1}$. This contradicts our construction of τ from τ_0. We now have found an $a \times 2$ cluster of hexagons, each of a different color. If there are already t hexagons here, we have our fundamental domain.

Otherwise, travel from $H_{1,2}$ in the SE direction to reach another hexagon $H_{1,3}$, and define $H_{1,3}, \ldots, H_{n,3}$ in the natural way. These new a hexagons have distinct colors; and as before, these colors must be distinct from those of $H_{1,1}, \ldots, H_{a,1}, H_{1,2}, \ldots, H_{a,2}$. We now have found an $a \times 3$ cluster of hexagons, each of a different color. If there already t hexagons, we have our fundamental domain.

Otherwise, we can continue on, although because t is finite, eventually we're bound to find our fundamental domain. Also, note that the N edges of the N-most hexagons of our fundamental domain our identified with the S edges of the S-most hexagons; similarly with the SE-most and NW-most edges. Indeed, the SE-most and NW-most edges are identified "naturally" with each other.

The next major step is to finding a short noncontractible path along the perimeter of the tiling.

It is sufficient to show that there exists a sequence $\mathcal{S} = (d_1, d_2, \ldots, d_k)$ of directions (along edges)

in $\{\texttt{e}, \texttt{se}, \texttt{sw}, \texttt{w}, \texttt{nw}, \texttt{ne}\}$ with the following properties:

(i) at most $n = \frac{1}{3}t$ of the d_i equal `e` or `w`, at most t of the d_i equal `se` or `nw`, and at most n of the d_i equal `sw` or `ne`;

(ii) from any hexagon's northwest vertex, travelling according to $\mathcal{S}$ gives a closed, noncontractible path around the torus.

(Note that if we have two vertices in the tiling of the plane which correspond to the same points on the torus, then any continuous path between corresponds to a contractible closed path on the torus. With a little work, this fact can be used to ensure that the paths we find below are noncontractible.)

We select a $(3a) \times b$ "grid" of hexagons in the tiling which serve as a fundamental domain, where all four sides of our fundamental domain contain `nw-se` edges and the tom and bottom sides are identified naturally. Note that it suffices to prove the claim starting from the northwest corner v_0 of the top edge. Also note that $n = ab$.

If $b \geq 3$, then $(d_1, d_2, d_3, d_4, \ldots, d_{6a}) = (\texttt{sw}, \texttt{se}, \texttt{sw}, \texttt{se}, \ldots, \texttt{sw}, \texttt{se})$ suffices.

If $(a, b) = (1, 2)$, then $(d_1, d_2, d_3, d_4) = (\texttt{e}, \texttt{se}, \texttt{e}, \texttt{ne})$.

Because $t \geq 6$, we cannot have $(a, b) = (1, 1)$. Thus, we may assume that $a \geq 2$ and $b \leq 2$. We begin our journey by travelling from v_0 along $2b - 1$ edges, alternating between travelling `e` and `se` so that we travel `e` b times and `se` $b - 1$ times. This brings us to a vertex v_1 on the left side of our fundamental domain.

If v_1 is the west vertex of one of the middle a hexagons of our grid, then our initial $2b - 1$ movements `e`, `se`, ..., `e` are the "same" as between $2a + 1$ and $4a - 1$ movements `sw`, `se`, ..., `se`. From v_1, travel `se`; then repeat our initial $2b - 1$ movements. Our journey is then the "same" as moving between $2(2a + 1) + 1 = 4a + 3$ and $2(4a - 1) + 1 = 8a - 1$ edges `sw`, `se`, ..., `se`. In other words, we are now at some vertex v_2 which is the west vertex of either (a) one of the top a hexagons of the first column; or (b) one of the bottom a hexagons of the first column.

In either case, we have travelled e at most $2b$ times and se at most $2b-1$ times. And we are now either (a) on the west vertex of one of the top a hexagons of the first column, or (b) on the west vertex of one of the bottom a hexagons of the first column.

In case (a), we may alternate between travelling ne and nw to reach v_0, travelling ne at most a times and nw at most $a-1$ times. Altogether, then, we have travelled: e at most $2b$ times, so total is $2b \leq ab$; se at most $2b-1$ times and nw at most $a-1$ times, so total is $a+2b-2 = ab-(a-2)(b-1) \leq ab$; ne at most a times so total is $a \leq ab$.

Hence, condition (ii) is satisfied.

In case (b), we may alternate between travelling se and sw to reach v_0, travelling se at most a times and sw at most $a-1$ times. Altogether, then, we have travelled: e $\leq 2b$ times, so total is $2b \leq ab$; se $\leq a+2b-1$ times, so total is $a+2b-1 = ab-(a-2)(b-1)+1 \leq ab+1$; sw $\leq a-1$ times, so total is $a-1 \leq ab$.

Hence, condition (ii) is satisfied if $a > 2$, because in this case the second inequality above is strict.

If $a = 2$, then it seems that we might have an incredible stroke of bad luck so that all our approximations above are sharp. Actually, though, this doesn't happen: one can verify that with a 6×1 or 6×2 grid, we can still find a path of the desired form. (To do so in the 6×2, we need to use the barbell property.)

We now make the final points to show that τ_0 has perimeter greater than 3.

For each of our t hexagons, start at the northwest vertex of the hexagon and travel along $(d_1, d_2, \ldots, d_k)$ as in the lemma of the previous section. Each of these t overlapping paths uses at most $n = t/3$ edges parallel to the e (resp. se or sw) direction, implying that they use a total of $t^2/3$ such edges. Furthermore, a symmetry argument implies that all of the t edges parallel to the e (resp. se or sw) direction are used the same number of times in these t paths. Therefore, each edge is used at most at most $(t^2/3)/t = t/3$ times in these overlapping paths.

Thus, the total perimeter P' of the t overlapping paths is at most $t/3$ times the perimeter P of τ_0.

Furthermore, each of the t overlapping paths has perimeter greater than 1, implying that altogether their total perimeter P' is greater than t.

It follows that

$$t < P' \leq (t/3)P,$$

or $P > 3$. QEF

Lemma 6.11. A hexagon tiling Υ with 6 components has perimeter strictly greater than 3.

Definition 6.1. Let a *barbell* be a pair of hexagons related as in Figure 6.12. Then the two hexagons each adjacent to both hexagons of the pair must be of different regions (B and C in Figure 6.12), and the two hexagons of the barbell must be of the same region (A in Figure 6.12).

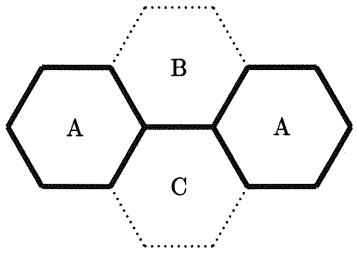


Figure 6.12: Regularity satisfying hexagon tilings exhibit the barbell-lift property.

Definition 6.2. Let a *web* W be a set of hexagons $\{x_1, x_2, ..., x_n\}$ in the lift of a tiling such that for any hexagon $x_i \in W$, another hexagon y is also in W if x_i and y are contained in the same barbell.

Then every hexagon in a web must be of the same region. Furthermore, a web contains exactly one third of the hexagons in the tiling, so all components of a region are contained in a single web.

Proof. By the perimeter bound, a double bubble whose set of boundary curves contains three non-contractible non-geodesic circles as disjoint subsets is not perimeter-minimizing. We show that Υ must contain three such circles by considering the lift of Υ to the plane.

Let the three regions of the tiling be labelled A, B, and C. Note that in the lift, there are exactly two unique hexagonal components of each region; let these components be labelled A_1, A_2, B_1, B_2, C_1, and C_2.

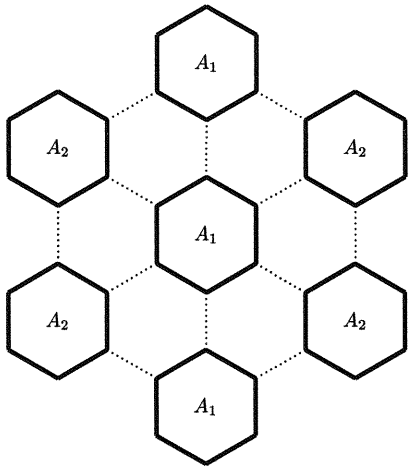


Figure 6.13: Structure of the lift of a six-hexagon tiling.

Now consider a seven-hexagon segment of a web as in Figure 6.13. Let the center hexagon be A_1. Then the other six hexagons must all be of region A, by the definition of a web. Further, at least one of the other six hexagons must also be A_1, since if all of them were A_2 then we would have the relation that every hexagon directly above or below A_2 in the web was also A_2, which would imply that the center hexagon was A_2, a contradiction. Let the hexagon immediately below the center hexagon be A_1. This immediately implies that the hexagon above the center hexagon is also A_1, again by identification. Now if any of the remaining four hexagons is A_1, then all of the hexagons in the segment, and therefore all of the hexagons in the web, must also be A_1. Thus, the remaining four hexagons must all be A_2. Now, considering the web as a whole we see that it consists of alternating vertical stripes of A_1 and A_2 hexagons. Now consider the hexagon adjacent and to the left of the center hexagon in the original web segment. Let this hexagon be the unique hexagon B_1. Then every A_1 hexagon has a B_1 hexagon adjacent and to its left, and we see that in the B web

we also have alternating vertical stripes of B_1 and B_2 hexagons. The same pattern holds for the C web. Thus, there is a single fundamental pattern in the arrangement of the six unique hexagons, up to rotation and relabelling. See Figure 6.13.

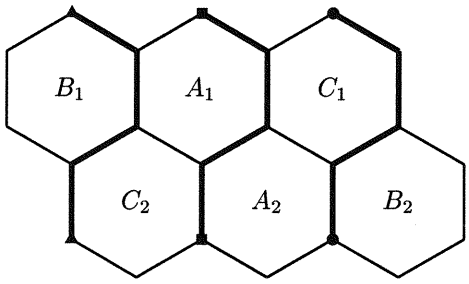


Figure 6.14: A fundamental domain for a six hexagon tiling has perimeter greater than three, since we can exhibit three non-trivial closed curves.

Now the set of six hexagons in Figure 6.14, with boundary curves identified as shown, has precisely the same structure and contains precisely the same perimeter as a six-hexagon tiling of a parallelogram torus. The three paths highlighted in Figure 6.14 are three disjoint non-contractible non-geodesic circles, since for each path the endpoints are in fact the same (identified) point. QEF

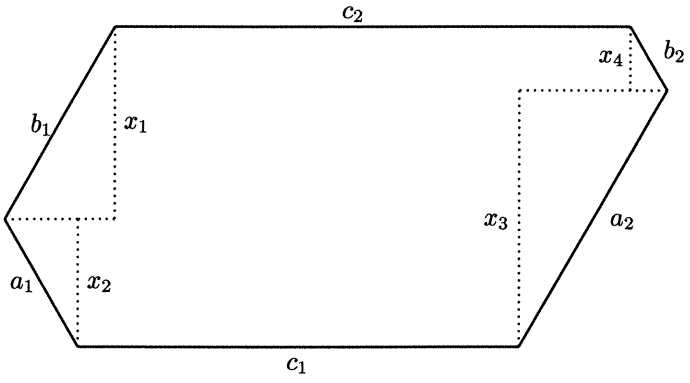


Figure 6.15: In a hexagon with regular angles, a simple geometric construction shows that opposite adjacent side pairs have perimeters which sum to the same total.

Lemma 6.12. In a hexagon that satisfies regularity, the sum of the lengths of two sides that are adjacent is the same as the sum of the lengths of the opposite pair of adjacent sides.

Proof. Let two adjacent sides of a regular hexagon be labelled a_1 and b_1, the opposite pair of adjacent sides be labelled a_2 and b_2, and the remaining two sides be labelled c_1 and c_2, as in Figure 6.15. Also, let the vertex between a_1 and b_1 be V_1 and the vertex between a_2 and b_2 be V_2. Then since c_1 and c_2 are parallel, the sum of the perpendicular distance from V_1 to each of c_1 and c_2 is the same as the sum of the perpendicular distance from V_2 to each of those sides; labelling these distances as in Figure 6.15, we get $x_1 + x_2 = x_3 + x_4$. Now, by the Law of Sines we have $a_1 = \frac{x_1}{sin(\frac{\pi}{3})}$, $b_1 = \frac{x_2}{sin(\frac{\pi}{3})}$, $a_2 = \frac{x_4}{sin(\frac{\pi}{3})}$, and $b_2 = \frac{x_3}{sin(\frac{\pi}{3})}$. Combining these equations, we get $a_1 + b_1 = a_2 + b_2$, as desired. QEF

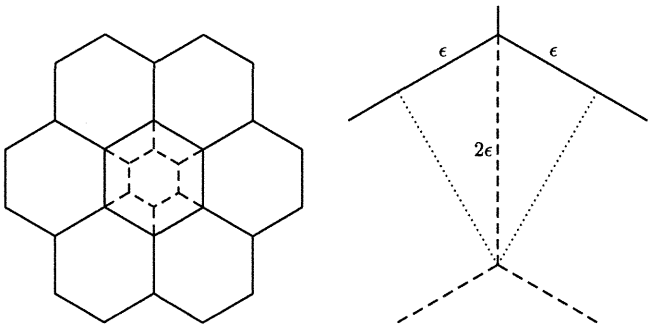


Figure 6.16: Shrinking or expanding components of a tiling of hexagons generates double bubbles enclosing different areas, but does not change perimeter.

Lemma 6.13. A 3-component straight-edge hexagon tiling enclosing arbitrary areas has the same perimeter as a 3-component straight-edge hexagon tiling of equal areas.

Proof. First, note that "shrinking" or "expanding" a single hexagon in a straight-edged hexagon tiling (as in Figure 6.16 is a smooth variation through equilibrium double bubbles and hence maintains perimeter.

Now consider the three unique hexagons in a 3-component straight-edge hexagon tiling, with boundary curves identified as in Figure 6.17. Let the nine unique sides, in three groups each of three sides parallel to one another, be labelled a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, and c_3, such that $a_1 \geq a_2 \geq a_3$, $b_1 \geq b_2 \geq b_3$, and $c_1 \geq c_2 \geq c_3$. Then if $a_1 = a_2 = a_3$, $b_1 = b_2 = b_3$, and $c_1 = c_2 = c_3$, the three hexagons are congruent and therefore have equal area. (In fact, any one of these equality equations implies the other two, by Lemma 6.12). We show that by shrinking or expanding hexagons we can transform any straight-edged hexagon tiling into one with congruent hexagons.

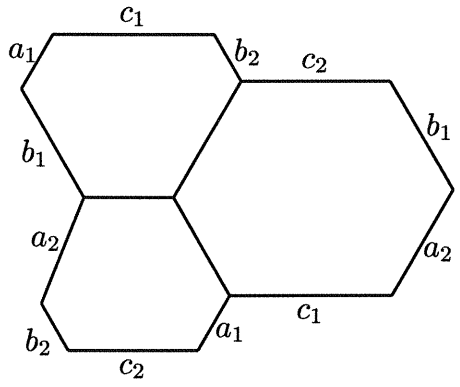


Figure 6.17: A generic fundamental region of a three-hexagon tiling looks something like this.

We start by showing that we can always label the nine sides in such a way that the hexagon H_1 with sides labelled a_1 and a_2 also has sides b_1, b_2, c_1, and c_2. Let the remaining sides of H_1 besides a_1 and a_2 be b_a, b_b, c_a, and c_b, as in Figure 6.17. Since $a_1 \geq a_2$, by Lemma 6.12 we know that $b_a \geq b_b$ and $c_a \geq c_b$. Similarly, since $a_1 \geq a_3$, we have $b_a \geq b_c$ and $c_a \geq c_c$, and since $a_2 \geq a_3$, we have $b_b \geq b_c$ and $c_b \geq c_c$. Thus we can always set $b_a = b_1$, $b_b = b_2$, $c_a = c_1$, $c_b = c_2$, and $c_c = c_3$ so that H_1 has sides with subscripts 1 and 2, while H_2 has sides all of subscript 1 or 3, and a hexagon H_3 has sides all of subscript 2 or 3.

Now consider a straight-edged three-hexagon tiling with hexagons of unequal area and side labels as described above. Then at least one of the inequalities in $a_1 \geq a_2 \geq a_3$ must be strict, or the tiling would be equal-area.

Suppose $a_1 > a_2$. Then know from Lemma 6.12 that $b_1 > b_2$ and $c_1 > c_2$. Now let us "expand" H_3 by lengthening a_2, a_3, b_2, b_3, c_2 and c_3 while shortening a_1, b_1, and c_1, until $a_1 = a_2$. We know we can do this smoothly, since a_1 is bounded below by a_2. Note also that by Lemma 6.12, $a_1 = a_2$ implies $b_1 = b_2$ and $c_1 = c_2$, so during the transformation we also have b_1 bounded from below by b_2, and c_1 bounded below by c_2. Thus none of the sides that we are shortening vanishes to zero length.

Now either $a_1 = a_2 = a_3$ and we are done, or $a_1 = a_2$ and $a_2 > a_3$. If the latter, then let us "shrink" H_1 by shortening a_1, a_2, b_1, b_2, c_1, and c_2 while lengthening a_3, b_3, and c_3, until $a_2 = a_3$. Again, we know we can do this smoothly, since a_1 and a_2 are bounded from below by a_3. As before, $a_2 = a_3$ implies $b_2 = b_3$ and $c_2 = c_3$, so b_1, b_2, c_1, and c_2 are all bounded from below as well.

Thus, by expanding H_3 and/or shrinking H_1 as necessary, we can continuously deform any straight-edged hexagon tiling into one with congruent hexagons, and thus equal areas. QEF

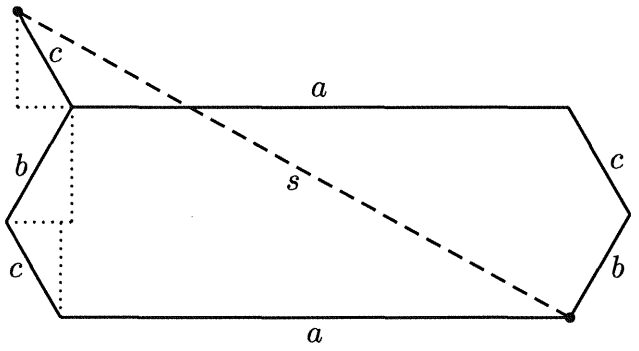


Figure 6.18: A closed geodesic in a three-hexagon tiling. It is a simple geometric argument to show that $a + b + c > s$, except in the case of the hexagonal torus, which implies that the hexagonal tiling in all other cases loses to the double band solution.

Lemma 6.14. A 3-component straight-edge hexagon tiling with congruent hexagons has perimeter strictly greater than 3 on all tori except the hexagonal torus.

Proof. A 3-hexagon tiling with congruent hexagons has three unique side lengths, $a \geq b \geq c$, and perimeter $P = 3(a + b + c)$. We show that unless $a = b = c$, $a + b + c$ is always greater than the length of a closed geodesic, and therefore the perimeter P is always greater than 3.

Consider some hexagon in such a tiling, and let V_1 be a vertex between sides of length a and b. Then the line connecting the diametrically opposed vertex V_2 (also between sides of length a and b) with the copy of V_2 that is connected to V_1 by a single side of length c, is a closed geodesic. See Figure 6.18.

Let s be the length of that geodesic. Then by the Pythagorean Theorem we have $s^2 = (\frac{\sqrt{3}}{2}c + \frac{\sqrt{3}}{2}b + \frac{\sqrt{3}}{2}c)^2 + (a + \frac{c}{2} - \frac{b}{2} + \frac{c}{2})^2$, which reduces to

$$\begin{aligned} s^2 &= a^2 + b^2 + 4c^2 + 2ac + 2bc - ab \\ &= (a^2 + b^2 + c^2 + 2ac + 2bc + 2ab) - 2ab + 3c^2 - ab \\ &= (a + b + c)^2 + 3c^2 - 3ab. \end{aligned}$$

Now suppose at least one of the inequalities in $a \geq b \geq c$ is strict. Then $c^2 < ab$, so $3c^2 - 3ab < 0$ and $(a + b + c)^2 > s^2$. Thus $P = 3(a + b + c) > 3s$, so the perimeter of the tiling is strictly greater than three unless $a = b = c$. Note from Figure 6.19 that the latter condition can only be satisfied on the hexagon torus, and that in this case the geodesic we constructed above is the shortest closed geodesic so the perimeter of the tiling is precisely 3. QEF

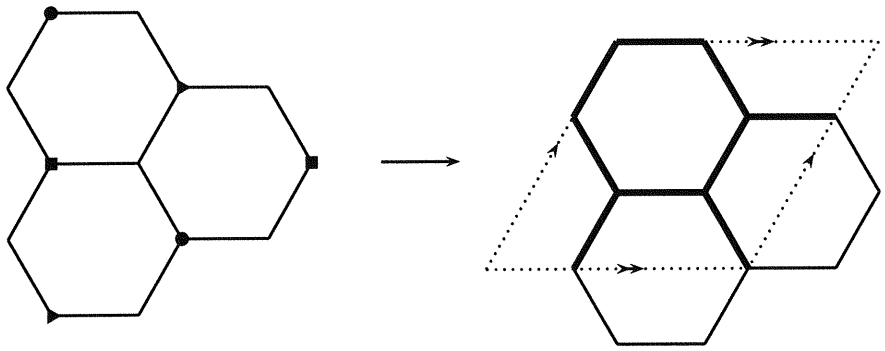


Figure 6.19: A three-hexagon tiling of congruent hexagons fits on the hexagonal torus, and is in fact a minimizer.

Proposition 6.13. The only hexagon tiling that is perimeter-minimizing is the 3-component straight-edge hexagon tiling on the hexagonal torus.

Proof. By Lemma 6.9, the number of components in a minimizing hexagon tiling must be a multiple of 3. By Lemma 6.10, a hexagon tiling with 9 or more components is not perimeter-minimizing. By Lemma 6.11, a hexagonal tiling with 6 components has perimeter strictly greater than 3, so by the perimeter bound it is not perimeter-minimizing. Thus, a perimeter-minimizing hexagon tiling must have exactly 3 components. By Proposition 6.12, a perimeter-minimizing hexagon tiling must have straight edges. By Proposition s 6.13 and 6.14, a 3-component straight-line hexagon tiling has perimeter strictly greater than 3 on all tori except the hexagonal torus, so by the perimeter bound the only hexagon tiling that can be perimeter-minimizing is the 3-component tiling on the hexagonal torus.

QEF

6.6 Formulas for Perimeter and Area

This section gives formulae for area and perimeter of the three potential minimizers from section 6.5: the standard double bubble Θ (Proposition 6.14), the two-component symmetric chain Ω (Propositions 6.15, 6.16), and the band-lens Φ (Proposition 6.18). We use these formulae to obtain the phase portrait of Figure 6.2. Propositions **??** and 6.19 show that the two-component symmetric chain and the band-lens are unique for prescribed areas.

6.6.1 Formulas for the Standard Double Bubble

We begin by stating a few simple geometric formulae that describe circular arcs meeting chords.

Remark 6.4. Given a circular arc λ and a chord of length C meeting the arc at (interior) angles θ, the area between the arc and the chord, the length of the arc, and the radius of the circle are given

by:

$$A(\theta, C) = \frac{C^2(\theta - \sin\theta\cos\theta)}{4\sin^2\theta} \tag{6.1}$$

$$L(\theta, C) = \frac{C(\theta)}{\sin\theta} \tag{6.2}$$

$$R(\theta, C) = \frac{C}{2\sin\theta} \tag{6.3}$$

Here area means the area between an arc and chord.

Formulae for perimeter and area of the standard double bubble as derived by Foisy [24, Section 2] are given in the following proposition.

Proposition 6.14. The area and perimeter formulae of a standard double bubble in terms of the separating chord length C and the angle θ at which it meets the interior arc are (smaller region R_1)

$$A_{R_1}(\theta, C) = A(\frac{2\pi}{3} - \theta, C) + A(\theta, C)$$

$$A_{R_2}(\theta, C) = A(\frac{2\pi}{3} + \theta, C) - A(\theta, C)$$

$$P(\theta, C) = L(\frac{2\pi}{3} + \theta, C) + L(\frac{2\pi}{3} - \theta, C) + L(\theta, C)$$

Proof. These formulae follow directly from Equations 6.1-6.3. QEF

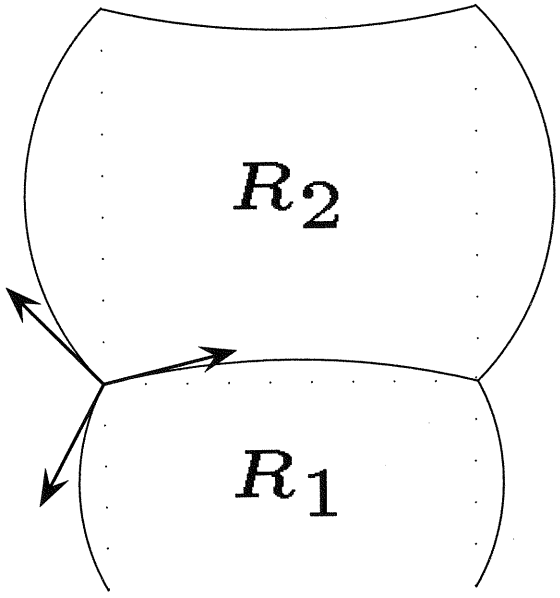


Figure 6.20: Important angles and lengths in a symmetric two component swath.

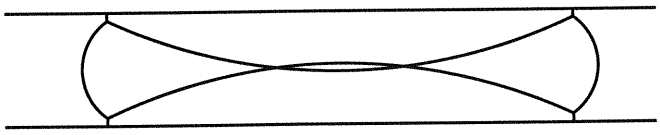


Figure 6.21: Arcs curve the right direction but intersect.

6.6.2 Formulas for the Standard 2-Component Symmetric Chain

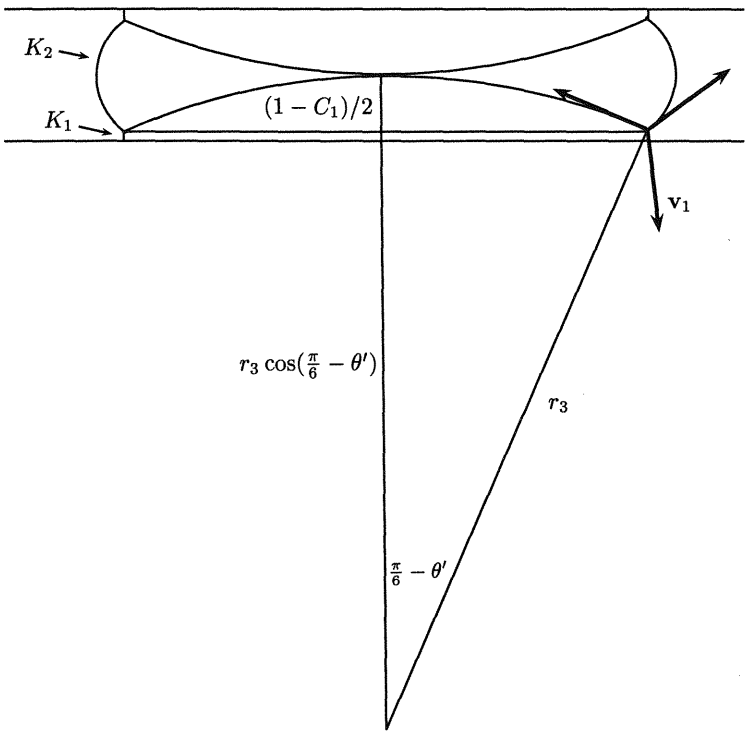


Figure 6.22: Conditions for tangency give the appropriate lower limit of the parameter θ_1.

Representing vertices by rectangles in a 2-component symmetric chain, gives a natural parameterization in terms of an angle θ_1 and a chord length C_1 determined by the boundary of the highest pressure region, R_1, with the exterior (Figure 6.20). We label this curve λ_1. Let θ_2 and C_2 be the corresponding parameters for each arc λ_2 between the exterior and R_2. Similarly, θ_3 and C_3 describe arcs λ_3 between R_1 and R_2.

Given θ_1 and C_1, we can determine κ_1. The λ_i's meet at 120°, so we get that $\theta_2 = \frac{\pi}{3} - \theta_1$ and $\theta_3 = \frac{\pi}{6} - \theta_1$. Since $C_1 + C_2 = 1$, we recover C_3 from the radius formula above using the fact that curvatures add to zero at each vertex. In sum, we have:

$$C_2 = 1 - C_1;$$

$$\kappa_3 = \kappa_1 - \kappa_2;$$

$$\frac{1}{\kappa_1} = \frac{C_1}{2\sin\theta_1};$$

$$\frac{1}{\kappa_2} = \frac{C_2}{2\sin(\frac{\pi}{3} - \theta_1)};$$

$$\frac{1}{\kappa_3} = \frac{C_3}{2\sin(\frac{\pi}{6} - \theta_1)}.$$

In terms of C_1 and θ_1, the formula for C_3 becomes

$$C_3 = \frac{C_1(1 - C_1)\sin(\frac{\pi}{6} - \theta_1)}{(1 - C_1)\sin\theta_1 - C_1\sin(\frac{\pi}{3} - \theta_1)}.$$

Now we must specify a range of values for θ_1 and C_1 where the swath makes sense, i.e. we find a parameter space where swaths do not violate regularity.

By symmetry, we can restrict C_1 to the open interval $(0, \frac{1}{2})$ (We exclude the endpoints to ensure uniqueness, and deal with these cases separately.) If we consider a frame of 3 vectors v_1, v_2, v_3 meeting at 120° we can model the set of tangent vectors to each arc at a given vertex (see Figure 6.21). In particular, we can place this frame at the intersection of the λ_i's and consider the structural change of a bubble under a small rotation of the frame. In the case of $C_1 = \frac{1}{2}$, v_3 must coincide with C_3 otherwise λ_3 will bow into the region of higher pressure. For instance, rotating the frame slightly counterclockwise makes $\kappa_2 > \kappa_1$ which implies R_2 has higher pressure with respect to the exterior. Since v_3 points into R_2, this construction cannot work. For an arbitrary C_1 (less than $\frac{1}{2}$), we rotate the frame until $\kappa_1 = \kappa_2$, this particular construction cannot work since v_3 must point horizontal if the regions have equal pressure. Rotating the frame slightly counterclockwise presents the same problem as before. If we rotate clockwise, we may initially produce a chain which intersects itself (see Figure 6.21)). At a certain value θ' of θ_1 the $\lambda_3's$ are tangent and beyond that, until $\theta_1 = \frac{\pi}{6}$,

the construction is regular. (Unless $C_1 = \frac{1}{2}$ the angle $\theta_1 = \frac{\pi}{6}$ is not valid. Again, we treat this case separately.)

We see from figure 6.20) that θ' satisfies

$$r_3 \cos(\frac{\pi}{6} - \theta') + \frac{1 - C_1}{2} = r_3$$

Upon further simplification, this condition reduces to

$$C_1 = \sin\theta'$$

Thus, θ_1 must lie in $(\sin^{-1}(C_1), \frac{\pi}{6})$. We can now write down formulae for the area and perimeter of chains.

Proposition 6.15. Using the parameterization of Section 6.6.2, the area and perimeter formulae for symmetric chains with components of unequal pressure are

$$P(\theta_1, C_1) = 2L(\theta_1, C_1) + 2L(\frac{\pi}{3} - \theta_1, C_2) + 2L(\frac{\pi}{6} - \theta_1, C_3)$$

$$A_{R_1}(\theta_1, C_1) = 2A(\theta_1, C_1) + 2A(\frac{\pi}{6} - \theta_1, C_3) + C_1C_3$$

$$A_{R_2}(\theta_1, C_1) = 2A(\frac{\pi}{3} - \theta_1, C_2) - 2A(\frac{\pi}{6} - \theta_1, C_3) + C_2C_3$$

where C_2 and C_3 are given by

$$C_2 = 1 - C_1$$

$$C_3 = \frac{C_1(1 - C_1)\sin(\frac{\pi}{6} - \theta_1)}{(1 - C_1)\sin\theta_1 - C_1\sin(\frac{\pi}{3} - \theta_1)}.$$

Proof. These formulae follow directly from the parameterization and Equations 6.1-6.2. QEF

Now we must treat the case where $C_1 = \frac{1}{2}$. We showed above that $\theta_1 = \frac{\pi}{6}$ at this point, otherwise the construction is irregular. However, this does not determine a unique value for C_3, in fact, it can take on any value greater than zero. Since all other arc lengths and angles in the chain are fixed, we have a simple formula relating volume and perimeter of such chains.

Proposition 6.16. In symmetric chains comprised of equal pressure components, perimeter as a function of the area of one region is given by

$$P(A_{R_1}) = 4A + \frac{2\pi + 3\sqrt{3}}{6}$$

where $A_{R_1} > 2A(\frac{\pi}{6}, \frac{1}{2}) = \frac{2\pi - 3\sqrt{3}}{24}$

Proof. By Equations 6.1-6.2, we know the perimeter and area in terms of C_3

$$P = 4L(\frac{\pi}{6}, \frac{1}{2}) + 2C_3 = \frac{2\pi}{3} + 2C_3$$

$$A_{R_1} = A_{R_2} = 2A(\frac{\pi}{6}, \frac{1}{2}) + \frac{C_3}{2} = \frac{2\pi - 3\sqrt{3}}{24} + \frac{C_3}{2}$$

This allows us to get perimeter as a function of area. Combining equations yields

$$P = 4A_{R_1} + \frac{2\pi + 3\sqrt{3}}{6}$$

For, $A_{R_1} > 2A(\frac{\pi}{6}, \frac{1}{2}) = \frac{2\pi - 3\sqrt{3}}{24}$, the chain is regular QEF

Conjecture 6.3. Given prescribed areas a_1 and a_2, and up to isometries of the torus, there is a unique standard two-component symmetric chain with homology $(1, 0)$ enclosing these areas.

Note that even if this conjecture fails, all we really need is the weaker result that the chain is unique *when it is minimizing*. This is probably not at all hard to show.

Proposition 6.17. For a given set of area pairs, and based on the parameterization described in this subsection, if the perimeter of an immersed symmetric is less than or equal to the perimeter of the embedded minimizer for those areas, then it is itself embedded.

Proof. We must consider the case when the symmetric chain overlaps itself around the torus. Clearly, C_3 must be less than .5, or else the double bubble would have perimeter greater than three. C_3 is perpendicular to the shortest side of the torus, so $C_3 < L \cdot \sin(\theta)$, where θ is the smallest interior angle of the torus, since otherwise two vertices which fall on a geodesic, g_1, in the short direction,

would wrap around past the other two vertices (which also lie on a geodesic, g_2) and there would be no exterior. This case falls outside of our calculation and thus we do not need to consider it.

Thus, there exists a distance $\epsilon = L \cdot \sin(\theta) - C_3 > 1 \cdot \sin(\frac{\pi}{3}) - 1/2 > .365$ that separates g_1 and g_2. But somewhere in the parallelogram that lies between g_1 and g_2, the double bubble intersects itself. Now there are three cases. Case 1 is when R_1 or R_2 intersects one of the vertices of the chain. Case 2 is when R_1 intersects R_2. Case 3 is when R_1 or R_2 intersects itself.

Case 1: Assume that R_1 intersects one of the vertices of the chain. Thus, given λ_1 and the height, h_1 that it makes with its chord, then $h_1 > \epsilon$.

Now, we'd like to calculate the length of λ_1,

$$L(\lambda_1) = \frac{C_1 \cdot \theta_1}{\sin(\theta_1)}.$$

But

$$\tan(\theta_1/2) = \frac{h_1}{(C_1/2)}.$$

So

$$C_1 = \frac{2 \cdot h_1}{\tan(\theta_1/2)} > .73/tan(\theta_1/2).$$

Thus,

$$L(\lambda_1) > \frac{2 \cdot \epsilon \cdot \theta_1}{\tan(\theta_1/2) \cdot \sin(\theta_1)}.$$

We put

$$f(x) = \frac{x}{\tan(x/2)\sin(x)} = \frac{x \cdot \cos(x) + x}{\sin(x)^2}.$$

Then

$$f'(x) = \frac{-(x \cdot (\cos(x))^2 - (\sin(x) - 2 \cdot x) \cdot \cos(x) - \sin(x) + x)}{\sin(x)^3}.$$

But since $x > sin(x)$ for $x > 0$, on the interval $[0, \pi/3]$, $f'(x) < 0$ and thus f is decreasing. Thus,

$$L(\lambda_1) > \frac{.73 \cdot \pi/3}{\tan(\pi/6) \cdot \sin(\pi/3)} > 1.528,$$

and so $P > 2 \cdot 1.528 > 3$.

Case 2: Assume that R_1 intersects R_2 but neither intersects one of the vertices of the chain. Thus, given λ_1 and λ_2, and the height, h_1 and h_2 that each makes with its respective chord, either $h_1 + h_2 > \epsilon$.

Now, we would like to calculate the sum of the lengths of λ_1 and λ_2,

$$L(\lambda_1) + L(\lambda_2) = \frac{C_1 \cdot \theta_1}{\sin(\theta_1)} + \frac{C_2 \cdot \theta_2}{\sin(\theta_2)}$$

. But

$$\tan(\theta_i/2) = \frac{h_i}{(C_i/2)}.$$

So

$$C_i = \frac{2 \cdot h_i}{\tan(\theta_i/2)}.$$

Thus,

$$L(\lambda_1) + L(\lambda_2) = \frac{2 \cdot h_1 \cdot \theta_1}{\tan(\theta_1/2) \cdot \sin(\theta_1)} + \frac{2 \cdot h_2 \cdot \theta_2}{\tan(\theta_2/2) \cdot \sin(\theta_2)}.$$

But, from above

$$f(x) = \frac{x}{\tan(x/2)\sin(x)} = \frac{x \cdot \cos(x) + x}{\sin(x)^2}$$

is decreasing. So if we assume, without loss of generality that $\theta_1 \geq \theta_2$, then

$$L(\lambda_1) + L(\lambda_2) \geq \frac{2 \cdot h_1 \cdot \theta_1}{\tan(\theta_1/2) \cdot \sin(\theta_1)} + \frac{2 \cdot h_2 \cdot \theta_1}{\tan(\theta_1/2) \cdot \sin(\theta_1)} > \frac{2 \cdot \epsilon \cdot \theta_1}{\tan(\theta_1/2) \cdot \sin(\theta_1)} > 1.528$$

for all θ in the interval $[0, \pi/3]$, as above. So $P > 2 \cdot 1.528 > 3$.

Case 3: Now assume that R_1 intersects itself. They will clearly intersect in a lens. We can remove the lens, and get the area back with less perimeter in the exterior. (This is where I wanted to have the isoperimetric inequality for free boundary closed surfaces - I employ a different method that needs tuning.) To see this, align the lens perpendicular to the axis of symmetry of the chain and center it over the ϵ-rectangle and where the exterior is widest. If the exterior is not wider at this point than the lens, separate the two circular arcs of the lens and slide them apart until the intersection of these arcs with the exterior encloses the same area that the lens had, and with less perimeter. If this

component of the exterior does not have enough area, take the whole component and repeat with a different component. If there is not enough area in the whole region, then the immersed double bubble lies outside our calculations. We must still consider the case where the exterior is too wide. If the lens fits, we are done. If not, then... I can't explain this without a picture. I will get on that for next time. QEF

This completes the description of symmetric chains. Note that when computing perimeter later, we restrict our parameter space even more by requiring $C_3 < \frac{1}{2}$, since the polygonal arc length of chains not satisfying this inequality is already greater than 3.

6.6.3 Formulae for the Band-Lens

The family of Band-Lens bubbles can be parameterized nicely in terms of two parameters, r, the radius of curvature ($\frac{1}{curvature}$) of the lens, and d, the width of the band. Of course, the lens can only be as tall as the cylinder, so r can be at most $\frac{1}{2\sin\frac{\pi}{3}} = \frac{1}{\sqrt{3}}$. For a given r, the width of the band must be at least half the width of the lens. Thus d is at least $\frac{r}{2}$. From this we can compute the area and perimeter of a Band-Lens.

Proposition 6.18. The area and perimeter formulae for a band-lens bubble in terms of r and d ($r < \frac{1}{\sqrt{3}}, d > \frac{r}{2}$) are

$$A_{Lens}(r) = 2r^2(\frac{\pi}{3} - \frac{\sqrt{3}}{4})$$

$$A_{Band}(r,d) = d - r^2(\frac{\pi}{3} - \frac{\sqrt{3}}{4})$$

$$P(r,d) = (\frac{4\pi}{3} - \sqrt{3})r + 2$$

Proof. Using Equations 6.1-6.2 we can compute the area of both components. The chord length C of the lens is $2r\sin\frac{\pi}{3} = \sqrt{3}r$. This gives us

$$A_{Lens}(r) = 2A(\frac{\pi}{3}, \sqrt{3}r) = 2r^2(\frac{\pi}{3} - \frac{\sqrt{3}}{4})$$

$$A_{Band}(r,d) = d - \frac{1}{2}A_{Lens}(r) = d - r^2(\frac{\pi}{3} - \frac{\sqrt{3}}{4})$$

The perimeter is given by

$$P(r,d) = 2L(\frac{\pi}{3}, \sqrt{3}r) + (2 - \sqrt{3}r) = (\frac{4\pi}{3} - \sqrt{3})r + 2$$

QEF

Proposition 6.19. Given prescribed areas a_1 and a_2, there exists a unique (up to isometry of the torus) minimal band-lens.

Proof. The lens is uniquely determined by the length, C, of the chord separating its vertices, and the interior angle θ between the two boundary arcs. By regularity θ is fixed. Moreover, the area of the lens is strictly increasing in C. Thus, given the smallest area A_1 of the three regions of the double bubble, there is a unique lens enclosing that area. The chord of the lens must lie parallel to the short side of the torus and thus the lens is positioned uniquely up to translation. Further, the vertices are connected by a uniquely determined line segment also parallel to the short direction of the torus and there exists a unique geodesic in the short direction which will enclose the proper area. QEF

Proposition 6.20. For a given set of area pairs, and based on the parameterization described in this subsection, if the perimeter of an immersed band-lens is less than or equal to the perimeter of the embedded minimizer for those areas, then it is itself embedded.

Proof. The first case we consider is the case in which the band with smaller area is too narrow, so that the shortest closed geodesic which is a component of its boundary intersects the boundary of the lens. But we can create a new, embedded, double bubble by eliminating the section of the geodesic which intersects the lens and sliding the rest of the geodesic inward to compensate for area. But this double bubble has perimeter less than the band lens, and further, contradicts regularity. So for such "difficult" area pairs, the band-lens is not a minimizer.

The second case to be considered is the case in which the lens does not fit along the short direction of the torus. In this case, the double bubble clearly has perimeter greater than three, and by Proposition 6.3, it cannot be a minimizer. Thus, if the immersed band-lens would minimize perimeter, then it must be embedded. QEF

6.7 Perimeter Minimizing Double Bubbles on the Flat 2-Torus and the Infinite Cylinder

Main Theorem 1. Each perimeter minimizing double bubble on a flat 2-torus is one of the following six types, depending on the areas to be enclosed:

i) the standard double bubble, Θ,

ii) the band-lens, Φ,

iii) the double band, Ξ,

iv) a symmetric two-component chain Ω,

v) the hexagon tiling, H, on the hexagonal torus, or

vi) an octagon-square tiling, X.

Conjecture 1 says that the octagon-square tiling is never perimeter-minimizing.

Proof. Proposition 6.6 shows that a minimizer must be either a a contractible double bubble, a tiling, a swath, a band adjacent to a set of components whose union is contractible, or the double band (after perhaps relabelling the two regions and the exterior).

First we consider all minimizers that do not fall in the class of tilings.

By Proposition 6.1, the only potential contractible minimizer is the standard double bubble Θ. By Proposition 6.10, the only potential swath minimizer is the symmetric two-component chain Ω. By Proposition 6.11, the only potential minimizer among bands adjacent to a set of components whose union is contractible is the band-lens Φ.

We focus here on potential minimizers that are tilings. Propositions 6.12 and 6.13 tell us that such a tiling must be either the special hexagon tiling H or an octagon-square tiling. QEF

Lemma 6.5. On the flat 2-torus, there exist area pairs for which the band-lens is perimeter minimizing, and separately for which the standard double bubble is perimeter minimizing.

Proof. Consider regions R_1 and R_2 with area A_1 and A_2 in a perimeter minimizing double bubble. By [21] , the perimeter minimizing single bubble for area A_1 is either a single band, a circle, or the complement of a circle, depending on the area A_1. Thus, in the limit through minimizing double bubbles as A_2 goes to zero, we must get back the single band, circle, or complement of a circle enclosing R_1. This is only possible if for very small A_2, the minimizing double bubble is the band-lens or the standard double bubble which will in the limit be the single band or the circle or complement of a circle, respectively. QEF

Corollary 6.1. Each perimeter minimizing double bubble on a torus with unit shortest side length and longest side of length greater than $\sqrt{3}$ is of one of the following four types of Figure 6.1, depending on the areas to be enclosed:

i) the standard double bubble Θ,

ii) the band-lens Φ,

iii) the symmetric two-component chain Ω, or

iv) the double band Ξ formed by three geodesic circles.

Proof. By Proposition 6.4, no tiling can occur on a torus with unit shortest side length and longest side length greater than $\sqrt{3}$. Thus, by the Main Theorem, each perimeter minimizing double bubble must be of one of the above types.

QEF

Corollary 6.2. Each perimeter minimizing double bubble on the infinite cylinder of unit circumference is of one of the following four types of Figure 6.1, depending on the areas to be enclosed:

i) the standard double bubble Θ,

ii) the band-lens Φ,

iii) the symmetric two-component chain Ω, or

iv) the double band Ξ formed by three geodesic circles.

Proof. Clearly there cannot be a tiling on the infinite cylinder. As for the non-tiling minimizers, all proofs carry over from the proof of the main theorem. This is because locally, the cylinder looks like a section of the torus. And none of the proof strategies require "wrapping" around on the torus, where something could go on to infinity on the cylinder.

QEF

Remark 6.5. Note that on the infinite cylinder and the torus with longest side greater than $\sqrt{3}$, each of the four minimizers occurs. Since we have paramterized each of them, the phase plots in Figure 6.2 provide a complete description of the minimizers for each area pair.

6.8 Numerical comparisons between Minimizer Finalists

Now that we have established what the minimizers can be, we will do a numerical comparison to show which double bubbles occur as minimizers given a arbitrary area pairs. To do this, we plot perimeter versus area for each minimizer (using the parameterizations from the previous section) and numerically find the actual minimizer. Figure 6.2 contains 2-d plots showing where each minimizer occurs for the whole parameter space. Further, in Figure 6.23 we have plotted the intersections of the standard double bubble graph and the band-lens graph as evidence that these surfaces do in fact intersect in well-defined curves.

Figure 6.23: This graph of the intersection of the area for the standard double bubble and the band-lens shows that the boundaries between regions in the phase plots are likely to be smooth curves.

It is important to note that because we have not parameterized the octagon square, we can not *a priori* assume that each of these potential minimizers is actually a minimizer since the octagon square *might* beat everything else. Lemma 6.5 proves that the band-lens and the standard double bubble do in fact occur as minimizers.

Chapter 7

t3

Charter. We present a conjecture, based on computational results, on the area minimizing way to enclose and separate two arbitrary volumes in the flat cubic three-torus $\mathbf{T}^3$. For comparable small volumes, we prove that an area minimizing double bubble in $\mathbf{T}^3$ is the standard double bubble from $\mathbf{R}^n$.

7.1 Introduction

Our Central Conjecture 2.1 states that the ten different types of two-volume enclosures pictured in Figure 1 comprise the complete set of surface area minimizing *double bubbles* in the flat cubic three-torus $\mathbf{T}^3$. Our numerical results, summarized in Figure 7.2, indicate the volumes for which we conjecture that each type of double bubble minimizes surface area. Our main theorem, Theorem 4.1, states that given any fixed ratio of volumes, for small volumes, the minimizer is the standard double bubble. This result applies to any smooth flat Riemannian manifold of dimension three or four with compact quotient by its isometry group.

The *double bubble problem* is a two-volume generalization of the famous isoperimetric problem. The isoperimetric problem seeks the least-area way to enclose a single region of prescribed volume. About 200 BC, Zenodorus argued that a circle is the least-perimeter enclosure of prescribed area in the plane (see [19]). In 1884, Schwarz [47] proved by symmetrization that a round sphere minimizes perimeter for a given volume in $\mathbf{R}^3$. Isoperimetric problems arise naturally in many areas of modern mathematics. Ros [44] provides a beautiful survey.

Soap bubble clusters seek the least-area (least-energy) way to enclose and separate several given volumes. Bubble clusters have served as models for engineers, architects, and material scientists (Emmer [13] surveys some architectual applications, and the text by Weaire and Hutzler [57], an introduction to the physics of space-filling bubble clusters or *foams*, discusses numerous other applications).

Existence and regularity The surface area minimizing property of bubble clusters can be codified mathematically in various useful ways, using the rectifiable currents, varifolds, or $(\mathbf{M}, \epsilon, \delta)$-minimal sets of geometric measure theory (see [34]). In three dimensions, mathematically idealized bubble clusters consist of constant-mean-curvature surfaces meeting smoothly in threes at 120°along smooth curves, which meet in fours at a fixed angle of approximately 109° ([53, Theorems II.4, IV.5, IV.8], or see [34, Section 13.9]).

Bubbles in the three-torus In comparison with $\mathbf{R}^3$ and $\mathbf{T}^2$, the double bubble problem in $\mathbf{T}^3$ appears to be more difficult. In the torus it is not possible to push through a component bound like Hutchings', since a key step in his proof is to show that the double bubble has an axial symmetry. Nor is a variational bound after Wichiramala forthcoming, due to the additional topological complications in three dimensions. Until some new approach provides a component bound, there probably

will be no definitive results. Indeed, the single bubble for the three-torus is not yet completely understood, although there are partial results. The smallest enclosure of half of the volume of the torus was shown by Barthe and Maurey [2, Section 3] to be given by two parallel two-tori. Morgan and Johnson [36, Theorem 4.4] show that the least-area enclosure of a small volume is a sphere. Spheres, tubes around geodesics, and pairs of parallel two-tori are shown to be the only types of area minimizing enclosures for most tori by the work of Ritoré and Ros ([42, Theorem 4.2], [43]).

Theorem about small volumes Theorem 4.1 states that any sequence of area minimizing double bubbles of decreasing volume and fixed volume ratio has a tail consisting of standard double bubbles. The central difficulty is to bound the curvature. This accomplished, we show that the bubble lies inside some small ball that lifts to $\mathbf{R}^3$, where a minimizer is known to be standard [23]. The result extends to any flat 3- or 4-manifold with compact quotient by the isometry group.

Our proof goes roughly as follows. From the original sequence of double bubbles we generate a new sequence by rescaling the manifold at each stage so that one of the volumes is always equal to one. We can then apply compactness arguments and area estimates to the rescaled sequence to show that certain subsequences of translates have non-trivial limits. These limits are used to obtain a curvature bound on the original sequence. With such a bound, we can apply monotonicity to conclude that if the volumes are small, the double bubble is contained in a small ball. We conclude that it must be the same as the minimizer in $\mathbf{R}^3$ or $\mathbf{R}^4$, *i.e.* that it must be the standard double bubble by [23] or [41].

Plan of the Paper. Section 2 reviews the methods leading to our Central Conjecture 2.1 and to Figures 7.1 and 7.2. Section 3 surveys some subconjectures. Section 4 focuses on the proof of Theorem 1 on small volumes. Section 5 shifts from the cubic to other tori and discusses other conjectures

and candidates, including a "Hexagonal Honeycomb."

7.2 The Conjecture

Generating Candidates. Many possibilities for double bubbles in the three-torus $\mathbf{T}^3$ were proposed in brainstorming sessions by participants in the Clay/MSRI Summer School. In order to classify the candidates we used the following method. Starting with a standard double bubble, we imagined one of the two volumes growing until the bubble enclosing it wrapped around the torus and encountered an obstruction. Following the principles of regularity for bubble clusters, if a bubble collided with itself we opened the walls up, whereas if two different bubbles collided we allowed them to stick together. We then repeated this procedure for these new double bubbles, sometimes changing the perspective slightly or becoming a bit more fanciful (e.g. the Center Cylinder of Figure 7.1 or Gary Lawlor's Fire Hydrant of Figure 7.3, also known as "Scary Gary").

Producing the Phase Diagram. Brakke's *Surface Evolver* [5] was used to closely approximate the minimal area that a double bubble of each type needs to enclose specified volumes. Our initial simulations gave us the approximate surface area for each candidate double bubble, tested on partitions $v_1 : v_2 : v_3$ of a unit volume taken in increments of 0.01. From the data obtained in these simulations, we found the least-area competitor for each volume triple. Figure 7.1 shows the candidates we found to be minimizing for some set of volumes. The phase diagram appearing in Figure 7.2 is the result of refining our initial computations along the boundaries with a 0.005 increment.

Central Conjecture 1. The ten double bubbles pictured in Figure 7.1 represent each type of surface area minimizing two-volume enclosure in a flat, cubic three-torus, and these types are minimizing

for the volumes illustrated in Figure 7.2.

Comments. One might expect that minimizers would be found among the various regularity-satisfying conglomerations of topological spheres and homotopically non-trivial tori, other possibilities being excessively complex. This was borne out in our computations. It is interesting to note, however, that not all of the simple possibilities appeared as minimizers (at least one of the enclosures in Figure 7.1 was always more efficient than the Transverse Cylinders pictured in Figure 7.3). The various double bubbles of Figure 7.3, while stable for a certain range of volumes, were never area minimizing. A challenging unsolved problem is to find all of the stable non-minimizing bubbles. Note that a given type from Figure 7.1 might be stable for a much wider range of volumes than those for which it actually minimizes surface area.

It is worth observing that all of the conjectured minimizers for the double bubble problem on $\mathbf{T}^2$ (a Standard Double Bubble, Band with Lens, Symmetric Chain, and Double Band) are echoed here in at least two ways. The Double Cylinder, the Slab Cylinder, the Double Slab, and the Cylinder String are $\mathbf{T}^2$ minimizers $\times$ $\mathbf{T}^1$. There are also more direct analogues, as is seen by comparing, for example, the three- and two-dimensional Standard Double Bubbles, or the Délauney and Symmetric Chains. See Corneli *et al.* [9] for more on the $\mathbf{T}^2$ minimizers.

7.3 Subconjectures

We now present a list of natural subconjectures, suggested either by the phase diagram or by examination of the pictures made with *Surface Evolver*.

One immediate observation is that the edges of our phase diagram appear to characterize single bubbles in $\mathbf{T}^3$.

Conjecture 7.1 (Ritoré and Ros [42], [43], [44]). The optima for the isoperimetric problem in a cubic $\mathbf{T}^3$ are the sphere, cylinder, and slab.

We still do not know proofs for the following intuitive conjectures about the double bubble problem (although Theorem 4.1 gives partial results on Conjecture 3.3):

Conjecture 7.2. An area-minimizing double bubble in $\mathbf{T}^3$ has connected regions and complement.

Conjecture 7.3. For two small volumes the standard double bubble is optimal.

Conjecture 7.4. For one very small volume and two moderate volumes, the Slab Lens is optimal.

Conjecture 7.5. The first phase transition as small equal, or close to equal, volumes grow is from the standard double bubble to a chain of two bubbles bounded by Délaunay surfaces.

Délaunay surfaces are constant-curvature surfaces of rotation, and as such have full rotational symmetry. From the *Surface Evolver* pictures, it appears that the conjectured surface area minimizers always have the maximial symmetry, given the constraints, a fact which leads to the following natural conjecture.

Conjecture 7.6. A minimizer is as symmetric as possible, given its topological type.

We will conclude this section with a proposition which represents the first step towards establishing a symmetry property. Specifically, we prove that for any double bubble, there is a pair of parallel planes (actually two-tori) that cut both regions in half. Hutchings [22, Theorem 2.6] used the fact that in $\mathbf{R}^n$ a double bubble has two perpendicular planes that divide both regions in half. This is an easy step in a difficult proof that the function that gives the least-area to enclose two given volumes is concave. Before proving a basic result of a similarly elementary flavor for the torus, we mention that we conjecture that the much deeper concavity result also holds:

Conjecture 7.7. The least area to enclose and separate two given volumes in the three-torus is a concave function of the volumes.

If one could prove Conjecture 3.7, one would then be able to apply other ideas in Hutchings' paper [22, Section 4] to obtain a functional bound on the number of components of a minimizing bubble.

Proposition 7.1 (Deluxe Ham Sandwich Theorem). In a rectangular torus, if a double bubble lies inside a cylinder $S^1 \times D^2$, then there is a plane that cuts both volumes in half.

Proof. This is a generalization of the standard argument for the ham sandwich theorem in Euclidean space. Assume that the identification occurs in the vertical direction. Take a circle in the xy plane indexed by $\alpha \in [0, \pi]$. Then for each α, rotate the surface by α in the xy plane and consider the family of pairs of flat tori parallel to the xz plane that are at distance 1/2 from each other. Then there is at least one such pair of planes that cuts the volume V_1 in half, for each α. There may be an interval of such pairs for a given α. However, some one parameter family of these planes may be chosen which varies continuously as a function of α, and one such plane cuts V_2 in half. QEF

7.4 Small Volumes

Conjecture 7.3 stated that small volumes are best enclosed by a standard double bubble. Theorem 4.1 proves that for any fixed volume ratio, the standard double becomes optimal when the volumes are sufficiently small. This result holds for many flat three- and four-dimensional manifolds (and see Remark 7.2).

The standard double bubble consists of three spherical caps meeting at 120 degrees. (If the volumes are equal, the middle surface is planar.) It is known to be minimizing in $\mathbf{R}^3$ [23] and $\mathbf{R}^4$ [41].

Theorem 7.1. Let M be a smooth flat Riemannian manifold of dimension three or four, such that M has compact quotient by the isometry group. Fix $\lambda \in (0, 1]$. Then there is an $\epsilon > 0$, such that if $0 < v < \epsilon$, a perimeter-minimizing double bubble of volumes $v, \lambda v$ is standard.

Remark 7.1. Solutions to the double bubble problem exist for all volume pairs in manifolds with compact quotient by their isometry group. The proof is the same as in Morgan [34, Section 13.7].

For the proof we will regard a double bubble as a pair of 3- or 4−dimensional rectifiable currents,

R_1 and R_2, each of multiplicity one, of volumes $V_1 = \mathbf{M}(R_1)$ and $V_2 = \mathbf{M}(R_2)$. The total perimeter of such a double bubble is $\frac{1}{2}(\mathbf{M}(\partial R_1) + \mathbf{M}(\partial R_2) + \mathbf{M}(\partial(R_1 + R_2)))$. Here $\mathbf{M}$ denotes the mass of the current, which can be thought of as the Hausdorff measure of the associated rectifiable set (counting multiplicities). For a review of the pertinent definitions, see notes from Morgan's course at the Summer School [37] or the texts by Morgan [34] or Federer [14].

Before proceeding it is helpful to fix some notation. By the Nash and subsequent embedding theorems, we may assume M is a submanifold of some fixed $\mathbf{R}^N$. We will consider a sequence of perimeter-minimizing double bubbles in M, containing the volumes v and λv, as $v \to 0$. For each v, M_v will denote $s_v(M)$ in $s_v(\mathbf{R}^N)$, where s_v is the scaling map that takes regions with volume v to similar regions with volume 1. In particular, s_v maps our perimeter-minimizing double bubble containing volumes v and λv to a double bubble which we call S_v which contains volumes 1 and λ, and which is of course perimeter-minimizing for these volumes.

We focus on the case of dimension three; the proof for dimension four is essentially identical.

Lemma 7.1. There is a $\gamma > 0$ such that if R is a region in an open Euclidean 3-cube K and $\mathrm{vol}(R) \leq \mathrm{vol}(K)/2$, then

$$\mathrm{area}(\partial R) \geq \gamma(\mathrm{vol}(R))^{2/3}.$$

Proof. Let γ_0 be such an isoperimetric constant for a cubical 3-torus, so that $\mathrm{area}(\partial P) \geq \gamma_0(\mathrm{vol}(P))^{2/3}$ for all regions $P \subseteq \mathbf{T}^3$. Such a γ_0 exists by the isoperimetric inequality for compact manifolds [34, Section 12.3]. Make the necessary reflections and identifications of the cube K to obtain a torus containing a region R' with eight times the volume and eight times the boundary area of R. The claim follows, with $\gamma = \gamma_0/2$. QEF

Proof of Theorem 4.1. The first step is to show that the sequence S_v, suitably translated and rotated, has a subsequence that converges as $v \to 0$ and has $V_1 \neq 0$ in the limit. Our argument also shows that there is a subsequence that converges to a limit with $V_2 \neq 0$, but does not show that

there is a subsequence where both volumes are non-zero in the limit. This is because while we are exerting ourselves trapping the first volume in a ball, the second one may wander off to infinity.

We first show the existence of a covering $\mathcal{K}_v$ of M_v with bounded multiplicity, consisting of cubes contained in M_v, each of side-length L. Lemma 7.1 will give us a positive lower bound on the volume of the part of R_1 that is inside one of these cubes for each S_v. We will then apply a standard compactness theorem to show that a subsequence of the S_v's converges.

Take a maximal packing by balls of radius $\frac{1}{4}L$. Enlargements of radius $\frac{1}{2}L$ cover M_v. Circumscribed cubes of edge-length L provide the desired covering $\mathcal{K}_v$. To see that the multiplicity of this covering is bounded, consider a point $p \in M_v$. The ball centered at p with radius $2L$ contains all the cubes that might cover it, and the number of balls of radius $\frac{1}{4}L$ that can pack into this ball is bounded, implying that the multiplicity of $\mathcal{K}_v$ is also bounded by some $m > 0$.

Now consider some such covering, with $L = 2$. By Lemma 7.1, there is an isoperimetric constant γ such that

$$\text{area}(\partial(R_1 \cap K_i)) \geq \gamma(\text{vol}(R_1 \cap K_i))^{2/3},$$

and therefore, since $\max_k \text{vol}(R_1 \cap K_k) \geq \text{vol}(R_1 \cap K_i)$ for any i,

$$\text{area}(\partial(R_1 \cap K_i)) \geq \gamma \frac{\text{vol}(R_1 \cap K_i)}{(\max_k \text{vol}(R_1 \cap K_k))^{1/3}}. \tag{7.1}$$

Note that the total area of the surface is greater than $1/m$ times the sum of the areas in each cube, and the total volume enclosed is less than the sum of the volumes, so summing Equation 7.1 over all the cubes K_i in the covering $\mathcal{K}_v$ yields

$$\text{area}(S_v) \geq \text{area}(\partial R_1) \geq m\gamma \frac{V_1}{(\max_k \text{vol}(R_1 \cap K_k))^{1/3}}$$

and

$$(\max_k(\text{vol}(R_1 \cap K_k)))^{1/3} \geq m\gamma \frac{V_1}{\text{area}(S_v)} \geq \delta > 0,$$

because $V_1 = 1$ and it is easy to show that $\text{area}(S_v)$ is bounded (since there is a bounded way of enclosing the volumes). Translate each M_v so that the cube where the maximum occurs is centered

at the origin of $\mathbf{R}^N$, and rotate so that the tangent space of each M_v at the origin is equal to a fixed $\mathbf{R}^3$ in $\mathbf{R}^N$. The limit of the M_v will be equal to this $\mathbf{R}^3$. Since a cube with edge-length L centered at the origin fits inside a ball of radius $2L$ centered at the origin, we have

$$\text{vol}(R_1 \cap B(0, 2L)) \geq \delta^3$$

for every S_v. By the compactness theorem for locally integral currents ([34, pp. 64,88], [48, Section 27.3, 31.2, 31.3]) we know that a subsequence of the S_v has a limit, which we will call D, with the property that $\text{vol}(R_1) \geq \delta^3$. This completes the first step.

Since D is contained in the limit of the M_v, namely, the copy of $\mathbf{R}^3$ chosen above, and each S_v is minimizing for its volumes, a standard argument shows that the limit D is the perimeter-minimizing way to enclose and separate the given volumes in $\mathbf{R}^3$ (cf. [34, 13.7]). In the limit, V_2 could be zero, in which case D is a round sphere. If both volumes are non-zero, D is the standard double bubble ([23] and [41], or see [34, Chapter 14]).

Our goal is now to prove that the double bubble of volumes v, λv lies inside a trivial ball, when v is small enough. This can be accomplished using the monotonicity theorem for mass ratio [1, Section 5.1(1)], which implies that for a perimeter-minimizing bubble cluster, a small ball around any point on the surface contains some substantial amount of area. This will limit the number of disjoint balls we can place on the surface. The monotonicity theorem applies only to surfaces for which the mean curvature is bounded, *i.e.*, for which there is a C such that for smooth variations

$$\frac{dA}{dV} \leq C.$$

Accordingly, the second major step in the proof is to obtain such bound on the curvature, for v small. It suffices to show that all smooth variation vector fields have the property that changes in the volume of S_v and in the area of ∂S_v are controlled. Take a smooth variation vector field F in $\mathbf{R}^n$ such that for D,

$$dV_1/dt = \int_{\partial R1} (F \cdot n)\, dA = c > 0$$

and

$$dV_2/dt = 0.$$

(Note that we need the first volume to be non-zero, or its variation could be zero.) For v small enough the subsequence of S_v headed towards the limit D has the property that dV_1/dt is approximately c and dV_2/dt is approximately 0.

By the argument in the first step, we can translate each S_v similarly, so that a subsequence of the subsequence above converges to a minimizer D' in $\mathbf{R}^3$ where the second volume is non-trivial. This time we take a smooth variation vector field F', such that for v small enough the subsequence of S_v headed to D' has the property that dV_1/dt is approximately 0 and dV_2/dt is approximately $c' > 0$. This proves that for this subsequence the change in volume is bounded below.

Now we need to show that the change in area is bounded above. This follows from the fact that every rectifiable set can be thought of as a varifold [34, Section 11.2]. By compactness for varifolds [1, Section 6], the S_v, situated so that the first volume does not disappear, converge as varifolds to some varifold, J. The first variation of the varifolds also converge, *i.e.*, $\delta S_v \to \delta J$, see [1]. The first variation of a varifold is a function representing the change in area. Therefore, far enough out in the sequence the change in area of the S_v under F is bounded close to the change in area of J under F, which is finite. Similarly, the change in area of the S_v under F' is bounded.

We conclude that the mean curvature of S_v is bounded for two independent directions in the two-dimensional space of volume variations, and hence for all variations. This completes the second step.

The third and final step is to show that all of the surface area is contained in some ball in M_v, of fixed radius for all v. Eventually, as v shrinks and M_v grows, this ball will have to be trivial in M_v. We will then use the result that the optimal double bubble in $\mathbf{R}^3$ is standard to show that our double bubble is standard as well.

As the S_v are approaching the minimizer in $\mathbf{R}^3$, there must be a bound A on the perimeters of

the S_v. By monotonicity of mass ratio [1, Section 5.1(1)], every unit ball centered at a point of S_v contains perimeter $\delta > 0$. Therefore there are at most A/δ such disjoint balls.

We claim that the S_v are eventually connected. There is an upper bound on the diameter of any component, $2A/\delta$. Since we are controlling curvature, our components cannot become too small. Since every unit ball contains at least δ area, we also have a lower bound on the area of each component, when unit balls are trivial. Unless eventually the S_v are connected, you can arrange to get in the limit a disconnected minimizer in $\mathbf{R}^3$, a contradiction.

Hence S_v is contained in a ball of radius $2A/\delta$ for all v. Since our original manifold has compact quotient by its isometry group, there is a radius such that balls in the original manifold of that radius or smaller are topologically trivial. Hence, as we expand the manifold, eventually balls of radius $2A/\delta$ can be lifted to $\mathbf{R}^3$, which means that they are Euclidean. Hence, S_v is eventually contained in a Euclidean ball, and is therefore the standard double bubble ([41], [23], [34, Chapter 14]).

Finally, since $S_v \subset M_v$ is simply a scaled version of the original double bubble in M, we conclude that the original double bubble is standard as desired. QEF

Remark 7.2. Given n, m, similar arguments show that for any smooth n-dimensional Riemannian manifold with compact quotient by the isometry group, given $0 \leq \lambda \leq 1$, there are C, $\epsilon > 0$, such that for any $0 < v < \epsilon$, a minimizing cluster with m prescribed volumes between λv and v lies inside a ball of diameter at most $Cv^{1/n}$.

To further deduce that the cluster smoothly approximates a Euclidean minimizer would require knowing that convergence weakly and in measure, under bounded mean curvature, implies C^1 convergence, as is known for hypersurfaces without singularities ([1, Section 8], see [35, Section 1.2])

7.5 Special Tori

Changing the shape of the torus, by stretching it or by skewing some or all of its angles, would certainly change the phase diagram of Figure 7.2:

Conjecture 7.8. In the special case of a very long T^3 the Double Slab is optimal for most volumes.

Special tori may have special minimizers:

Conjecture 7.9. For the special case of a torus based on a relatively short $\frac{\pi}{3}$-rhombic right prism, the Hexagonal Honeycomb prism of Figure 7.4 is a perimeter-minimizing double bubble for which both regions and the exterior each have one third the volume, or when two volumes are equal and the third is close.

Indeed, for such volumes the Hexagonal Honeycomb ties the Double Slab, just as in the $\frac{\pi}{3}$-rhombic two-torus a Hexagonal Tiling ties the Double Band [9].

In the *triple bubble problem* for the Face Centered Cubic (FCC) and Body Centered Cubic (BCC) tori, if not for the standard cubic torus, we would expect to find minimizers that lift to $\mathbf{R}^3$ as periodic foams with cells of finite volume. For example, the Weaire-Phelan foam would be expected to appear as lifts of some solutions to the *seven* bubble problem in the BCC torus. The Weaire-Phelan foam, a counterexample to Kelvin's conjectured best way to divide space into unit volumes ([56], [27]; see [34, Chapter 15], [6]). Kelvin's foam might appear as a minimizer for the seven bubble problem in the FCC torus. It is interesting to note that the Weaire-Phelan foam has essentially the same shape as chlorine hydrate crystale.

In contrast with the triple bubble problem, it would be extremely unlikely if there was a solution to the double bubble problem in any torus that lifted to a division of $\mathbf{R}^3$ into finite volumes, since by regularity, singular curves meet in fours. This means that singular points look locally like the cone over a tetrahedral frame, and hence have four volumes coming together. Since a region will never be adjacent to another component of the same region (because the dividing wall could be removed

to decrease area and maintain volumes), a foam generated using a fundamental domain coming from a double double in the three-torus would have to exhibit the strange property that the singular curves never meet.

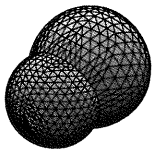
Conjecture 7.10. There are no least-area divisions of $\mathbf{T}^3$ into three volumes that lift to a foam in $\mathbf{R}^3$.

This conjecture suggests in addition that it is not likely that there will be are any other special minimizers.

Conjecture 7.11. The double bubbles of Figure 7.1 together with the Hexagonal Honeycomb of Figure 7.4 comprise the complete set of area minimizing double bubbles for all three tori.

As a final remark, in light of the fact that the adding more bubbles seems likely to produce many interesting candidates, we would like to mention one final conjecture.

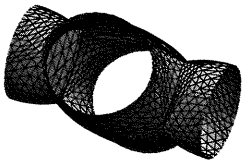
Conjecture 7.12. For the triple bubble problem in a cubic $\mathbf{T}^3$ in the case where one of the volumes is small, the minimizers will look like the double bubbles of Figure 7.1 with a small ball attached. The phase diagram will look just like our Figure 7.2.



STANDARD DOUBLE BUBBLE

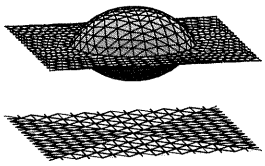
DÉLAUNEY CHAIN

CYLINDER LENS

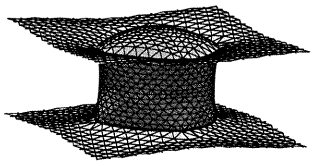


CYLINDER CROSS

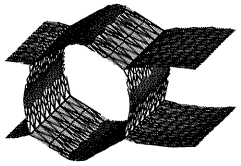
DOUBLE CYLINDER



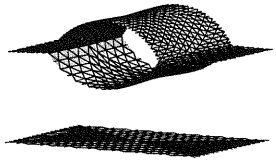
SLAB LENS



CENTER BUBBLE



CYLINDER STRING



SLAB CYLINDER

DOUBLE SLAB

Figure 7.1: Catalog of Conjectured Area-Minimizing Double Bubbles in the Flat Cubic 3-Torus.

SDB = Standard Double Bubble
DC = Delauney Chain
CL = Cylinder Lens
CC = Cylinder Cross
2C = Double Cylinder
SL = Slab Lens
CB = Center Bubble
CS = Cylinder String
SC = Slab Cylinder
2S = Double Slab

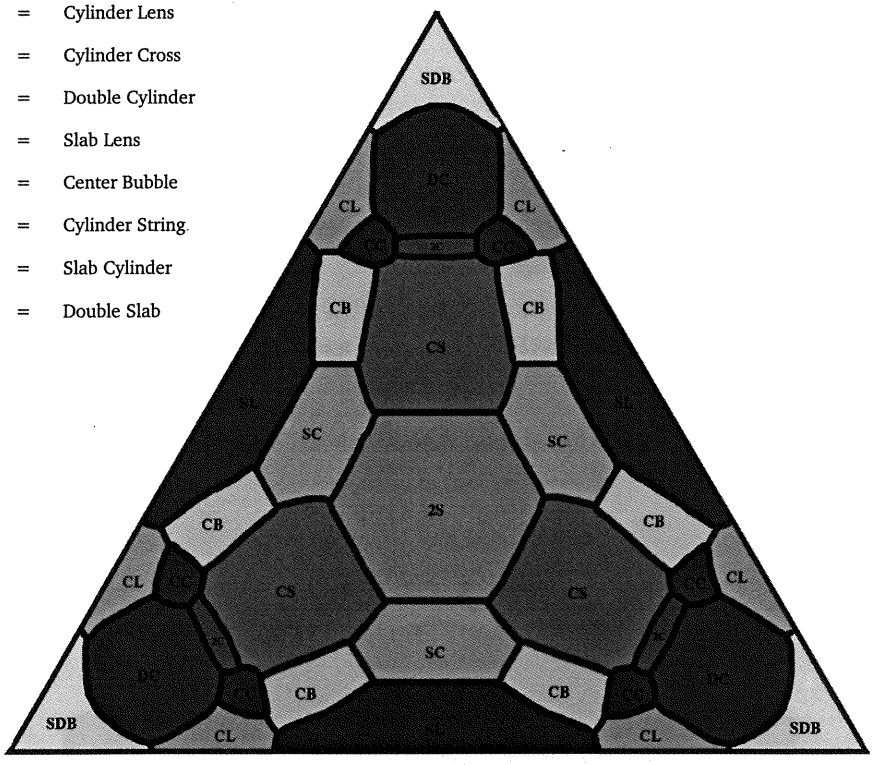
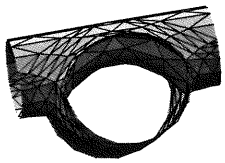
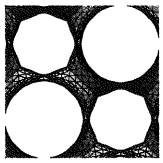


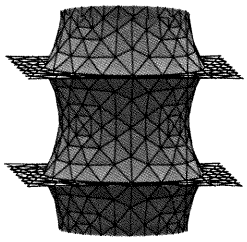
Figure 7.2: Phase portrait: volumes and corresponding double bubble. In the center both regions and the complement have one third of the total volume; along the edges one volume is small; in the corners two volumes are small.



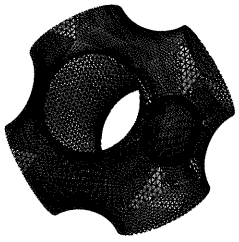
TRANSVERSE CYLINDERS



DOUBLE HYDRANT



CENTER CYLINDER



HYDRANT LENS

Figure 7.3: Inefficient double bubbles

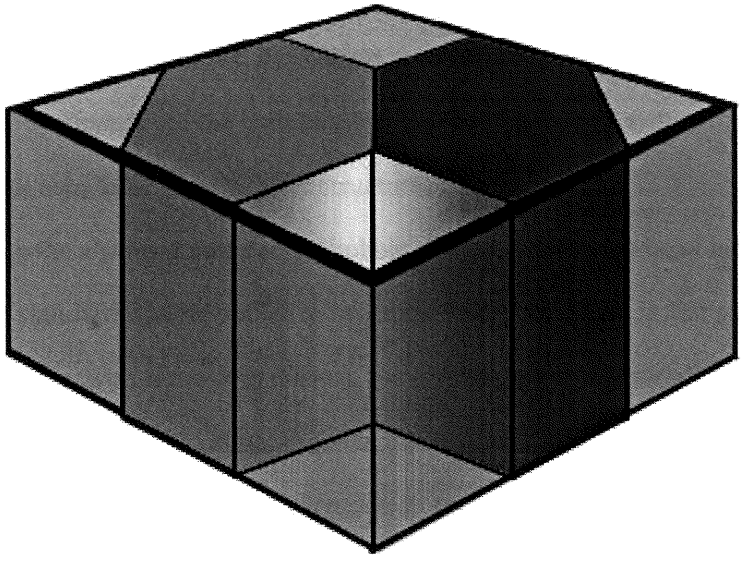


Figure 7.4: Hexagonal Honeycomb

conjectural epilog

These are some of the thoughts I have had while working on this thesis.

Any resembalence of the things discussed here to the current views of my collaborators or of the mathematical community at large or to the actual facts of the matter has to be regarded as coincidental at best.

Symmetry and algebraic structure Comparing the results from the various chapters leads to the conjecture that area-minimizing double bubbles in constant curvature spaces always have some of the same symmetries as the space, and more generally maximize symmetry in some sense.

We now consider the possibility that by seeking to maximize symmetry, bubble clusters not only take on some of the symmetries of the ambient, but also exhibit certain symmetries (*i.e.* invariants) associated with their own 'intrinsic' algebraic structure.

If this is the case, the algebraic structure of bubble clusters is bound to be of interest. Recall from Chapter 4 that the standard double bubble is invariant under inversion in the separating cap. For standard bubble clusters with more components, we can again move between adjacent bubbles by inversions – although doing so sometimes requires two moves, since if B_3 curves into B_2, we might send a point $p \in B_1$ to a point $q \in B_3$ by inversion in the wall separating B_1 and B_2; all such 'q's are however mapped into B_2 by inversion in the wall separating B_2 and B_3. When there are more than two bubbles, points fixed by the entire group of these (generalized) adjacency transformations

(or, the *cluster group*) are always in the singular set. When there are more than four bubbles, there may be no points that are fixed by the entire cluster group, but all singular curves are fixed by some subgroup of order three, and all singular points are fixed by some subgroup of order four.

The intrinsic algebraic structure of foams may may be particularly interesting (or, possibly, but not likely, particularly easy to understand) since periodic foams have a natural group acting on them already.

Maximizing symmetry here should for the time being be understood in a weak sense, say, maximizing symmetry while taking into consideration certain other factors. The cluster group may be hard to define or even meaningless in clusters that have non-contractible components (although I suspect that that won't be much of an obstacle). When the boundary surfaces are constant-mean-curvature sufaces other than pieces of spheres, things certainly *sound* less attractive, but the important ideas still seem to carry through.

At the bottom of this line of inquiry lies the fundamental assumption that there are no factors besides symmetry that can come into play to reduce area.

As one specific example that might give evidence as to the truth or falsity of this assumption, here is a problem that I think is open: the Kelvin foam *seems* to have 'more symmetry' than the Weaire-Phelan foam that takes less area. However, it might be that the Weaire-Phelan foam actually has more symmetries when one looks at its cluster group. Decide this question.

Is the m-bubble problem *inductive*? The standard double bubble is clearly a relaxation of the solution to the single bubble problem in space minus a round ball. In addition, it seems that all of the other conjectured minimizers for the double bubble problem in $\mathbf{T}^k$ $(k = 1, 2, 3)$ are relaxations of the solution to the single bubble problem in the torus minus a single bubble solution. This seems to be plainly obvious, albeit ever-so-hard to prove, for everything but the hexagonal tiling and hexagonal honeycomb. And upon further reflection it does not seem at all unlikely that the hexagonal tiling (resp. hexagonal honeycomb) is a relaxation of the solution to the single bubble problem in a 'nut',

i.e. a hexagonal torus minus a round disk (resp. tubular neighborhood of a short geodesic).

These considerations lead to the following *meta-Conjecture*.

Meta-Conjecture.

1. The solution to the m-bubble problem in any space of constant curvature C is bounded by regularity satisfying relaxations of some collection of (a) tubular neighborhoods of totally geodesic submanifolds (counting points as a degenerate case) and (b) tubular neighborhoods of totally geodesic submanifolds (not allowing points this time) regarded as manifolds with boundary that then have some noncontractible (with respect to their boundary) holes drilled out of them; moreover,

2. the boundary of the union of the closure of the holes in each tubular neighborhood T_0 in (1b) bounds a totally geodesic submanifold of T_0 (again not allowing points); and, ultimately,

3. the problem is inductive in the sense that the m-bubble problem is given by relaxing the solution to the 1-bubble problem for the current with free boundary given by designating some solution to the $(m-1)$-bubble problem in C as boundary.

appendices

A Algebraic preliminaries

Definition A.1. A *vector space* is a quadruple $(X, F, +, *)$, where X is a set, F is a field, and $+$ and $*$ are operations such that

i. $\forall x \in X, f \in F : x * f \in X$,

ii. $\forall x, y \in X : x + y \in X$.

Furthermore, addition $(+)$ is commutative, there exists an additive identity (written 0) and there are additive inverses. Scalar multiplication $(*)$ is associative and distributes over addition both ways.

We write $V = (X, F, +, *)$, and when we say $v \in V$ we mean $v \in X$. The field F is called the *base field*. In this thesis, the base field is usually the real or the complex numbers.

The notion of a *basis* and of how to do computations with a basis (using *matrices*) are assumed to be familiar to the reader, at least for finite dimensions.

Definition A.2. A *bilinear form* on a finite dimensional vector space V is a map b from $V \times V$ into the base field that is linear in each factor. If we chose a basis for the vector space, then the bilinear form is given by matrix multiplication as follows: $b : (x, y) \mapsto y^T B x$, for some matrix B.

Remark A.1. The two dimensional case is

$$((x_1, x_2), (y_1, y_2)) \mapsto \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = ax_1y_1 + bx_1y_2 + cx_2y_1 + dx_2y_2.$$

Definition A.3. An *inner product* on a vector space is a bilinear form that is also *positive definite* (which means that $b(x, x) \geq 0$ and is equal to zero only when $x = 0$) and *symmetric* ($b(x, y) = b(y, x)$).

Remark A.2. Given an abstractly defined linear map on a finite dimensional vector space, to determine the entries of the matrix representing this map it suffices to compute the action of the map on a basis. Similarly for bilinear forms and thus for inner products. The matrix representing a bilinear form b with respect to the basis $\{x_i\}_{i=1}^n$ is $(b(x_i, x_j))$.

Definition A.4. A *vector space isomorphism* is a linear, bijective map between two vector spaces.

Definition A.5. The *dual* of a vector space V is the set of linear maps from V into the base field. We denote the dual of V by the symbol V^*.

Remark A.3. It is easy to check that vector space duals are themselves also vector spaces. It is also easy to check that in the finite dimensional case a vector space is isomorphic to its dual. This is not generally true for infinite dimensional vector spaces.

Definition A.6. A *group* is a set X together with a binary operation $*$ with the properties

i. $\forall x, y, z \in X : x * (y * z) = (x * y) * z$,

ii. $\exists e \in X : \forall x \in X : e * x = x * e = x$,

iii. $\forall x \in X : \exists x^{-1} \in X : x * x^{-1} = x^{-1} * x = e$,

We write $G = (X, *)$ and when we say $g \in G$ we mean $g \in X$. It is easy to deduce that a vector space is an *additive group* (that is, the operation $(+)$ in Definition A.1 plays the role of $(*)$ in Definition A.6). We now define another important example:

Definition A.7. The *symmetric group on n elements* $\mathfrak{S}^n$ is equal to the set of permutations of n elements together with the operation of composition.

Remark A.4. So, for example, $\mathfrak{S}^3$ can be written down as $\{(), (12), (23), (13), (123), (321)\}$ (do nothing, sent 1 to 2 and 2 to 1, send 2 to 3 and 3 to 2, etc.) and then it is easy to check that with the operation $\circ$, $\mathfrak{S}^3$ really is a group (for example, $(123)\circ(321) = ()$, so $(321) = (123)^{-1}$). The symmetric group is important in the theory of determinants (and many other places). Indeed, using the *sign* of permutations (even if the an even number of elements are permuted, odd if an odd number of elements are permuted) it is not too difficult to show (or at least beleive) that for an $n \times n$ matrix A,

$$det(A) = \sum_{\sigma \in \mathfrak{S}^n} \text{sign}(\sigma) A_{\sigma 1,1} \cdot \ldots \cdot A\sigma n, n);$$

see [20, Chapter 5] for the details.

Definition A.8. If a group G acts on a set X (in the sense that $\mathfrak{S}^n$ acts on the set $\{1, \ldots, n\}$) the *orbit* of a point $x \in X$ is defined to be the set of points in X that are hit by applying elements of G to x (with repeated application of the same element allowed).

B Topological preliminaries

Definition B.1. A *topological space* is a pair (X, τ) such that τ is comprised of subsets of X and has the further additional properties

i. $\emptyset \in \tau$ and $X \in \tau$.

ii. finite intersections of elements of τ are again in τ

iii. arbitrary unions of elements of τ are again in τ.

A collection of subsets τ satisfying these conditions make τ into a *topology*. The elements of τ are then called the *open sets of X in the τ topology*.

Remark B.1. If $T = (X, \tau)$ is a topological space, when we say $p \in T$ we mean $p \in X$.

Definition B.2. A map between topological spaces is *continuous* if the inverse image of every open set is an open set.

Definition B.3. A map between topological spaces is a *homeomorphism* if it is a continuous bijection such that in addition the image of every open set is an open set. (So, it is a bijection of points and it induces a bijection on open sets.)

Since they are a bijection of points and sets, we conclude that homeomorphisms respect topological structure. In this sense it would be appropriate to consider them to be 'topological isomorphisms'.

We will need the following two 'special' definitions.

Definition B.4. A topological space T is *Hausdorff* if it has the property that any pair of distinct points $p, q \in T$ belong to disjoint open sets.

(Spaces without this property include easy examples like $X = \{a, b, c\}, \tau = \{\emptyset, \{a\}, \{a, b, c\}\}$ and more complicated examples, like $(\mathbf{R} \cup (0, 1) \subset \mathbf{R}^2, \{U; U \cap \mathbf{R}$ such that U is open in the metric topology on $\mathbf{R}$ and if $p \in U$ then $(U \cap \mathbf{R}) \cup 0$ is a neighborhood of 0 in $\mathbf{R}\})$. In general we will exclude non-Hausdorff spaces like this because they cause trouble.)

Definition B.5. A topological space $T = (X, \tau)$ is *paracompact* if there is a countable set $\mathcal{B}$ of open sets in τ with the property that every open set in τ is the union of elements of $\mathcal{B}$.

Remark B.2. In general, a set $\mathcal{B}$ that contains the set of open sets is called a basis for the topology. Every topology has a basis (any topology contains all of its open sets). We note that the basis *condition* is equivalent to the condition that any point in an open set is also in an element of $\mathcal{B}$. The *word* paracompact is synonymous with the phrase *second countable*.

This begs the question, what was *first countable*? This condition is not needed elsewhere in this thesis, but so that the reader leaves with a satisfied mind, here it is anyway: each point in X must

have the property that there is a countable set $\mathcal{B}_p$ of open sets containing p such that every open set containing p also contains an element of $\mathcal{B}_p$. The set $\mathcal{B}_p$ is then called a *local basis*.

Definition B.6. A *metric space* is a pair (X, d) such that d is a function defined on $X \times X$ with the following additional properties:

i. $0 \leq d(x, y) < \infty$

ii. $d(x, y) = 0 \Leftrightarrow x = y$

iii. $d(x, y) = d(y, x)$

iv. $d(x, y) \leq d(x, z) + d(z, y)$

Definition B.7. The set of *metric balls* $\{\mathbf{D}_x(r) = \{y \in X; d(y, x) < r\}; x \in X, 0 \leq r \in \mathbf{R}\}$ form a topology for X. This is called the *metric topology*.

Remark B.3. We remark that $\mathbf{R}^n$ with the metric topology (the *standard topology on* $\mathbf{R}^n$) is both paracompact and Hausdorff. Metric balls with a rational radius centered at rational points form a countable basis for the topology, and by chosing ρ small enough, any two points p and q can be enclosed in disjoint metric balls of radius ρ.

The dividing line between topology and geometry moves around a lot. It shows up right here this time.

C Geometric Preliminaries

For our purposes, Riemannian geometry is the study of

I manifolds, with

II a differentiable structure,

III admitting the definition of vector fields,

IV which can themselves be differentiated;

V with also a notion of lengths and angles associated with vectors based at a point in the manifold,

VI leading to the notion of shortest paths in the manifold,

VII and to many other results having to do with global structure.

Item VII includes answers to questions like 'What are the least surface area objects in my manifold?'.

Geometry of manifolds

Manifolds generalize the idea of parametrized surfaces. This is useful, because, for example, it is all fine and good to study the upper-hemisphere, considered as the graph of $\sqrt{r^2 - x_2 - y^2}$ over a disk of radius r, but if we can't talk coherently about the whole sphere, our mathematical lives will be considerably less interesting. Manifold theory lets us talk about the whole sphere and about lots of other things. One of the central ideas in the definition of a smooth manifold is that it stitches together 'local parameterizations' in a nice way.

Definition C.1. A *manifold* is a paracompact Hausdorff topological space such that about each point $p \in M$ there is some open set $U \ni p$ such that there exisits a homeomorphism φ from U to $\mathbf{R}^n$.

The pairs of open sets and homeomorphisms (U, φ) from Definition C.1 are called *charts*. If we say that a point p is in a chart (U, φ), what we mean is that $p \in U$. All of the charts taken together are called an *atlas*. Our atlases are henceforth assumed to be *complete* in the sense that they are never contained in some other atlas. This is a viable assumption, because we can keep adding in charts until we complete the atlas. We denote the complete atlas for our manifold by $\mathcal{A}$.

We write M for the triple $(X, \tau, \mathcal{A})$, and when we say $p \in M$ we mean $p \in X$. Because this convention is so prevalently and widely used, it is sometimes easy to lose sight of the important fact that the topology τ as well as the point-set X are fixed; in particular, you should always be ready to ask 'with what topology?' if someone describes a manifold to you in terms of its point set only and their choice of topology was not obvious.

The atlas makes it possible to compute things; we can use a chart (U, φ) to define a coordinate system on U which we then use to study small pieces of the manifold analytically. If we are interested in what is going on around a point $p \in U \subset M$, we can assume p is mapped by the associated homeomorphism φ to $0 \in \mathbf{R}^n$, since wherever it is mapped we can compose with a translation to recenter $\mathbf{R}^n$ on the image of p. We then make the following extremely useful definition.

Definition C.2. If $x_1, \ldots, x_n$ are the functions which give the coordinates of the point x in $\mathbf{R}^n$, we say that the set of functions $x_1 \circ \varphi, \ldots, x_n \circ \varphi$ are *local coordinates centered at p* or just *local coordinates at p*.

Remark C.1. For notational convenience, we will frequently make use of the alternate notation for local coordinates at p, namely $u_1 = x_1 \circ \varphi, \ldots u_n = x_n \circ \varphi$.

In order to compute anything very interesting, we need some more structure.

Definition C.3. A manifold's atlas is a *differentiable structure on the manifold* if the charts of Definition 1 satisfy the following property: if (U_1, φ_1) is a chart about p and (U_2, φ_2) is a chart about q then whenever $U_1 \cap U_2 \neq \emptyset$, the map $\varphi_1 \circ \varphi_2^{-1}$ from $\varphi_2(U_1 \cap U_2) \subset \mathbf{R}^n$ onto $\varphi_1(U_1 \cap U_2) \subset \mathbf{R}^n$ is differentiable.

We say a manifold has a C^k *structure* if the transition functions $\varphi_j \circ \varphi_i^{-1}$ are k times continuously differentiable. We say a manifold has a *smooth* structure if the transition functions are infinitely differentiable.

The existence of differentiable structures allows us to talk about differentiable and smooth maps between manifolds.

Definition C.4. If M_1 and M_2 are manifolds (say, of dimensions m and n respectively) then $f : M_1 \to M_2$ is said to be a *differentiable map* if for every chart (U, φ) of M_1 and every chart (V, ψ) of M_2 such that $f(U) \subset V$ the map $\psi \circ \varphi^{-1}$ from $\varphi(U) \subset \mathbf{R}^m$ into $\psi(V) \subset \mathbf{R}^n$ is C^k as a map between Euclidean spaces (see section 2.1.1).

If we take $M_1 = J$ to be an open interval in $\mathbf{R}$ then a *differentiable curve* in M_2 is any differentiable map γ of a closed interval in M_1 into M_2. For the rest of the appendix, curves will always be differentiable.

Definition C.5. Let $\gamma : [a, b] \to M$ be a smooth curve in M that passes through the point p at the time t_0. Let $\mathcal{F}$ be the set of differentiable functions from M to $\mathbf{R}$. The *tangent vector to γ at p* is a linear map $L_{\gamma,p}$ defined on $\mathcal{F}$ by the rule

$$L_{\gamma,p}(f) = \left.\frac{df(\gamma(t))}{dt}\right|_{t_0} \in \mathbf{R}.$$

Remark C.2. Tangent vectors give us a directional derivative on functions (we can use them to compute how fast a function is changing in a given direction). We can conveniently compute the tangent vector to a curve γ at p in terms of local coordinates at p. If $\gamma_i(t) = x_i \circ \varphi \circ \gamma(t)$ are the coordinates of the curve, and we know f as a function of local coordinates at p, then the chain rule tells us that

$$\begin{aligned}\left.\frac{df(\gamma(t))}{dt}\right|_{t_0} &= \sum_{i=1}^{n} \left.\frac{df}{d(x_i \circ \varphi \circ \gamma)}\right|_{t_0} \left.\frac{d\gamma_i(t)}{dt}\right|_{t_0} \\ &= \sum_{i=1}^{n} \left.\frac{df}{d(x_i \circ \varphi)}\right|_{p=\gamma(t_0)} \left.\frac{d\gamma_i(t)}{dt}\right|_{t_0}\end{aligned}$$

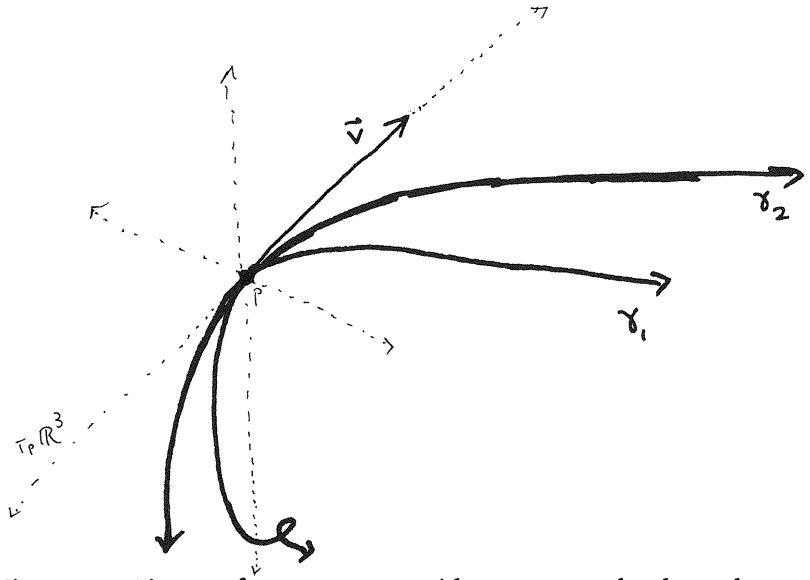


Figure 7.5: Picture of tangent space with two curves that have the same tangent vector.

$$= \sum_{i=1}^{n} \frac{df}{du_i}\bigg|_{p=\gamma(t_0)} \frac{d\gamma_i(t)}{dt}\bigg|_{t_0}.$$

Definition C.6. The *tangent space of M at p*, denoted T_pM, is equal to the set of all tangent vectors of curves in M through p.

Remark C.3. The computation of Remark C.2 shows that the tangent space at a point is an n-dimensional real vector space with basis $\{\frac{d}{du_i}\big|_p\}$. We frequently will identify this space with $\mathbf{R}^n$.

The functional definition of tangent vectors and tangent spaces given here is slightly less popular than the more geometric-seeming definition from subsection 2.1.1 in the text that identifies curves whose tangent vectors are the same to make the tangent space. The computation of Remark C.2 shows that these two definitions are equivalent, since if we regard tangent vectors as being given by a linear combination of '$\frac{d}{du_i}$'s, then to get a given tangent vector we must specify the right '$\frac{d\gamma_i(t)}{dt}$'s, that is, we must specify the derivatives found upon differentiating a curve in some equivalence class; whereas on the other hand, clearly if we specify on equivalence class of curves, we get a functional tangent vector (something that differentiates functions) using the given formula.

Definition C.7. The *tangent bundle TM* of a manifold is the union over x in M of all the the tangent spaces.

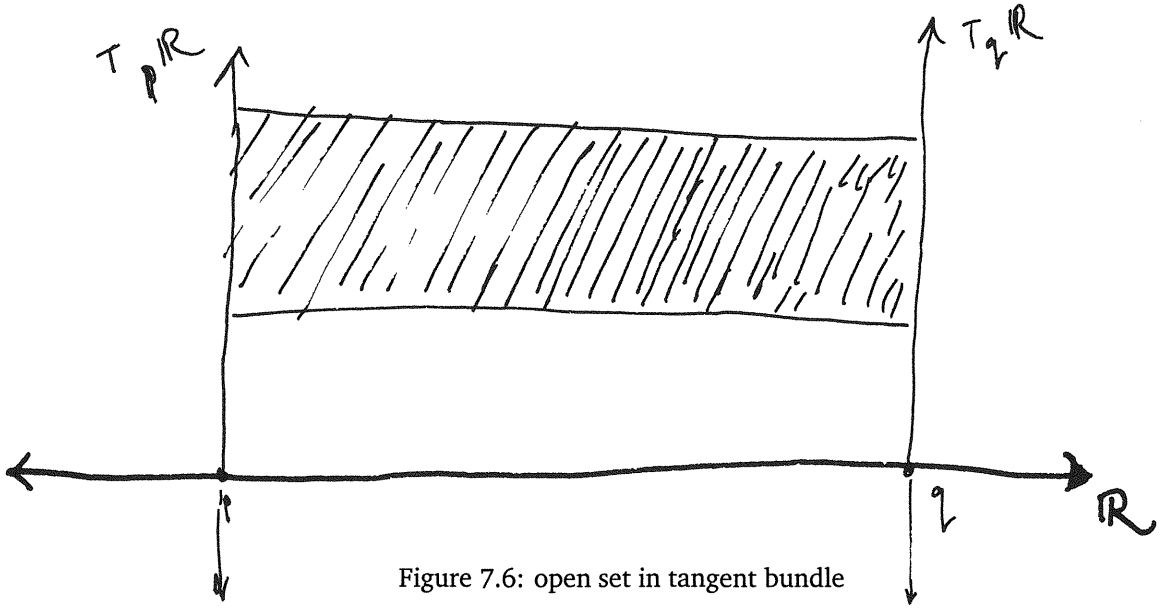


Figure 7.6: open set in tangent bundle

Definition C.8. We define the *standard projection operator* π from the tangent bundle to the manifold which takes each vector to its base point.

Remark C.4. Typically we want to think of the tangent bundle as a manifold with a differentiable structure that can tell us something about the base manifold, not just as the set of all vectors tangent to the manifold 'bundled together' indiscriminately. We distiguish the points in TM carefully by writing $TM = \{(p, v) : p \in M, v \in T_pM\}$; then a natural topology on TM is induced by the topology of M, namely, an open set in TM is given by $U \times V$ where U is an open set of M that comes from a chart and V is an open set of $\mathbf{R}^n$. We are using the identification of the tangent spaces and $\mathbf{R}^n$. The picture is easy in one dimension; see Figure 7.6. The tangent bundle also has a natural smooth structure on it, since if (U, φ) is a chart on M then every element $u \in \{(p, v) : p \in U, v \in T_pM\} \subset TM$ can be written in the form $u = \sum_{i=1}^{n} a_i(v) \left.\frac{d}{du_i}\right|_{p=\pi(v)}$. The map

$$T_pM \ni u \mapsto (u_1(\pi(v)), \ldots, u_n(\pi(v)), a_1(v), \ldots, a_n(v)) \in \mathbf{R}^{2n}$$

is a homeomorphism from $\{(p, v) : p \in U, v \in T_pM\}$ to $\mathbf{R}^{2n}$.

One checks that the conditions of Definition C.3 are satisfied ([49, volume I, page 3-32]).

Definition C.9. A *vector field* on a manifold consists of a choice of one vector from each tangent space. Hence vector fields are functions from a base manifold to its tangent bundle. We write $\text{vect}(M) = \{\text{vector fields on } M\}$. Another way to talk about a vector field to call it a *section* of the tangent bundle, where the word section means a map s from the manifold to the bundle such that $\pi \circ s$ is the identity. If $\mathcal{V}$ is a vector field on M, $\mathcal{V}|_x$ denotes the vector associated with $T_x M$.

Remark C.5. Since tangent vectors are differential operators, vector fields are too; they take differentiable functions defined on the manifold to functions on the manifold given by the directional derivatives specified by the tangent vector at each point. They also allow us to compute lots of other things, but as before, to compute anything really interesting, we want more structure.

Definition C.10. Refering to DefinitionC.3, we note that the smooth structure on TM developed in Remark C.4 makes it possible to define

$$\text{vect}^k(M) = \{C^k \text{ vector fields on } M\}$$

and

$$\text{vect}^\infty(M) = \{C^\infty \text{ vector fields on } M\}.$$

We will usually just write $\text{vect}(M)$ and understand a degree of smoothness appropriate for the needs at hand.

Remark C.6. The idea of the nabla coming up is that somehow the derivative of a smooth section of TM with respect to a smooth section of TM is again a smooth section of TM. We don't necessarily expect that to happen; we have to work just the right way. (See Kobaishi and Nomizu [39] for a much more general way to go about doing this.)

Definition C.11. A *connection* is a map $\nabla : TM \times \text{vect}(M) \to TM$ which satisfies the following properties:

i. if $\xi \in TM, \mathcal{V} \in \text{vect}(M)$, then $\nabla(\xi, \mathcal{V})$ is in the same tangent space as ξ.

ii. if $a, b \in \mathbf{R}^n$, $x \in M$, $\xi, \eta \in T_x M$, and $\mathcal{V} \in \text{vect}(M)$, then

$$\nabla(a\xi + b\eta, \mathcal{V}) = a\nabla(\xi, \mathcal{V}) + b\nabla(\eta, \mathcal{V}) \tag{C.1}$$

iii. if $x \in M$, $\xi \in T_x M$, $\mathcal{V}, \mathcal{W} \in \text{vect}(M)$, and $f \in C^1(M)$, then

$$\nabla(\xi, \mathcal{V} + \mathcal{W}) = \nabla(\xi, \mathcal{V}) + \nabla(\xi, \mathcal{W}) \tag{C.2}$$

$$\nabla(\xi, f\mathcal{V}) = (\xi f)(\pi(\xi))\mathcal{V}|_{\pi(\xi)} + f(\pi(\xi))\nabla(\xi, \mathcal{V}) \tag{C.3}$$

iv. ∇ is smooth in the sense that if $\mathcal{V}, \mathcal{W} \in \text{vect}^\infty(M)$ then $\cup_{x \in M} \nabla(\mathcal{V}|_x, \mathcal{W})$, which we call $\nabla(\mathcal{V}, \mathcal{W})$, is again in $\text{vect}^\infty(M)$.

Remark C.7. When it is clear from context where we are working, or when we want to be mobile, we use this more casual notation for the second part of item (iii):

$$\nabla(\xi, f\mathcal{V}) = (\xi f)\mathcal{V} + f\nabla(\xi, \mathcal{V}) \tag{C.4}$$

where it is understood that everything on the right-hand side that needs to be is to being evaluated at $\pi(\xi)$.

Definition C.11 is far from being the most general definition of a connection. In some ways it is much more like the definition of what is called a *covariant derivative associated with a connection*. We leave such matters aside; this definition will suffice for all our needs in this thesis. At any rate, it is clear from item (iii) above that whatever it is called, we now have a method for differentiating vector fields.

At this point we note that we must have

$$\nabla\left(\frac{d}{du_k}, \frac{d}{du_j}\right) = \sum_{i=1}^{n} \Gamma_{jk}^i \frac{d}{du_i} \tag{C.5}$$

for some set of 'Γ_{jk}^i's. (The order and placement of the various indices here is traditional.) If we 'know' a connection somehow, we know its action on all of $TM \times \text{vect}(M)$ and we can compute

these; alternatively, and more realistically, we can just specify them to begin with and get some connection as a result. In particular, this shows that connections exist.

In order to actually compute $\nabla(\xi, \mathcal{V})$ all we need to know is $\mathcal{V}$ restricted to some smooth curve γ through $p = \pi(\xi)$ with tangent vector ξ at p. In terms of the basis for the tangent space at p that we chose in Remark C.1, $\xi = \sum_{i=1}^{n} \xi_i \frac{d}{du_i}$.

By the linearity in the first component (condition (ii) above) this means that

$$\nabla(\xi, \mathcal{V}) = \sum_{i=1}^{n} \xi_i \nabla(\frac{d}{du_i}, \mathcal{V}).$$

Since we know $\mathcal{V}$ along γ, for each $q \in \gamma$ we can express $\mathcal{V}$ at each point in terms of local coordinates at q as $\mathcal{V}|_q = \sum_{i=1}^{n} \eta_i(q) \frac{d}{du_i}|_q$, or, more informally, $\mathcal{V} = \sum_{i=1}^{n} \eta_i \frac{d}{du_i}$. Combining this expression with the previous, we have

$$\nabla(\xi, \mathcal{V}) = \sum_{i=1}^{n} \xi_i \nabla \left(\frac{d}{du_i}, \sum_{j=1}^{n} \eta_j \frac{d}{du_j} \right).$$

We now enter a long chain of manipulations that will in the end give us a nice formula for $\nabla(\xi, \mathcal{V})$. First applying the first part of condition (iii) above

$$\sum_{i=1}^{n} \xi_i \nabla \left(\frac{d}{du_i}, \sum_{j=1}^{n} \eta_j \frac{d}{du_j} \right) = \sum_{i=1}^{n} \xi_i \sum_{j=1}^{n} \nabla \left(\frac{d}{du_i}, \eta_j \frac{d}{du_j} \right)$$

(and then the second part of condition (iii))

$$= \sum_{i=1}^{n} \xi_i \sum_{j=1}^{n} \left(\left(\frac{d\eta_i}{du_i} \right) \frac{d}{du_j} + \eta_j \nabla \left(\frac{d}{du_i}, \frac{d}{du_j} \right) \right)$$

(and the definition of the 'Γ_{jk}^{i}'s)

$$= \sum_{i=1}^{n} \xi_i \left(\sum_{j=1}^{n} \left(\left(\frac{d\eta_j}{du_i} \right) \frac{d}{du_j} + \eta_j \sum_{k=1}^{n} \left(\Gamma_{ji}^{k} \frac{d}{du_k} \right) \right) \right)$$

(collecting like terms)

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left(\left(\xi_i \frac{d\eta_j}{du_i} \right) \frac{d}{du_j} + \sum_{k=1}^{n} \left((\xi_i \eta_j \Gamma_{ji}^{k}) \frac{d}{du_k} \right) \right) \right)$$

(rewriting the summation without quite so many parentheses)

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\left(\xi_i \frac{d\eta_j}{du_i} \right) \frac{d}{du_j} \right) + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \left((\xi_i \eta_j \Gamma_{ji}^{k}) \frac{d}{du_k} \right)$$

(renaming one of the indices of summation does not effect the sum)

$$= \sum_{i=1}^{n}\sum_{k=1}^{n}\left(\left(\xi_i \frac{d\eta_k}{du_i}\right)\frac{d}{du_k}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left((\xi_i\eta_j\Gamma_{ji}^k)\frac{d}{du_k}\right)$$

(and now rearranging the summation)

$$= \sum_{k=1}^{n}\left(\sum_{i=1}^{n}\left(\left(\xi_i \frac{d\eta_k}{du_i}\right)\frac{d}{du_k}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\left((\xi_i\eta_j\Gamma_{ji}^k)\frac{d}{du_k}\right)\right)$$

(and finally, factoring out the elements of the basis)

$$= \sum_{k=1}^{n}\left(\sum_{i=1}^{n}\left(\left(\xi_i \frac{d\eta_k}{du_i}\right)\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}(\xi_i\eta_j\Gamma_{ji}^k)\right)\frac{d}{du_k}$$

we get the promised formula for $\nabla(\xi, \mathcal{V})$. In terms of a curve γ that goes through the point p at time t_0 with the tangent vector ξ, this is

$$\sum_{k=1}^{n}\left(\sum_{i=1}^{n}\left(\frac{d(\eta_k \circ \gamma)}{dt}\bigg|_{t_0}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}(\xi_i\eta_j(p)\Gamma_{ji}^k(p))\right)\frac{d}{du_k}\bigg|_p . \tag{C.6}$$

So we have a word, 'connection', and we know that it is not a meaningless one, because all we need are some n^3 functions, the 'Γ_{ij}^k's to specify one. (The official, undescriptive and even misleading name for the 'Γ_{ij}^k's is the *Christoffel symbols* – I prefer to use the more descriptive but less popular name *connection coefficients*, but even this does not sufficiently stress the fact that these are functions that can change as we move around the manifold.) We also know a connection is not an *a priori* useless thing, because if we have a curve with the right tangent vector, we use it and the connection to compute something.

But the real reason we care about Definition C.11 will not be seen for a few pages, that is, until we work throught the most important example. Along the way we will introduce some further terminology that will take us closer to our goal, but understanding Definition C.11 will be our main objective for the moment. The main idea is that connections will enable us bring vectors from one tangent space to another tangent space to compare them. In this way a connection is sort of like a set of pnuematic tubes that run between every point on the manifold, through which we can send and receive tangent vectors. As one might guess things can happen to the tangent vectors while in

transit. In particular, one can send a tangent vector around a closed loop, and it will not necessarily come back the same as it was when it started. This is not a defect of the system, but one of its strengths.

For such a complicated structure to be useful, we first need a fruitful way to compare vectors in a fixed tangent space. There are a number of other things we could define without this added structure, but I suspect that they would all be better motivated if we make everything concrete first, rather than afterwards. In light of this change of mood, we will close the current subsection and begin a new one now.

Riemannian manifolds

Definition C.12. A *Riemannian manifold* is a pair, a smooth manifold M, together with a map g defined on the manifold that associates to each point an inner product, that is,

$$g(p) = \langle \cdot, \cdot \rangle_p : T_pM \times T_pM \to \mathbf{R};$$

g must satisfy the additional property that for any open set $U \subset M$ and any fixed $\mathcal{V}, \mathcal{W} \in \text{vect}(U)$ the function

$$\langle \mathcal{V}, \mathcal{W} \rangle : U \ni p \mapsto g(p)(\mathcal{V}_p, \mathcal{W}_p) = \langle \mathcal{V}_p, \mathcal{W}_p \rangle_p \in \mathbf{R}$$

is differentiable on U. Any g satisfing all of these conditions is called a *smooth Riemannian metric.*

Remark C.8. The idea of a Riemannian metric is (as the name suggests) the heart of Riemannian geometry. This is what lets us compare vectors at a point: in accordance with our experience from Euclidean space, which as we have remarked the tangent space *is*, the *length* of a vector ξ based at the point p is

$$|\xi|_p = \sqrt{\langle \xi, \xi \rangle_p}. \tag{C.7}$$

The *angle* between two vectors ξ and η based at the point p is

$$\cos^{-1}\left(\frac{\langle \xi, \eta \rangle_p}{|\xi|_p |\eta|_p}\right). \tag{C.8}$$

The difference from what we might be used to is that the metric can change as we move from place to place. This is what makes comparing vectors from different tangent spaces such a subtle problem.

The Riemannian metric makes available to us for the first time some concrete geometric content. In addition to making local comparisons, we can begin to move around in the manifold and talk, for example, about the *length of a curve* $\gamma : [a, b] \to M$ in analogy with our experience in Euclidean space (see Chapter 2) as being

$$\int_a^b \left| \frac{d}{dt}\gamma(t) \right| \, dt. \tag{C.9}$$

If we define the *distance* between two points in the manifold by $d(p, q) = \inf_\gamma \text{length}(\gamma)$ then we get a distance metric on the manifold, and the manifold therefore can be viewed as a metric space. (Note that Riemannian metrics and distance metrics are very, very different things!)

As we made explicit in the section on Algebraic Preliminaries, each inner product on a finite dimensional vector space is determined by a matrix. We will write the matrix associated with the inner product at the point p as $\left(g_{ij}(p)\right)$ where the '$g_{ij}(p)$'s are determined by the inner products of basis vectors as in remark A.2. We write (g_{ij}) when we want to be mobile.

The inverse matrix often comes up in computations. It is traditional to write $(g_{ij})^{-1} = (g^{ij})$. Since these matrices are inverses, the sum $\sum_{i=1}^{n} g^{ij} g_{jk} = \delta_{ik}$.

We are getting very close to an explicit statement of the 'correct' way to compare tangent vectors from different tangent spaces; in fact, there is just one more definition we need before we can state our big theorem.

Definition C.13. For $U \subset M$ an open subset, if $\mathcal{V}, \mathcal{W} \in \text{vect}(U)$ we define their *Lie Bracket* or *commutator* $[\mathcal{V}, \mathcal{W}]$ to be the unique vector field such that

$$\forall p \in U : [\mathcal{V}, \mathcal{W}]|_p = \mathcal{V}|_p \circ \mathcal{W}|_p - \mathcal{W}|_p \circ V|_p.$$

Remark C.9. Note that in this and in the previous definition, the reason we talk about an open set U is that we will want to use local vector fields a lot of the time.

Theorem C.1 (Levi-Civita). If M is a Riemannian manifold then there exists a unique connection ∇^{LC} for which all $\mathcal{U}, \mathcal{V}, \mathcal{W} \in \mathrm{vect}(M)$ satisfy

i. $\nabla^{\mathrm{LC}}(\mathcal{V}, \mathcal{W}) = \nabla^{\mathrm{LC}}(\mathcal{W}, \mathcal{V}) + [\mathcal{V}, \mathcal{W}]$

ii. $\mathcal{U}\langle \mathcal{V}, \mathcal{W}\rangle = \langle \nabla^{\mathrm{LC}}(\mathcal{U}, \mathcal{V}), \mathcal{W}\rangle + \langle \mathcal{V}, \nabla^{\mathrm{LC}}(\mathcal{U}, \mathcal{W})\rangle$.

Remark C.10. 'LC' stands of course for 'Levi-Civita'; ∇^{LC} is called the *Levi-Civita connection for the Riemannian metric on* M. The two conditions say respectively that the Levi-Civita connection is *torsion-free* (in the sense that commutation is respected; this implies that the 'Γ^k_{ij}'s depend on the metric in a particularly simple fashion, and are, as we will see, symmetric under exchange of i and j) and that it is *compatible with the metric* in the sense that it behaves like a standard derivative of the metric (for example, if v and w are tangent vectors to a curve, $\frac{d\langle v,w\rangle}{dt} = \left\langle \frac{dv}{dt}, w\right\rangle + \left\langle v, \frac{dw}{dt}\right\rangle$).

Proof of Theorem C.1. Assuming such a connection ∇^{LC} exists, we compute like mad (see Table C.1). Upon collecting like terms we have that

$$\begin{aligned}\langle \nabla^{\mathrm{LC}}(\mathcal{U}, \mathcal{V}), \mathcal{W}\rangle = \frac{1}{2}\big(&\mathcal{U}\langle \mathcal{V}, \mathcal{W}\rangle + \mathcal{V}\langle \mathcal{W}, \mathcal{U}\rangle - \mathcal{W}\langle \mathcal{U}, \mathcal{V}\rangle \\ &- \langle \mathcal{W}, [\mathcal{V}, \mathcal{U}]\rangle + \langle \mathcal{U}, [\mathcal{W}, \mathcal{V}]\rangle - \langle \mathcal{V}, [\mathcal{U}, \mathcal{W}]\rangle\big)\end{aligned} \tag{C.10}$$

which gives us the exact form of $\langle \nabla^{\mathrm{LC}}(\mathcal{U}, \mathcal{V}), \mathcal{W}\rangle$ and therefore fixes $\nabla^{\mathrm{LC}}(\mathcal{U}, \mathcal{V})$.

To get the form explicitly, we have to compute the 'Γ^k_{ij}'s in terms of the metric; condition (i) of the theorem implies that

$$\nabla^{\mathrm{LC}}\left(\frac{d}{du_j}, \frac{d}{du_k}\right) = \nabla^{\mathrm{LC}}\left(\frac{d}{du_k}, \frac{d}{du_j}\right) + \left[\frac{d}{du_j}, \frac{d}{du_k}\right] \tag{C.11}$$

but $\left[\frac{d}{du_j}, \frac{d}{du_k}\right] = \frac{d}{du_j} \circ \frac{d}{du_k} - \frac{d}{du_k} \circ \frac{d}{du_j} = 0$ by equality of mixed partials*, so by definition (Equation C.5) we have $\Gamma^k_{ij} = \Gamma^k_{ji}$. Then Equation C.10 implies that

$$\left\langle \nabla^{\mathrm{LC}}\left(\frac{d}{du_i}, \frac{d}{du_j}\right), \frac{d}{du_k}\right\rangle = \frac{1}{2}\left(\frac{d}{du_i}\left\langle \frac{d}{du_j}, \frac{d}{du_k}\right\rangle + \frac{d}{du_j}\left\langle \frac{d}{du_k}, \frac{d}{du_i}\right\rangle\right.$$

*That is all this is, really.

$$
\begin{aligned}
& -\frac{d}{du_k}\left\langle \frac{d}{du_i}, \frac{d}{du_j} \right\rangle - \left\langle \frac{d}{du_k}, \left[\frac{d}{du_j}, \frac{d}{du_i} \right] \right\rangle \\
& + \left\langle \frac{d}{du_i}, \left[\frac{d}{du_k}, \frac{d}{du_j} \right] \right\rangle - \left\langle \frac{d}{du_j}, \left[\frac{d}{du_i}, \frac{d}{du_k} \right] \right\rangle \Big) \\
= \; & \frac{1}{2}\left(\frac{d}{du_i}\left\langle \frac{d}{du_j}, \frac{d}{du_k} \right\rangle + \frac{d}{du_j}\left\langle \frac{d}{du_k}, \frac{d}{du_i} \right\rangle \right. \\
& \left. -\frac{d}{du_k}\left\langle \frac{d}{du_i}, \frac{d}{du_j} \right\rangle \right) \\
= \; & \frac{1}{2}\left(\frac{d}{du_i} g_{jk} + \frac{d}{du_j} g_{ki} - \frac{d}{du_k} g_{ij} \right) \qquad \text{(C.12)}
\end{aligned}
$$

But by the definition of the 'Γ^k_{ij}'s and the properties of inner products, we also have

$$
\begin{aligned}
\left\langle \nabla^{\mathrm{LC}}\left(\frac{d}{du_i}, \frac{d}{du_j} \right), \frac{d}{du_k} \right\rangle &= \left\langle \sum_{i=1}^{n} \Gamma^i_{jk} \frac{d}{du_i}, \frac{d}{du_k} \right\rangle \\
&= \sum_{i=1}^{n} \left\langle \Gamma^i_{jk} \frac{d}{du_i}, \frac{d}{du_k} \right\rangle \\
&= \sum_{i=1}^{n} \Gamma^i_{jk} \left\langle \frac{d}{du_i}, \frac{d}{du_k} \right\rangle \\
&= \sum_{i=1}^{n} \Gamma^i_{jk} g_{ki} \qquad \text{(C.13)}
\end{aligned}
$$

Cleverly, we note that

$$
\sum_{k=1}^{n} g^{kl}\left(\sum_{i=1}^{n} \Gamma^i_{jk} g_{ki} \right) = \sum_{k=1}^{n}\sum_{i=1}^{n} g^{kl} g_{ki} \Gamma^i_{jk} = \sum_{k=1}^{n} \delta_{ik} \Gamma^i_{jk} = \Gamma^i_{jk}. \qquad \text{(C.14)}
$$

So combining everything we must have

$$
\Gamma^i_{jk} = \frac{1}{2}\sum_{j=1}^{n} g^{kj}\left(\frac{d}{du_i} g_{jk} + \frac{d}{du_j} g_{ki} - \frac{d}{du_k} g_{ij} \right) \qquad \text{(C.15)}
$$

On the other hand, these '$\Gamma^i_j k$'s specify a connection, and the reverse of the argument just given shows that this connection satisfies the conditions of the theorem. QEF

Corollary C.1. The connection coefficients for the Levi-Civita connection in $\mathbf{R}^n$ are all zero.

Proof. In Euclidean space the coefficients of the metric (the 'g_{ij}'s) are all constant. Therefore in equation C.15 we are differentiating constants and summing. The result of this operation is always zero. QEF

$$
\begin{aligned}
\langle \nabla^{\mathrm{LC}}(u,v), w \rangle &= u\langle v, w\rangle - \langle v, \nabla^{\mathrm{LC}}(u,w)\rangle \\
&= u\langle v, w\rangle - \langle v, \nabla^{\mathrm{LC}}(w,u)\rangle - \langle v, [u,w]\rangle\rangle \\
&= u\langle v, w\rangle - w\langle v,u\rangle + \langle \nabla^{\mathrm{LC}}(w,v),u\rangle - \langle v,[u,w]\rangle\rangle \\
&= u\langle v, w\rangle - w\langle v,u\rangle + \langle \nabla^{\mathrm{LC}}(w,v),u\rangle + \langle [w,v],u)\rangle - \langle v,[u,w]\rangle\rangle \\
&= u\langle v, w\rangle - w\langle v,u\rangle + v\langle w,u\rangle - \langle w, \nabla^{\mathrm{LC}}(v,u)\rangle + \langle [w,v],u\rangle - \langle v,[u,w]\rangle\rangle \\
&= u\langle v, w\rangle - w\langle v,u\rangle + v\langle w,u\rangle - \langle w, \nabla^{\mathrm{LC}}(v,u)\rangle - \langle w,[v,u]\rangle + \langle [w,v],u\rangle - \langle v,[u,w]\rangle\rangle.
\end{aligned}
$$

Table C.1: Computing the form of the Levi-Civita connection

Having gone through the hard work of defining the 'right' connection, we can now begin to put our system to work. (We will then see why this connection is the right one.)

Definition C.14. Given a chart (U, φ) and a smooth curve $\gamma : [a, b] \to U \subset M$ such that $U \supset \gamma$ we define the *derivative of a vector field* $\mathcal{V}$ *along* γ in coordinates as follows: we write $\mathcal{V}(t) = \sum_{j=1}^{n} \xi_j(t) \frac{d\gamma}{du_j}|_{\gamma(t)}$ and then write

$$\nabla^{\mathrm{LC}}\left(\frac{d\gamma}{dt}, \mathcal{V}(t)\right) = \sum_{i=1}^{n}\left(\left(\left.\frac{dxi_i}{dt}\right|_{t_0} + \sum_{j=1}^{n}\sum_{k=1}^{n}(\Gamma^i_{jk} \circ \gamma)\xi_j(t_0)\left(\left.\frac{d\gamma_j}{dt}\right|_{t_0}\right)\right)\left.\frac{d\gamma}{du_i}\right|_{\gamma(t_0)}\right). \tag{C.16}$$

Remark C.11. The same formal definition works for other connections, in fact the definition is essentially the same, but since the connection coefficients are coming from the Levi-Civita connection, this derivative respects the metric and is torsion free.

Definition C.15. If $\gamma : (a, b) \to M$ is a path in M, we say that $\mathcal{V}$ is *parallel along* γ if

$$\nabla^{\mathrm{LC}}\left(\frac{d\gamma}{dt}, \mathcal{V}(t)\right) = 0 \tag{C.17}$$

for all $t \in (a, b)$.

Definition C.16. We say a path $\gamma : (a, b) \to M$ is a *geodesic* if

$$\nabla^{\mathrm{LC}}\left(\frac{d\gamma}{dt}, \frac{d\gamma}{dt}\right) = 0 \tag{C.18}$$

for all $t \in (a, b)$, *i.e.* if the tangent vector to γ is parallel along γ on the entire interval on which γ is defined.

Remark C.12. To write the equation for a geodesic in local coordinates, we set $\gamma_j = x_j \circ \gamma$ as above. Then our definition for the derivative of a vector field along a curve (Definition C.14) implies that Equation C.18 becomes

$$\sum_{i=1}^{n}\left(\left(\left.\frac{d^2xi_i}{dt^2}\right|_{t_0} + \sum_{j=1}^{n}\sum_{k=1}^{n}(\Gamma^i_{jk} \circ \gamma)\left(\left.\frac{d\gamma_j}{dt}\right|_{t_0}\right)\left(\left.\frac{d\gamma_j}{dt}\right|_{t_0}\right)\right)\left.\frac{d\gamma}{du_i}\right|_{\gamma(t_0)}\right) = 0. \tag{C.19}$$

We can rewrite this second order equation as a first order system, and after that, rewrite the condition of being a geodesic in terms of the condition of being an integral curve for a certain vector field.

To each chart (U, φ) on M we associate a chart (Q, ψ) on TM by putting $Q = \pi^{-1}(U)$ and

$$\psi(\xi) = (\varphi \circ \pi(\xi), \xi(u_1), \ldots, \xi(u_n)) \in \mathbf{R}^{2n} \tag{C.20}$$

on Q. We write $q(\xi) = \varphi \circ \pi(\xi) \in \mathbf{R}^n$ and $\dot{q}(\xi) = (\xi(u_1), \ldots, \xi(u_n)) \in \mathbf{R}^n$. Using this notation we have

$$\xi = \sum_{j=1}^{n} \dot{q}_j(\xi) \left.\frac{d}{du_j}\right|_{\pi(\xi)}. \tag{C.21}$$

The differential equation above can be rewritten as

$$\frac{dq_j}{dt} = \dot{q}_j \tag{C.22}$$

$$\frac{d\dot{q}_j}{dt} = -\sum_{j=1}^{n}\sum_{k=1}^{n}(\Gamma^i_{jk} \circ \pi)\dot{q}_j\dot{q}_k. \tag{C.23}$$

The solution to this system are integal curves of the vector field $\mathcal{G}$ defined on Q by

$$\mathcal{G} = \sum_{i=1}^{n}\left(\dot{q}_j\frac{d}{dq_j} - \sum_{j=1}^{n}\sum_{k=1}^{n}(\Gamma^i_{jk} \circ \pi)\dot{q}_j\dot{q}_k\frac{d}{d\dot{q}_i}\right). \tag{C.24}$$

Definition C.17. Let $\varphi(t, \xi)$ denote the geodesic flow of the vector field $\mathcal{G}$ of the previous remark (*i.e.* the map that tames points in M to the points in the direction given by $\mathcal{G}$).

Let

$$\gamma_\xi : [0, 1] \ni t \mapsto \pi \circ \varphi(t, \xi) \in M.$$

Let T^1M denote

$$\{\xi \in TM : 1 \in \text{maximal interval on which } \gamma_\xi \text{ is defined}\}.$$

We define

$$\exp : T^1M \ni \xi \mapsto \gamma_\xi(1) \in M.$$

Remark C.13. Frequently the manifolds we work with in this thesis will be *complete*, that is, their geodesics will run on forever (although they may, as in the case of spheres, come back and trace over themselves again, or, as in the case of tori, come back and intersect themselves infinitely often). *In the case of complete manifolds,* exp *is defined on all of* TM.

We note that one sometimes sees the related definition

$$\exp t\cdot : T^1M \ni \xi \mapsto \gamma_\xi(t) \in M, \tag{C.25}$$

for example in the proof of Theorem 2.1.

Differential forms

Recall from the section on algebraic preliminaries that the dual of a finite dimensional vector space is the set of linear maps from that vector space to the base field, and that it is isomorphic to the original vector space. Since tangent spaces are finite dimensional real vector spaces, it is very easy to talk about their duals, which are called cotangent spaces[†]. Just to point out exactly where things stand, tangent vectors are a special kind of linear maps from functions on the manifold to the real numbers – and now we have cotangent vectors which are all of the linear maps taking that set of maps into the real numbers! The basis of the cotangent space that is dual to the standard basis of the tangent space is denoted $\{du^i\}$.

Definition C.18. The *cotangent bundle* of a manifold is the union over x in M of all the the cotangent spaces.

If we wanted to, we could write this in the careful way we wrote the tangent bundle above. All of the same ideas apply. In particular, we can define sections of the cotangent bundle. We are especially interested in the smooth sections, which we give a special name.

[†]This name has even less to do with the cotangent function from trigonometry than tangent spaces have to do with the tangent function – which is, admittedly, something, but not a lot.

Definition C.19. A smooth m-section of the cotangent bundle (that is, a choice of m vectors in each cotangent space that varies smoothly with respect to the smooth structure on the cotangent bundle) is called a *differential m-form* (this is usually just shortened to m-form). We extend this in a slightly odd way by say that smooth functions on the manifold are *zero forms*[‡]. The collection of all m-forms ($m \in 0, 1, \ldots$) is the set of *differential forms* (or sometime, just *forms*). After lumping them all together, we sort them back out again by saying a differential form has *degree* m if it is a m-form.

The fundamental binary operation on differential forms is the $\wedge$ product ('wedge product'). This is defined by the action of differential 1-forms on tangent vectors and its important because it defines the action of differential m-forms for higher m on lists of tangent vectors, *i.e.* on sections of the tangent bundle. (The key point is that the m-form is associated with the wedge product of all of the 1-forms that make it up.)

Remark C.14. The standard short-hand for lists of m vectors is *m-vectors*.

If ω is a m-form and υ is a p-form, their wedge product $\omega \wedge \upsilon$ takes an $(m+p)$-vector $(\xi_1, \ldots, \xi_{m+p})$ to the number

$$\frac{1}{m!p!} \sum_{\sigma \in \mathfrak{S}^{m+p}} \operatorname{sign}(\sigma) \omega(\xi_{\sigma(1)}, \ldots, \xi_{\sigma(m)}) \upsilon(\xi_{\sigma(m+1)}, \ldots, \xi_{\sigma(m+p)}). \tag{C.26}$$

If ω and υ are *1-forms* this complicated expression reduces to

$$\sum_{\mathfrak{S}^2} \operatorname{sign}(\sigma) \omega(\xi_{\sigma(1)}) \upsilon(\xi_{\sigma(2)}) = \omega(\xi_1)\upsilon(\xi_2) - \omega(\xi_2)\upsilon(\xi_1); \tag{C.27}$$

In the case that one of the factors is a zero form, things get even easier since the wedge product reduces to standard multiplication.

In the case of a one form and a two form, the expression is almost as simple as it is for two one forms; and from here, various properties can be deduced[§] by induction.

[‡]Actually, I think that this might not be *so* odd; it might make sense for it to be standard practice to say that smooth functions are zero sections of any bundle

[§]Or, induced.

There are three fundamental unary operations on differential forms; their names are ι, d, and δ.

Actually, to be precise, ι is map defined from vector fields *into* the space of operators on differential forms. The definition is that $\iota(\mathcal{V})$ takes m-forms to $(m-1)$-forms that act by the following rule:

$$(\iota(\mathcal{V})\omega)(\xi_1, \ldots, \xi_{m-1}) = \omega(\mathcal{V}, \xi_1, \ldots, \xi_{m-1}).$$

The map ι is called the *substitution operator*. (At last, a nice, descriptive name! – the reason we use the *symbol* 'ι' is perhaps less clear at this point; the reason is that ι is associated with integration.)

The operator d is called the *coboundary operator*; it is a map from m-forms to $(m+1)$-forms, and it is defined by these properties:

1. $d(\omega + \upsilon) = d\omega + d\upsilon$

2. $d(\omega \wedge \upsilon) = d\omega \wedge \upsilon + (-1)^{\text{degree}(\omega)} \wedge d\upsilon$

3. $dd\omega = 0$

4. for all zero forms f, $df = \sum_{i=1}^{n} \frac{df}{du_i} du_i$

The 'd' appearing at the beginning of 'du_i' is a special case of this 'd'.

At this point we go on to define a couple of essential operations that have something to do with forms. *The following two definitions are almost impossible to keep straight, so pay attention*[¶].

Definition C.20. The *pushforward* of a map $\varphi : M \to N$ is a map $\varphi_* : TM \to TN$ defined so that so that if γ is a curve in M with tangent vector ξ at time t_0, then φ_* is the map that takes $\xi \in TM$ to

[¶] In the spirit of SOHCAHTOA, the acronym BOTCGNTTOFFEPF (Bottom On Tangents, Curve Gives New Tangent; Top On Forms, Form Eats Push Forward) may serve as a mneumonic tool to help the reader remember which one is which, and what they do.)

the vector in TN that is tangent to the curve $\varphi(\gamma(t_0))$.

The map φ_* is sometimes called the *differential of* φ and written as $d\varphi$. One can explore the relationship of this 'd' to the one above.

Definition C.21. The *pullback* of a differential form is defined whenever we have a smooth map φ between two manifolds, say $\varphi : M \to N$. If v is a form on N, then $\varphi^*(v)$ is a form on M determined by

$$(\varphi^*(v))(\xi_1, \ldots, \xi_m) = v(\varphi_*\xi_1, \ldots, \varphi_*\xi_m).$$

Since have been talking about a *coboundary* operator, it might seem that we should have already at least mentioned the *boundary* operator. In fact, we have used this operator throughout the thesis, although we have not been terribly explicit about what we were doing. The boundary operator we are refering to just finds boundary of subsets in the manifold, in symbols, $\partial A = \bar{A} \setminus A^\circ$ for all $A \subset M$. It is a basic topological property that

$$\partial\partial A = \emptyset \tag{C.28}$$

for all subsets A. We are using the prefix 'co' approximately as we used it to talk about the cotangent bundle, *i.e.* we are talking about objects that are dual to one another in some sense. So far, we have defined the coboundary operator to act on the space of differential forms, which in turn act on the space of sections. We now use the definition of the action of differential forms on sections to define their action on domains. The way we let a differential m-form ω act on a m-dimensional domain C is by integrating; we define $\int_\Omega \omega$ to be $\int_{\{p\in\Omega\}} \omega(\xi_1(c), \ldots, \xi_n(c))$, where $\{\xi_k(p)\}_{k=1}^n$ is the ordered basis for $T_p\Omega$.

The coboundary operator d is dual to the boundary operator ∂ in the sense that

$$\int_{\partial\Omega} \omega = \int_\Omega d\omega. \tag{C.29}$$

This important equation has a name: it is called Stokes' Theorem. A nice proof of Stokes' Theorem is written down in [18]. We note that in the special case that Ω is the boundary of some set

A, C.28 implies that the left hand side of C.29 is zero. Hence, we deduce that the integral over a boundary of the coboundary of any form is zero.

Our last word on special operators applies to the operator δ. It extremely interesting, and we can not get away without mentioning that it is extremely useful for talking about the shape and structure of manifolds, and that there is a dual operator that does the same thing for domains. But that is where we have to stop.

There are various inner products and norms defined on differential forms, and we close this section with a list of the most important ones.

Definition C.22. We say an m-form is *simple* if it can be written as a one-term sum.

We can define an inner product between simple forms $\omega = w_1 \wedge \ldots \wedge w_m$ and $v = v_1 \wedge \ldots \wedge v_m$ based at p by

$$\langle\langle \omega, v \rangle\rangle = \sum_{\sigma \in \mathfrak{S}} \operatorname{sign}(\sigma) \prod_{k=1}^{m} \langle w_{\sigma(k)}, v_k \rangle_p. \tag{C.30}$$

(We will only use the double angle bracket notation in the appendix.)

The term on the right of Equation C.30 is clearly equal to

$$\sum_{\sigma \in \mathfrak{S}} \operatorname{sign}(\sigma^{-1}) \prod_{k=1}^{m} \langle w_k, v_{\sigma^{-1}(k)} \rangle_p,$$

and so our definition is symmetric. Linearity of the standard inner product on vectors extends the definition to all m-vectors. The norm induced by this inner product is called the *Riemannian norm*.

Remark C.15. Since we can easily take an m-vector and define a m-form (by taking the wedge product of all of the vectors in the list), we can use the above definition to define a product between m-forms and m-vectors, which we also denote by $\langle\langle \cdot, \cdot \rangle\rangle$.

Definition C.23. The *mass* of a differential m-form ω is

$$\operatorname{mass}(\omega) = \sup\{\langle\langle \omega, \Xi \rangle\rangle; \Xi \text{ is an } m-\text{vector, and } \operatorname{comass}(\Xi) \leq 1\}$$

where comass is about to be defined.

Definition C.24. The *comass* of a m-vector Ξ is

$$\text{comass}(\Xi) = \sup\{\langle\langle \omega, \Xi \rangle\rangle; \omega \text{ is an } m-\text{form}, \xi \text{ is simple , and } |\omega| \leq 1\}$$

(here $|\cdot|$ denotes the Riemannian norm).

Isometries

Definition 7.1. A map φ from M to M is an *isometry* if for all $p \in M$, if $\xi, \eta \in T_pM$ then $g(\xi, \eta) = g(\varphi_*\xi, \varphi_*\eta)$.

We state without proof the fact that the set of isometries for a manifold form a group; this group is appropriately called the *isometry group*.

Definition C.25. The *quotient* of a manifold by its isometry group is the set obtained by identifying each point with the point(s) in its orbit(s) under the isometry group.

D Analytic Preliminaries

Charter Review of the basic results of Analysis germane to our discussion.

Measures

Note the similarities between the following definition and the definition of a topological space (see appendix on Topological Preliminaries).

Definition D.1. *Measureable space* This is a pair $(X, \mathcal{M})$ such that $\mathcal{M}$ is comprised of subsets of X and has the further additional properties

i. $X \in \mathcal{M}$.

ii. The complement in X of any $A \in \mathcal{M}$ is again in $\mathcal{M}$.

iii. countable unions of elements of $\mathcal{M}$ are again in $\mathcal{M}$.

If $\mathcal{M}$ satisfies all of these properties, it is called a *σ-algebra*. The elements of $\mathcal{M}$ are then the *measurable sets of X with respect to $\mathcal{M}$* or just *measurable sets*.

Definition D.2. A *measure* is a function μ taking a σ-algebra $\mathcal{M}$ into the complex numbers $\mathbf{C}$ with the property that $\mu(\emptyset) = 0$, and that satisfies an additional criterion, namely that for all countable disjoint collections $\{A_i; A_i \in \mathcal{M}\}$,

$$\mu\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Functions with this property are said to satisfy *countable additivity*.

Remark D.1. Reasonably, we say that such a function μ is a *real* measure if its image is contained in the reals. Slightly less reasonably, but rather, according to convention, we say that a measure μ is *positive* if its image lies in the interval $[0, \infty]$.

The definition of a positive measure is unreasonable for two reasons: the first is easy to catch, namely it might be better to call such measures *non-negative* rather than positive, because zero is not positive. The second reason is that even though we allow ∞ into the range of a positive measure, $\infty \notin \mathbf{C}$, so by the above definition for measure it would never be hit. We make the special dispensation that μ may assign ∞ as the measure of a set as long as μ is positive.

We put this all together into one main definition as follows.

Definition D.3. A *measure space* is a measurable space $(X, \mathcal{M})$ together with a measure μ defined on $\mathcal{M}$.

Remark D.2. The definition of a measurable space is frequently written out as a triple, $(X, \mathcal{M}, \mu)$, rather than as a pair $((X, \mathcal{M}), \mu)$. Per usual, if $S = (X, \mathcal{M}, \mu)$ is a measure space, $p \in S$ means $p \in X$ (and in particular, a map defined on S is a map defined on X.).

The structure of maps on measure spaces are understood in terms of the measurable sets of the underlying σ-algebra.

Definition D.4. A mapping from a measure space $(X, \mathcal{M}, \mu)$ into a topological space (Y, τ) is *measurable* if the inverse image of every open set $U \in \tau$ is an element of $\mathcal{M}$, *i.e.*, is a measurable set.

Remark D.3. When talking about a measure space, we say that a set is μ-measurable if it is in the underlying σ-algebra. The terms 'almost everwhere', 'almost all', or, more specifically, 'μ-almost all', 'μ-almost everywhere', etc. when used to describe some property of a measure space mean that the property holds at all points except some set of μ-measure zero.

Example. A *simple function* on a measure space S is a map taking S into $[0, \infty) \subset \mathbf{C}$ that has only finitely many points in its range.

Since the range of a simple function ρ defined on S is finite, S splits into a finite number of subsets $A_1, \ldots, A_n$ such that $\rho(A_1) = a_1, \ldots, \rho(A_n) = a_n$. We can therefore write

$$\rho = \sum_{i=1}^{n} a_i \chi_{A_i},$$

where the 'χ's are *indicator functions*, defined in general by $\chi_A : S \to \{0, 1\}, p \mapsto 1$ if $p \in A$, $p \mapsto 0$ else.

A simple function ρ is measurable if and only if all of the 'A_i's are measurable, since the inverse image of an open interval will be a union of some subset of the 'A_i's and so measurable if all of the 'A_i's are by item (iii) of the definition of a σ-algebra; on the other hand, if A_1, say, is not measurable, then the inverse image of some small open interval about a_1 is equal to A_1 and ρ then not measurable.

(We can show that there is a sequence of simple functions that approximate any given measurable function if need be.)

Abstract integration

Rudin [45] treats this very thoroughly; the reader is refered to him for most of the proofs.

Definition D.5. If $S = (X, \mathcal{M}, \mu)$ is a measure space, $E \in \mathcal{M}$ and $\rho = \sum_{i=1}^{n} a_i \chi_{A_i}$ is a measurable simple function on S, we define the *integral of ρ over E with respect to the measure μ* by

$$\int_E \rho \, d\mu = \sum_{i=1}^{n} a_i \mu(A_i \cap E).$$

In general, if $f : S \to [0, \infty]$ is measurable, we define the *(Lebesgue) integral of f over E with respect to the measure μ* by

$$\int_E f \, d\mu = \sup_{\substack{\rho \text{ simple}, \\ 0 \le \rho \le f}} \int_E \rho \, d\mu.$$

In the case of measurable simple functions, it is easy to see that the two definitions give the same result

We get all the standard properties we expect of integrals (*i.e.*, integration respects majorization, is linear, etc.). These things are fairly easy to see for simple functions (we use countable additivity as needed) and then they carry over to the rest of the measurable positive functions because we can approximate such functions pointwise from below by a sequence of simple functions (so the supremum in the definition of the integral is). A proof that we can approximate measurable functions by simple functions in this manner is not at all difficult, but for our purposes a sketch will suffice: the idea is to approximate f from below wherever its values lie in $[0, n]$ by a simple function whose accuracy of approximation depends on n. Wherever f takes values greater than n, it is approximated by a step of size n. As n grows, we specify a finer and finer approximation on a growing finite interval, and so we converge wherever f takes finite values. If f takes the value infinity at some point or on some interval, it is approximated there by a sequence of steps of size n.

We also get a number of new consequences that show the superiority of the Lebesgue integral to the Riemann integral.

Theorem D.1 (Monotone Convergence). Let $\{f_n\}$ be a sequence of measurable functions on a measure space $S = (X, \mathcal{M}, \mu)$ and suppose that

i. $0 \leq f_1(x) \leq \ldots \leq \infty$ for every $x \in S$.

ii. $f_n(x) \to \infty$ as $n \to \infty$ for every $x \in S$.

Then f is measurable and $\int_S f_n \, d\mu \to \int_S f \, d\mu$.

Theorem D.2 (Fatou's Lemma). If $\{f_n : S \to [0, \infty]\}$ is a sequence of measurable functions on a measure space $S = (X, \mathcal{M}, \mu)$, then for each positive integer n we have

$$\int_S \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int_S f_n \, d\mu$$

We now extend the definition of integration to cover integration of complex functions. This is the general definition of the Lebesgue integral.

Definition D.6. The *Lebesgue integral* over any measurable set is defined by breaking the function into its positive real, negative real, positive imaginary, and negative imaginary parts (which then all give positive functions), and integrating each seprarately, then multiplying by $1, -1, i$, or $-i$ as appropriate and summing.

Associated with this is the class of 'integrable functions'.

Definition D.7. The space $L^1(\mu)$ is the set of functions defined on all of $S = (X, \mathcal{M}, \mu)$ with the property that

$$\int_S |f| \, d\mu < \infty.$$

Theorem D.3 (Dominated Convergence). Suppose that $\{f_n : S \to \mathbf{C}\}$ is a measurable sequence of functions on a measure space $S = (X, \mathcal{M}, \mu)$ with the property that $\lim_{n\to\infty} f_n(x)$ exists for every $x \in S$. We call this function $f(x)$. If there is a function $g \in L^1(\mu)$ such that for all positive integers n and all $x \in S$ we have

$$|f_n(x)| \leq g(x)$$

then we conclude the following properties must hold for f:

1. $f \in L^1(\mu)$
2. $\lim_{n\to\infty} \int_S |f_n - f|\, d\mu = 0$
3. $\lim_{n\to\infty} \int_S f_n\, d\mu = \int_S f\, d\mu$.

Finally, we state one other related theorem:

Theorem D.4 (Fubini's Theorem). Let (X, S, μ) and (Y, T, λ) be σ-finite measure spaces and let f be an $(S \times T)$-measureable function on $X \times Y$.

- If $0 \leq f \leq \infty$ and if for all $x \in X$ and $y \in Y$ we have

$$\varphi(x) = \int_Y f_x\, d\lambda$$

and

$$\psi(x) = \int_X f^y\, d\mu$$

then φ is S-measurable, ψ is T-measurable, and

$$\int_X \varphi\, d\mu = \int_{X\times Y} f\, d(\mu \times \lambda) = \int_Y \varphi\, d\lambda$$

- If f is complex and if

$$\varphi*(x) = \int_Y |f|_x\, d\lambda$$

and

$$\int_X \varphi*\, d\mu < \infty$$

then $f \in L^1(\mu \times \lambda)$

- If $f \in L^1(\mu \times \lambda)$, the $f_x \in L^1(\lambda)$ for almost all $x \in X$, $f^y \in L^1(\mu)$ for almost all $y \in Y$ and the functions φ and ψ

Lebesgue measure

Definition D.8. A *Borel measure* is a measure on open sets.

Theorem D.5. Up to a constant multiple, there is a unique translation invariant positive Borel measure on $\mathbf{R}^n$.

Remark D.4. This measure is called the *Lebesgue measure*. In general, an isometry invariant measure is called a *Harr measure*. Each of $\mathbf{S}^n$ and $\mathbf{H}^n$ has, again up to a constant multiple, a unique Harr measure.

Remark D.5. Note that the Harr measures are in some sort of abstract sense like the Levi-Civita connection. Both respect the basic geometrical structures of the spaces they are defined on.

Miscellaneous

Definition D.9. The lim inf (lim sup) of a sequence of functions is...

Definition D.10. A function f is *upper semi-continuous with respect to a metric* δ, if

$$\delta(a_k, a) \to 0 \Rightarrow \limsup_{k\to\infty} f(a_k) \leq f(a).$$

Definition D.11 (Alternative definition, see Rudin, p. 39). A function f on a manifold M is *upper semi-continuous* if $\{x : f(x) < \alpha\}$ is open for every real α.

Definition D.12. A function or form or whatever has *compact support* if the set on which it takes nonzero values is compact.

Definition D.13. A function from one metric space (X) to another (Y) is said to satisfy a Hölder condition of order α or to be *Hölder continuous* if there exist M and α such that for all $x, y \in X$ we have $d_Y(x, y) < M(d_X(x, y))^\alpha$.

Definition D.14. A function is *Lipschitz* if the α in Definition D.13 is equal to one.

Definition D.15. A function is *Hölder contiuously differentiable* or $C^{1,\alpha}$ if is differentiable and the differentiated function satisfies a Hölder condition of order α.

Definition D.16. A *semi-norm* is a real valued function $p(x)$ defined on a vector space such that

i. $0 \leq p(x) < \infty$,

ii. $p(x, y) \leq p(x) + p(y)$,

iii. $p(ax) = |a|p(x)$.

Definition D.17. A function is *real analytic* in a neighborhood of p if there is a convergent power series with real coefficients with center p such that at every point in the neighborhood the value of the function and the sum of the power series coincide. (Complex analytic functions are defined similarly.)

E Variational Calculus

Charter. We should understand how to find extrema for functionals defined on a set of geometric objects.

We won't be able to fully deal with this charter, but we will do what we can. We begin with a definition that has little or nothing to do with variational calculus.

Definition E.1. The *second fundamenatal form* of an n-dimensional manifold M is defined in terms of local coordinates at a point $p \in M$ by using the fact that by embedding M in some higher-dimensional $\mathbf{R}^N$, we may view M locally as a graph of some function $f_p : T_pM \to \mathbf{R}$, where at each

point $\mathbf{R}$ is viewed as a line in the direction of some normal vector to M at p. Specifically, we put

$$\mathbf{Sec}_p(M) = \begin{pmatrix} \frac{\partial^2 f_p}{\partial^2 x_1}\Big|_p & \cdots & \frac{\partial^2 f_p}{\partial^2 x_1 \partial^2 x_m}\Big|_p \\ \frac{\partial^2 f_p}{\partial x_n \partial x_1}\Big|_p & \cdots & \frac{\partial^2 f_p}{\partial^2 x_n}\Big|_p \end{pmatrix} \tag{E.31}$$

where $x_n, \ldots, x_n$ are coordinates for T_pM. The matrix $\mathbf{Sec}_p(M)$ defines a bilinear form on T_pM, which we call by the same name.

Remark E.1. Since $\mathbf{Sec}_p(M)$ is symmetric, we can choose a basis for T_pM so that it becomes diagonal (see for example [20, page 266], and think about the geometrical content of the Corollary there). The numbers on the diagonal are called the *principal curvatures of M at p* and are traditionally labeled $\kappa_1, \ldots, \kappa_n$.

Definition E.2. The *mean curvature of M at p* is defined to be

$$H = \text{trace}(\mathbf{Sec}_p(M)).$$

The name 'mean curvature' might make more sense of course if H was defined to be $\text{trace}(\mathbf{Sec}_p(M))$; however, several nice equalities depend our definition.

Definition E.3. Given a smooth, compactly supported vectorfield $\mathcal{V}$ defined on M *first variation of area for a manifold with respect to* $\mathcal{V}$ is defined to be

$$\delta^1(M) = \frac{d}{dt}\Big|_{t=0} \text{area}(\text{M} + \text{t}\mathcal{V}),$$

where $M + t\mathcal{V}$ means the geodesic flow at time t of points in M in the directions given by $\mathcal{V}$.

The first variation of area measures the infinitesimal change in area due to an infinitesimal change in the shape of the manifold.

Theorem E.1. The first variation of area is given by integrating against mean curvature as follows:

$$\delta^1(M) = -\int_M \mathcal{V} \cdot H\nu,$$

where ν is a unit normal to M

We need a simple lemma.

Theorem E.1. Changing the radius of a circle of radius $r = 1/\kappa$ by dr changes the arclength of an infinitesimal sector from ds to $(1 - \kappa dr)ds$.

Proof. Putting the new radius equal to $r' = 1/k - dr$ and the length of the new sector equal to $(ds)'$, we need to prove that $\frac{r'}{r} = \frac{(ds)'}{ds}$. But $ds = rd\theta$, so $(ds)' = r'd\theta$. QEF

Proof of Theorem E.1. Since integration is linear, we can decompose $\mathcal{V}$ into normal and tangential components, and integrate each separately. In the case of a purely tangential variation, the surface does not change area, so the theorem is correct. In the case of a purely normal variation, we may write $\mathcal{V} = V\nu$; then, by refering the the preceeding lemma to see what happens in each direction, we find that an infinitesimal area $du_1 \wedge \ldots \wedge$ is distorted to

$$(1 - V\kappa_1)du_1 \wedge \ldots \wedge (1 - V\kappa_n)du_n, \tag{E.32}$$

where the 'κ_i's are the principal curvatures. Equation E.32 is equal to

$$(1 - V\nu \cdot H\nu)du_1 \wedge \ldots \wedge \tag{E.33}$$

i.e.

$$(1 - \mathcal{V} \cdot H\nu)du_1 \wedge \ldots u_n \wedge . \tag{E.34}$$

The initial rate of change in area is measured by integrating this over the manifold and then subtracting off the integral of $du_1 \wedge \ldots u_n \wedge$ over the manifold. The theorem follows. QEF

With the idea of a variation now in mind, we might, if we had time, go on to prove lots of things. For example, geodesics can be studied from the variational viewpoint. However, we don't have time for that.

The main thing we are interested in showing is that $H = d(Area)/d(Volume)$; we use the result of Theorem E.1, and state this as a proposition.

Theorem E.1. The mean curvature of a surface S at a point is given by the derivative of area with respect to volume $d(Area)/d(Volume)$.

Proof. Since we are concerned with the behavior at a point, we can assume that the change in volume comes from a variation vector field that is defined on an infinisimal piece of area, and so can be considered constant. From Theorem E.1, we then have $dA/dt = H \int_S f$. Since everything is inifintesimal, we have $dV/dt = \int_S f$. The chain rule $dA/dV = (dA/dt)(dt/dV) = (dA/dt)(dV/dt)$ then finishes the proof. QEF

This proposition means that for our problem, we can think of mean curvature as an exchange rate (how much area is obtained for a certain amount of volume, or vice versa). A little thought then shows that the each of the surfaces separating two given components of a minimizer for the least-area problem with volume constraints have *globally* constant mean curvature. Of course, as with the standard double bubble, different pieces of surface may have different curvatures, because as we showed at the end of Chapter 2, curvature is proportional to pressure difference.

The ideas behind the following two definitions are undercurrents of the entire thesis:

Definition E.4. A surface is *stable* for the least-area problem if small variations do not decrease area.

Definition E.5. A surface is in *equilibrium* for the least-area problem if for any variation that fixes volumes, the first variation of area is zero.

This has been a very quick and incomplete introduction to the subject of variational calculus, giving us only the bare minimum of what we need for this thesis. There are many places to read more; [?] is good for further applications to area-minimization and geometric measure theory. In particular, there (in Part B) we find a very general and elegant description of the category of variational problems – which again, there is not time to talk about here.

indices

Dramatis Personæ

$\mathbf{R}^n$	n-dimensional Euclidean space
$\mathbf{S}^n$	n-dimensional Spherical space
$\mathbf{H}^n$	n-dimensional Hyperbolic space
$\mathbf{D}_p(r)$	a disk of radius r centered at p
B	a double bubble
I	an isoperimetric profile
$\mathit{\Pi}$	a two-volume least-area function
$\mathbf{G}^n$	Gauss space
Y	the standard partition of $\mathbf{G}^n$ into three equal pieces
$\mathbf{T}^2$	a two-torus
$\mathbf{T}^3$	a three-torus
Θ	a standard double bubble
Ξ	a double band
Φ	a band lens
O	an octagon square

the characters of the appendix

X	a set of points
τ	a topology
M	a manifold
p	a point in M
T_pM	the tangent space to M at p
f	a differentiable map
$\mathcal{M}$	a σ-algebra

Also necessary throughout are various numbers, unary and binary operators, delimiters, etc.

The play has been performed in its entirity only once, in the Great Oklahoma Theater, where it was directed by Joseph K. (see [26]). That production included a cast of thousands. An abriged version is to be presented in the Soo Bong Chae Memorial Auditorium, May 6, 2002.

Index of notational conventions used in this thesis

Euclidean space

Many authors (Dubrovin *et al.* [11] for example) use a notation such as $\mathbf{E}^n$ for what I am calling $\mathbf{R}^n$, and they write their definitions in terms of the structure of the isometry group. However, in one of the more interesting turns of mathematical fate not treated it this thesis due to time constraints, we can *recover* the isometry group structure from the geometrical/vector space definition.

Subscripts

Whenever possible, I avoid them and write things out using parentheses. Exceptions include a finite enumerated set, a countable enumerated set, and an uncountable set indexed by a real parameter,

and vector spaces associated with points on a manifold (T_pM, $\mathcal{N}_pM$, etc.)

This results in a non-standard notation for connections and derivations.

Volumes and Regions

We denote a volume by a lower case letter, and a given region containing the volume by the corresponding upper case letter. For example, if we are considering a double bubble containing volumes v and w in $\mathbf{S}^3$ we let $u = |\mathbf{S}^3| - v - w$ be the volume of 'U', the complement of the double bubble. The space has then been divided into three regions, U, V and W. Similarly with other compact spaces. For non-compact spaces, we typically refer to the complement of a bubble cluster as *the exterior*.

Note that the space itself does not have to follow the casing rule, e.g. as in this paragraph we will use $\mathbf{S}^3$ to denote the unit radius three-sphere, and $|\mathbf{S}^3|$ to denote its volume.

Finally, we note that if we start with V (an old-style 'U') and then go to W, it is only fitting to have the m^{th} volume be V^m.

Proofs

QEF stands for *quad erat faciendum* which translates as 'what was to be done'. (QED stands for *quad erat demonstratum*, 'what was to be demonstrated'.) Both used to be used, and were even found interspersed in the same document sometimes. One text that used both is Galileo's *Dialogs Concerning Two New Sciences* [15]. I use QEF to mark the end of proofs first as an homage to that book, and, secondly, because I feel that that if the thesis demonstrates anything, it does so as a whole, and certainly not as frequently as after every proof.

Bibliography

[1] William K. Allard. On the first variation of a varifold. *Ann. of Math.(2)*, 95:418–446, 1972.

[2] F. Barthe and B. Maurey. Some remarks on isoperimetry of Gaussian type. *Ann. Inst. H. Poincaré Probab. Statist.*, 36:419–434, 2000.

[3] Roberto Bonola. *Non-Euclidean Geometry, and the Theory of Parallels by Nikolas Lobachevski, with a Supplement Containing The Science of Absolute Space by John Bolyai.* Dover, 1955.

[4] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30:207–216, 1975.

[5] K. Brakke. Surface evolver. http://www.susqu.edu/facstaff/b/brakke/evolver/evolver.html.

[6] K. Brakke. Century-old soap bubble problem solved! *Imagine That!*, 3:1–3, Fall, 1993. Publication of The Geometry Center.

[7] Isaac Chavel. *Eigenvalues in Reimannian Geometry*. Academic Press, 1984.

[8] Isaac Chavel. *Introduction to Riemannian Geometry*. Cambridge University Press, 1993.

[9] Joseph Corneli, Paul Holt, Nicholas Leger, and Eric Schoenfeld. The double bubble conjecture on the flat 2-torus, 2001. Williams College NSF "SMALL" Geometry Group undergraduate research report, `http://lanfiles.williams.edu/~fmorgan/geom.group/torus.ps`.

[10] Andrew Cotton and David Freeman. The double bubble problem in spherical and hyperbolic space. Williams College NSF "SMALL" Geometry Group undergraduate research report, 2000.

[11] Fomenko Dubrovin and Novikov. *Modern Geometry – Methods and Applications: Part I. The Geometry of Surfaces, Transformation Groups, and Fields*. Springer, 1984.

[12] E. B. Vinberg (Ed.). *Geometry II: Spaces of Constant Curvature*. Springer, 1993.

[13] Michele Emmer. Architecture and mathematics: Soap bubbles and soap films. In Kim Williams, editor, *Nexus: Architecture and Mathematics*, pages 53–65. Edizioni dell'Erba, Fucecchio, 1996.

[14] Herbert Federer. *Geometric Measure Theory*. Springer-Verlag, Berlin, 1969.

[15] Galileo Galilei. *Dialogues Concerning Two New Sciences*. Dover, 1954.

[16] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian Geometry*. Springer-Verlag, 1990.

[17] Alfred Gray. *Tubes*. Addison-Wesley, 1990.

[18] Victor Guillemin and Allan Pollack. *Differential Topology*. Prentice-Hall, Inc., 1974.

[19] Thomas Heath. *A History of Greek Mathematics*. Oxford University Press, Oxford, 1960. (Vol. II).

[20] Kenneth Hoffman and Ray Kunza. *Linear Algebra*. Prentice-Hall, Inc, 1965.

[21] Hugh Howards. Soap bubbles on surfaces. Undergraduate thesis, Williams College, Williamstown, MA, 1992.

[22] Michael Hutchings. The structure of area-minimizing double bubbles. *J. Geom. Anal.*, 7:285–304, 1997.

[23] Michael Hutchings, Frank Morgan, Manuel Ritoré, and Antonio Ros. Proof of the double bubble conjecture. *Ann. Math. 155*, pages 459–489, March 2002.

[24] Jeffrey Brock Nickelous Hodges Joel Foisy, Manual Alfaro and Jason Zimba. The standard double soap bubble in r^2 uniquely minimizes perimeter. *Pac. J. Math.*, 159(1):47–49, 1993.

[25] F. J. Almgren Jr. Geometric measure theory and elliptic variational problems. In *Geometric measure theory and minimal surfaces (C.I.M.E., III Ciclo, Varenna, 1972)*, pages 31–117. Edizioni Cremonese, Rome, 1973.

[26] Franz Kafka. *Amerika*. New Directions, 1962.

[27] R. B. Kusner and J. M. Sullivan. Comparing the Weaire-Phelan equal-volume foam to Kelvin's foam. *Forma*, 11:233–242, 1996.

[28] Joseph D. Masters. The double bubble on the flat torus. Unpublished notes on file with F. Morgan, Williams College, 1994.

[29] Patrick McDonald. Differential geometry. Lecture notes on an introductory course given at New College, Sarasota, Florida, Fall 2000.

[30] H. P. McKean. Geometry of differential space. *Ann. Prob.*, 1:197–206, 1973.

[31] Robert Meyers. Average exit time moments of geometric graphs with boundary. New College Thesis, Division of Natural Sciences, 2000.

[32] Frank Morgan. Soap bubbles in $\mathbf{R}^2$ and in sufaces. *Pac. J. Math*, 165(2):347–361, 1994.

[33] Frank Morgan. *Riemannian Geometry: a Beginner's Guide*. A. K. Peters, Ltd., Wellesley, MA, second edition, 1998.

[34] Frank Morgan. *Geometric Measure Theory: a Beginner's Guide*. Academic Press Inc., San Diego, CA, third edition, 2000.

[35] Frank Morgan. Small perimeter-minimizing double bubbles in compact surfaces are standard. *Electronic Proceedings of the 78th annual meeting of the Lousiana/Mississippi Section of the MAA, Univ. of Miss., March 23-24*, 2001. to appear.

[36] Frank Morgan and David L. Johnson. Some sharp isoperimetric theorems for Riemannian manifolds. *Indiana U. Math J.*, 49:1017–1041, 2000.

[37] Frank Morgan and Manuel Ritoré. Geometric measure theory and the proof of the double bubble conjecture. Lecture notes by Ritore on Morgan's lecture series at the Clay Mathematics Institute Summer School on the Global Theory of Minimal Surfaces, at the Mathematical Sciences Research Institute, Berkeley, California 2001 http://www.ugr.es/ ritore/preprints/course.pdf.

[38] Frank Morgan and Wacharin Wichiramala. The standard double bubble is the unique stable double bubble in $\mathbf{R}^2$. *Proc. AMS, to appear.*

[39] Katsumi Nomizu and Shoshichi Kobayashi. *Foundations of Differential Geometry*. Number 15, Volume 1 of Interscience Tracts in Pure and Applied Mathematics. John Wiley and Sons, Inc., New York, 1963.

[40] CRC Press. Handbook of mathematical tables.

[41] Ben W. Reichardt, Cory Heilmann, Yuan Y. Lai, and Anita Spielmann. Proof of the double bubble conjecture in R^4 and certain higher dimensional cases. *Pacific J. Math., to appear.*

[42] Manuel Ritoré. Applications of compactness results for harmonic maps to stable constant mean curvature surfaces. *Math. Z.*, 226:465–481, 1997.

[43] Manuel Ritoré and Antonio Ros. The spaces of index one minimal surfaces and constant mean curvature surfaces embedded in flat three manifolds. *Trans. Amer. Math. Soc.*, 348:391–410, 1996.

[44] Antonio Ros. The isoperimetric problem. Lecture notes on a series of talks given at the Clay Mathematics Institute Summer School on the Global Theory of Minimal Surfaces June 25, 2001 to July 27, 2001, at the Mathematical Sciences Research Institute, Berkeley, California, `http://www.ugr.es/~aros/isoper.pdf`.

[45] Walter Rudin. *Real and Complex Analysis*. McGraw Hill, Inc., New York, second edition, 1974.

[46] E. Schmidt. Beweis der isoperimetrischen Eigenschaft der Kugel im hyperbolischen und sphärischen Raum jeder Dimensionenzahl. *Math. Z.*, 49:1–109, 1943.

[47] H. A. Schwarz. Beweis des Satzes, dass die Kugel kleinere Oberfläche besitzt als jeder andere Körper gleichen Volumens. *Nachrichten Königlichen Gesellschaft Wissenschaften Göttingen*, pages 1–13, 1884.

[48] Leon Simon. *Lectures on Geometric Measure Theory*. Proc. Centre Math. Anal, Australian Nat. U. Vol. 3. Centre for Mathematical Analysis, Australian National University, Australia, 1984.

[49] Michael Spivak. *A Comprehensive Introduction to Differential Geometry (5 vols.)*. Publish or Perish, Inc., 6 Beacon St., Boston Mass., 1970-197?

[50] J. Steiner. Über Maximum und Minimum bei den Figuren in der Ebene, auf der Kugelfläche und im Raume überhaupt. *J. Math. Pure Appl.*, 6:105–170, 1842. Reprinted in Ges. Werke, p. 245–308 (1882).

[51] V. N. Sudakov and B. S. Tsirelson. Extremal properties of half-spaces for sperically invariant measures. *J. Soviet Math.*, 9:9–18, 1978.

[52] J. Lindenstrauss T. Figiel and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta. Math.*, 139:53–94, 1977.

[53] Jean Taylor. The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces. *Ann. Math*, pages 489–539, 1976.

[54] Jean Taylor. The structure of singularities in solutions to ellipsoidal variational problems with constraints in $\mathbf{R}^3$. *Ann. Math*, pages 541–546, 1976.

[55] William P. Thurston. *Three-Dimensional Geometry and Topology, volume 1*. Princeton University Press, 1997.

[56] D. Weaire and R. Phelan. A counter-example to Kelvin's conjecture on minimal surfaces. *Phil. Mag. Lett.*, 69:107–110, 1994.

[57] Denis Weaire and Stefan Hutzler. *The Physics of Foams*. Oxford University Press, Oxford, 2001.

[58] Jeff Weeks. The shape of space. A video production of the Geometry Center.

[59] Joseph A. Wolf. *Spaces of Constant Curvature*. Publish or Perish, 1974.

Technical details.

The final form of this thesis was composed using *xemacs*, LaTeX, *dvips*, and *xmodmap* on a Hewlett-Packard Pavillion 6330 running *Mandrake Linux 8* together with a keyboard from Kinesis Corporation. The line figures were produced using the *pstricks* package for LaTeX, except for the various conjectured minimizers appearing in Figure 7.1 which were drawn by *Surface Evolver*, and for a few which were drawn by hand.

Raytraced pictures of the tiled Euclidean plane and of the sphere appearing in Chapter 1 do so with the permission of their creators (in theory). The screenshot of *Asteroids* was taken with *xv* and edited slightly with the *GIMP*.

The phase diagrams in Chapter 6 were produced with *Mathematica*. The phase diagram in Chapter 7 was produced using *Surface Evolver*, *perl*, *GNUPlot*, and the *GIMP*. Computing the data for this phase plot in *Surface Evolver* took about 8 hours on 26 Dell PCs running in parallel.

Typeface. The text of this thesis is set in the "Charter" font available in standard LaTeX. Not only is this font nice and readible, but its name makes it fit nicely into one of the major themes of the thesis.