A. N. Whitehead's Geometric Algebra

Stephen Blake
stebla@ntlworld.com

February 19, 2005

Contents

List of Figures

Preface

The motivation for this work grew out of the writer's dissatisfaction with the vector algebra that is taught to students of physics and engineering[1]. The unsatisfying aspect of vector analysis is that it attempts to be a *geometric algebra* that enables one to solve geometric problems in a coordinate-free way, yet it falls short of this goal by being unable to express many of the notions that are needed by a geometric algebra. For example, the vector language does not have a syntax for the join and intersection of linear subspaces. Consequently, it is not straightforward to write down the point at the intersection of a line and a plane in 3-d space. Another deficiency is that it is asymmetric with respect to translations and rotations. Translations are easily represented in a coordinate-free way, but there is no syntax for rotations. In order to represent a rotation, one has to introduce a coordinate system and use matrices.

Similar concerns about the deficiencies of vector algebra have been expressed elsewhere [2, 3, 4]. These authors all suggest, in various ways, that a geometric algebra should be based on a Grassmann algebra [7] of vectors. However, it seems wrong to choose vectors as the basic elements of a geometric algebra because a vector is something which naturally arises in the context of an affine space - a space with a notion of parallelism. In contrast, the current work argues that the Grassmann algebra should be based on points and hyperplanes instead of vectors. This follows from the nature of geometry as expressed by Félix Klein's *Erlangen Programme* [8, 9]. Vectors are displacements. As already noted, they arise naturally in an affine space. According to Klein's genealogy of geometries, which is shown overleaf, all the geometries in which space is homogeneous are constructed from projective geometry by specifying certain structures to be absolute (invariant). Affine geometry is not fundamental from the Kleinian point of view. Instead projective geometry is fundamental, and the elements of projective geometry are points and, dually, hyperplanes.

[1]See reference [1] for a standard account.

```
                         projective
            __________________|_________________
           |                  |                 |
        affine            elliptic          hyperbolic
      _____|______
     |            |
 Euclidean    Minkowskian
```

Interestingly, in 1898, A. N. Whitehead published [10] an exposition of a geometric algebra built from a Grassmann algebra of points instead of vectors. Whitehead's algebra of points naturally represents projective space in accord with Klein's genealogy of geometries. However, in Whitehead's book, points have a privileged status compared to hyperplanes. This is contrary to the *principle of duality* [9] which holds in projective geometry. According to which, any theorem involving points can be turned into a theorem involving hyperplanes if the points are replaced by hyperplanes and a few words altered in the statement of the theorem. The privileged status of points in Whitehead's theory complicates matters. He finds two products which he names *progressive* and *regressive*. However, when points and hyperplanes are put on an equal footing according to the principle of duality, there is only one product and the theory has an appealing simplicity.

Whitehead's algebra is related to the *double algebra* expounded by Barnabei, Brini and Rota in reference [5]. The double algebra has also been advocated by Faugeras and Luong in chapter 3 of the book [6] where it is also called *Grassmann-Cayley algebra*. The exposition of the double algebra by Faugeras and Luong closely follows the treatment given by Barnabei et al. Consequently, in order to point out how Whitehead's algebra differs from the double algebra, it is sufficient to consider the double algebra as described in the paper of Barnabei et al. The kernel of the difference is that in the introduction to their paper [5], they write:

> Whitehead, however, did not take the decisive step of clearly distinguishing the two kinds of products, let alone axiomatizing them.

The two products referred to are the progressive and the regressive. Whitehead used the same notation for both products because he clearly felt that there should be only one product in the algebra. The current text agrees wholeheartedly with Whitehead - and Grassmann in [7] - that there is only one product in the algebra; this single product is called the *antisymmetric product* in the current text. The use of two products in [5] destroys the beauty of the theory which comes from the central rôle played by the principle of duality. In fact, the double algebra is not really an algebra of projective geometry because its group is the special linear group - the subgroup of the linear group with unity determinant; according to the Erlangen Programme, the group of projective geometry is the general linear group.

The text sets out to be an exposition of Whitehead's algebra[2], incorporating the writer's simplifications and developments. However, it has turned out to be a schizophrenic text. On one hand it is a text on 3-d Euclidean computational geometry intended to be used in engineering applications. On the other hand, the methods of Whitehead's algebra enable us to readily deal with Euclidean and non-Euclidean spaces of any dimension, so it is natural hat applications to the physics of space-time continually present themselves, inviting study. Rather than supressing one of these, perhaps discordant, aspects, we choose to live with them. The reader who is only interested in engineering applications can simply omit anything to do with space-time. The writer makes his living building computer vision systems, and the motivation for much of this work has been to produce a language for computational geometry and vision to make it easier to program applications. So, in addition to theoretical material, the text describes a function package, written for the scientific interpreter *Yorick* [12], which implements Whitehead's algebra as a practical tool for programming applications in geometry and vision. In the later chapters, the function package is used to illustrate the theoretical material with numerical examples. The following paragraphs briefly describe the contents of each chapter, highlighting the interesting results.

Chapter 1 sets out a geometric algebra of $(n-1)$-d projective space along the lines suggested by Whitehead, but modified to place the principle of duality in the centre of the stage. The most important result is that Whitehead's regressive and progressive products are replaced by a single antisymmetric product through the use of duality.

Chapter 2 shows how the theory of chapter 1 can be used to make a function package for computational geometry on the projective plane. The aims are to consolidate the theory from chapter 1 for the particular case $n = 3$, and to introduce the function package in this simple case. Chapter 2 also motivates the introduction of the important notion of polarity which is then taken further in chapter 3.

Chapter 3 describes the theory of congruences in Whitehead's algebra. This chapter provides the theoretical background for the study of elliptic, hyperbolic, Euclidean and Minkowski spaces. These spaces, unlike projective space and affine space, admit a notion of distance. The key result of this

[2]The geometer H. G. Forder also wrote a book [27] on Whitehead's algebra. However, Forder's work does not go beyond the material in Whitehead's original treatise [10].

chapter is a closed formula (3.26) for a simple congruence in an $(n-1)$-d space with an arbitrary polarity.

Chapter 4 considers calculus in Whitehead's algebra. This chapter introduces the methods that are needed to deal with hypersurfaces, parallel transport and curvature which are considered in more depth in chapter 9.

Chapter 5 shows how the theory of chapter 3 can be applied to the study of space-time. This chapter considers the problem of the non-appearance of dilatations in our world and concludes that elliptic de Sitter space is the most natural space-time from the point of view of Whitehead's algebra.

Chapter 6 is a digression to prove Müller's theorems which are needed for some derivations in the subsequent chapters.

Chapter 7 gives the theory of $(n-2)$-d Euclidean geometry modelled on a quadric hypersurface in $(n-1)$-d elliptic space. In addition, the Yorick package for 3-d Euclidean geometry is described and used to illustrate the theoretical material with numerical examples. The power of Whitehead's algebra is demonstrated by closed formulae for the centre (7.28) and radius (7.33) of the subspace formed by the intersection of r hyperpsheres in $(n-2)$-d Euclidean space. Also, a proof is given of Soddy's generalised formula concerning mutually touching hyperspheres in Euclidean space.

Chapter 8 shows how camera calibration of a computer vision system should be carried out using Whitehead's algebra.

Chapter 9 gives the theory of an arbitrary smooth hypersurface embedded in a higher-dimensional Euclidean space. However, the theory works for hypersurfaces embedded in a general flat space. The Euclidean (or flat) space is modelled on the absolute quadric of a bulk elliptic space as in chapter 7. The normal to the hypersurface plays a prominent rôle and the theory is similar to classical differential geometry of 2-d surfaces. However, the Gauss map is invertible and this means that we can replace the physical hypersurface embedded in the Euclidean (or flat) space with a hypersurface in the bulk elliptic space. It turns out that curvature looks much simpler when studied in the bulk elliptic space. Using this result, we are able to show how the theory relates to the other approach to differential geometry - Riemannian geometry - which is the mathematical arena for general relativity.

Chapter 10 is an application of the hypersurface theory of chapter 9 to space-time. It sets up a theory of the free gravitational field (i.e. a universe without any matter) by assuming that the physical space-time is a curved 4-d hypersurface embedded in a flat 5-d Minkowski space. Gravity is then a manifestation of the curvature of the space-time just as in general relativity. The theory is an attempt to see what general relativity would look like if Einstein had been a student of the Cayley-Klein school of geometry instead of the Riemannian school as happened. Interestingly, the solution of the equations of empty space-time turns out to be de Sitter space. Furthermore, contrary to the situation in general relativity, this solution appears without a cosmological constant having been put in by hand.

Appendix A describes a function package that implements Whitehead's algebra for general $(n-1)$-d space using the scientific interpreter *Scilab* [26]. The description of the Scilab package is relegated to the appendix because, unlike the Yorick package, it is too slow to be of practical value in engineering applications. However, since it works in higher dimensions, whilst the number of dimensions is fixed in the Yorick package, the writer felt it to be a useful tool deserving a place in this text. Consequently, some illustrative computations on hyperspheres in 4-d Euclidean space are performed with the Scilab package.

Appendix B covers the same material as chapter 1 of the main text. Chapter 1 develops the theory heuristically, guided by geometrical intuition. Appendix B is an attempt to develop the theory in a more formal way.

This is effectively the second edition of the text. Having put the first edition on my site [21] in November 2004, I received penetrating questions by email from George Craig and Issac Trotts about the algebra of projective geometry in chapter 1. As a result, I have rewritten chapter 1 in an attempt to remove some sloppy thinking and to make it clearer. I have also added appendix B which attempts to cover the same material as chapter 1 in an axiomatic way. I also received encouraging comments about the text by email from Danny Ross-Lunsford, and so for this second edition I thought it worthwhile to add an index.

S.B.
February 19, 2005

Chapter 1

An Algebra of Projective Geometry

1.1 Notational Conventions

This work adopts two notational conventions in order to reduce the number of brackets appearing in the formulae, thus making them easier to read.

A dot (.) notation is used to indicate the position of an application of the antisymmetric product which is the only product which appears in Whitehead's algebra as formulated in this text. Physically, the antisymmetric product builds and intersects linear subspaces. For example, in a 3-dimensional space, suppose we have two points a and b. The line through these points is ab where the points are multiplied together by the antisymmetric product. Similarly, a plane through the three points p, q and r is given by the product pqr. These two examples show how linear subspaces are built up using the antisymmetric product. The intersection of the line and the plane is the point $(ab)(pqr)$ which is written in the dot notation as $ab.pqr$. The dot can be is used whenever the product of some elements is not associative. In pqr, the products are associative so that $p(qr)$ and $(pq)r$ are equal and there is no ambiguity in writing pqr. However, it can happen that $(ab)(pqr)$ is not associative because it is not the same as $(abp)(qr)$. The dot resolves the ambiguity by indicating the position of the final application of the antisymmetric product. So, $(ab)(pqr) = ab.pqr$ and $(abp)(qr) = abp.qr$. If elements of the algebra are adjacent to one another, and are not separated by a dot or a bracket, then the elements should multiply associatively. The dot should not be confused with the use of a dot to denote an inner product, such as occurs in vector algebra. There is only one product in Whitehead's algebra and this is the antisymmetric product.

The other bit of notation concerns functions or operators. Suppose that T maps an element X to some other element $T(X)$. We often think of T as an operator acting on X by writing $T(X)$ as TX. The operator convention adopted is that $TXYZ = T(XYZ)$ and XYZ must, by the first convention, multiply associatively. The operator acts on the element to its right which multiplies associatively.

The dot and operator conventions still need to be augmented with brackets, but the formula are made easier to read as the following examples show. In these examples, p, L, M are elements of the algebra and I is an operator.

$$\begin{aligned} ILM.LM &= (I(LM))(LM) \\ (p.ILM).LM &= (pI(LM))(LM) \\ ILM.(p.LM) &= I(LM)(p(LM)) \end{aligned}$$

1.2 Points

Our arena is a set of points $a, b, c, \ldots, x, y, z$. The basic idea is that new points are linear combinations of given points. Consider two reference points a_1, a_2. A point p on the line through a_1 and a_2 is the linear combination,

$$p = \xi_1 a_1 + \xi_2 a_2$$

with ξ_1 and ξ_2 numbers. The weights ξ_1, ξ_2 draw the point towards reference points a_1 or a_2 according to the relative strength of the weights. If we set $\xi_2 = 0$ we get $p = \xi_1 a_1$ and this holds for any weight ξ_1. Therefore, a_1 and $\xi_1 a_1$ represent the same point.

Suppose we have two observers Alice and Bob. Alice sees a point $p = \xi_1 a_1 + \xi_2 a_2$. Let's have a transformation $f(.)$ that takes points into points as the rule for transforming from Alice's to Bob's viewpoint. The situation is shown below where the transformation f is physically the motion of Alice with respect to Bob.

$$\begin{array}{ccc} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ fp & & p \end{array}$$

Now, if Alice sees a_1, Bob sees this point as $f(a_1) = b_1$. Also, if Alice sees a_2, Bob sees $f(a_2) = b_2$. Consider the point $a_1 + a_2$ seen by Alice. Bob sees $f(a_1 + a_2)$. However, Alice could also transform the reference points and let Bob form the linear combination. We assume that the operation "+" is a scalar operation that does not change. Then Bob sees $f(a_1) + f(a_2)$. Hence we must have $f(a_1 + a_2) = f(a_1) + f(a_2)$. Now consider

$f(2a_1) = f(a_1 + a_1) = f(a_1) + f(a_1) = 2f(a_1)$. By continuity we have $f(\xi a) = \xi f(a)$. Therefore $f(.)$ must be a linear transformation, $f(\xi_1 a_1 + \xi_2 a_2) = \xi_1 f(a_1) + \xi_2 f(a_2)$. Thus, the transformations between observers are linear maps. According to Klein's Erlangen Programme [8, 9] , this means that Alice and Bob inhabit projective space and so our algebra of points is naturally an algebra of projective geometry[1]. Incidently, we see that the weighting numbers transform as scalar numbers.

Generalizing, let $a_1, a_2, \ldots, a_n$ be n reference points. An arbitrary point in this $(n-1)$-dimensional projective space is p and,

$$p = \sum_{i=1}^{n} \xi_i a_i \tag{1.1}$$

where the ξ_i are numerical weights.

1.3 The Antisymmetric Product

We now assume that points can be multiplied together to produce lines, planes and higher dimensional linear subspaces. So, the line through the points a_1 and a_2 is $a_1 a_2$ and the plane through the points a_1, a_2, a_3 is $a_1 a_2 a_3$. We assume that the products represent complete subspaces, so, for example, $a_1 a_2$ is the entire line drawn through the two points, and not just the segment which terminates on the two points. In general, the element $X = a_1 \ldots a_r$ represents the $(r-1)$-dimensional linear subspace through the r points.

We have already seen in section 1.2 that a_1 and ξa_1 represent the same point. Here ξ is a numerical weight. The point a_1 may be thought of as a linear subspace of dimension zero. So, if an element X of the algebra is to represent a linear subspace, then this same linear subspace must also be represented by ξX where ξ is a numerical weight.

Now consider the line $l = a_1 a_2$. A point on the line is $p = a_1 + a_2$ so that $p a_2$ must also represent the line l. Since the algebraic elements $a_1 a_2$ and $p a_2$ represent the same line (linear subspace), they must be equal up to some weight ξ. Therefore, we have the equation,

$$p a_2 = (a_1 + a_2) a_2 = \xi a_1 a_2$$

and assuming the product is distributive, we can write,

$$p a_2 = (a_1 + a_2) a_2 = a_1 a_2 + a_2 a_2 = \xi a_1 a_2$$

[1] Klein's Erlangen Programme says that each type of geometry is characterised by the group of transformations which connect equivalent observers. In this case, each transformation f is a member of the general linear group and this group, by definition, specifies the geometry as that of projective space.

and then we must have $\xi = 1$ and $a_2 a_2 = 0$. Now a_2 is an arbitrary reference point, so that, in general, the product of a point with itself must vanish[2]. Consider the vanishing product $(a_1 + a_2)(a_1 + a_2) = 0$. Expanding,

$$0 = (a_1+a_2)(a_1+a_2) = a_1a_1+a_1a_2+a_2a_1+a_2a_2 = 0+a_1a_2+a_2a_1+0 = a_1a_2+a_2a_1$$

which shows that the product is antisymmetric,

$$a_1 a_2 = -a_2 a_1 \ . \tag{1.2}$$

1.4 Grassmann's Algebra of Points

Let us summarise what we know so far. We have a set of reference points $a_1, \ldots, a_n$ for an $(n-1)$-dimensional projective space. We can multiply the reference points to make algebraic elements such as $X = a_1 \ldots a_r$ which represents the $(r-1)$-dimensional linear subspace through the r points. The product of the points is antisymmetric,

$$a_i a_j = -a_j a_i \text{ and } a_i a_i = 0 \ . \tag{1.3}$$

Let us take the case $n = 4$ as an example and list the linear subspaces formed by taking products of the reference points. There are 4 reference points a_1, a_2, a_3, a_4 for the 3-dimensional projective space. Next we can make lines of the form $a_i a_j$. In general there are 16 elements of grade 2 such as $a_1a_1, a_1a_2, \ldots, a_4a_4$. However, the antisymmetry (1.3) means that, for example, the elements $a_1 a_2$ and $a_2 a_1$ are not independent because $a_2 a_1 = -a_1 a_2$ and an element such as $a_1 a_1 = 0$. So, antisymmetry means that for $n = 4$ we only get 6 independent elements a_1a_2, a_1a_3, a_1a_4, a_2a_3, a_2a_4, a_3a_4 of grade 2 (lines).

Continuing the $n = 4$ example, the elements of grade 3 of the form $a_i a_j a_k$ are planes. Ordinarily there would be 64 of these elements but antisymmetry means that only the 4 elements $a_1a_2a_3$, $a_1a_2a_4$, $a_1a_3a_4$, $a_2a_3a_4$ are independent. For example,

$$a_4a_3a_1 = -a_4a_1a_3 = a_1a_4a_3 = -a_1a_3a_4 \ .$$

Finally, antisymmetry means that for $n = 4$, there is only 1 independent element $a_1a_2a_3a_4$ of grade 4. This element represents the entire 3-dimensional projective space.

[2]The fact that $a_1 a_1 = 0$ should alert us to the fact that certain combinations of elements can evaluate to a number, or more generally, a weight. The other case of this behaviour is explained in section 1.6.

So, for $n = 4$, the reference points with their antisymmetric product have generated an algebra with 15 basis elements (linear subspaces). These are the 4 basis elements a_1, a_2, a_3, a_4 of grade 1 (points), the 6 basis elements a_1a_2, a_1a_3, a_1a_4, a_2a_3, a_2a_4, a_3a_4 of grade 2 (lines), the 4 basis elements $a_1a_2a_3$, $a_1a_2a_4$, $a_1a_3a_4$, $a_2a_3a_4$ of grade 3 (planes) and a single basis element $a_1a_2a_3a_4$ of grade 4 (the entire 3-dimensional projective space). This algebra is called a *Grassmann algebra.*

Having seen how the linear subspaces form a basis for a Grassmann algebra when $n = 4$, let us now study the general case of $(n-1)$-dimensional projective space. The reference points are $a_1, \ldots, a_n$. The basis elements are of the form, $a_i, a_ia_j, a_ia_ja_k, \ldots$. Antisymmetry means that a basis element of the algebra cannot contain any repeated reference points; the reference points in a basis element must be all different. Furthermore, a permutation in the order of the reference points making up a basis element can do no more than just change the sign of the element. For, suppose we have a basis element $a_1 \ldots a_r$ of grade $r \leq n$. Let σ be a permutation of the indices $1, 2, \ldots, r$ as $\sigma(1), \sigma(2), \ldots, \sigma(r)$ then it follows from (1.3) that,

$$a_{\sigma(1)} \ldots a_{\sigma(r)} = \mathrm{sgn}(\sigma) a_1 a_2 \ldots a_r \ . \tag{1.4}$$

In other words, a permutation of the order of the reference points making up a basis element does not produce an independent basis element. The number of independent basis elements of grade r is just the number ${}^nC_r$ of ways of picking r reference points, all different, without regard to order, from the complete set of n reference points $a_1, a_2, \ldots, a_n$. The total number of independent basis elements for the Grassmann algebra is,

$$\sum_{r=1}^{n} ({}^nC_r) = 2^n - 1 \ .$$

The sequence of basis elements terminates after the single element $a_1 \ldots a_n$ of grade n which represents the entire $(n-1)$-dimensional projective space.

We can write down a typical member of the ${}^nC_r$ basis elements of grade r as follows. Pick a set of r labels $1 \leq j_1 < j_2 < \ldots < j_r \leq n$ from the n labels $1, 2, \ldots, n$. The basis element with these labels is, $a_{j_1} a_{j_2} \ldots a_{j_r}$. General elements of the algebra can be made by forming linear combinations of the basis elements. The general element of grade r is the linear combination,

$$X = \sum_{1 \leq j_1 \ldots j_r \leq n} \xi_{j_1 \ldots j_r} a_{j_1} \ldots a_{j_r} \tag{1.5}$$

where the $\xi_{j_1 \ldots j_r}$ are weights.

1.5 Transformation of the Algebraic Elements

We now examine the way in which the basis elements (linear subspaces) transform. The antisymmetric product is assumed to be a scalar operation that does not change.

$$\begin{array}{rcl} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ fa_1 & & a_1 \\ fa_2 & & a_2 \\ f(a_1)f(a_2) & & a_1a_2 \end{array}$$

Thus, as the above diagram shows, if Alice sees the product a_1a_2 then Bob sees this as $f(a_1)f(a_2)$. For consistency, Alice can also transform a_1a_2 directly as $f(a_1a_2)$ so that $f(a_1a_2) = f(a_1)f(a_2)$ and in general this gives the action of f on higher grade products as,

$$f(a_1a_2 \dots a_r) = f(a_1)f(a_2) \dots f(a_r) \tag{1.6}$$

which is known as the *outermorphism* property. We have already seen in section 1.2 that a transformation f between observers is linear. Therefore, the transformation of an arbitrary element of the algebra follows from that of the basis elements by linearity. The diagram,

$$\begin{array}{rcl} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ f(a_1 \dots a_r) & & a_1 \dots a_r \\ = fa_1 \dots fa_r & & \\ \alpha fa_{j_1} \dots fa_{j_r} + \beta fa_{k_1} \dots fa_{k_r} & & \alpha a_{j_1} \dots a_{j_r} + \beta a_{k_1} \dots a_{k_r} \end{array}$$

summarises how the elements transform. In the diagram, α and β are numerical weights. They transform as scalars.

Notice that a transformation f (element of the linear group) always maps elements of grade r into elements of grade r. Hence a subspace of grade r in the Grassmann algebra is an invariant subspace under the action of the transformation group. Tentatively, this fact seems to imply that there will never be a use for elements of mixed grade in the algebra. In other words, elements of the Grassmann algebra like $2a_1 + 3a_2a_3$ will never appear. A corollary of this observation is that all the terms that are summed to form an element should transform in the same way. For example, a valid element is $2a_1a_2 + 3a_2a_3$.

1.6 Pseudonumbers

The basis element $a_1 \dots a_n$ represents the entire $(n-1)$-dimensional projective space. Since there is only one basis element $a_1 \dots a_n$ of grade n, it follows from (1.5) that all elements of grade n are proportional to the single basis element. So, if X is an element of grade n, then,

$$X = \xi a_1 \dots a_n \tag{1.7}$$

where ξ is a numerical weight. It follows that the grade n elements have similarities to numbers because an arbitrary number χ is proportional to some basis number α as $\chi = \xi\alpha$ and so $\xi = \chi/\alpha$. In the same way, the notion of formal division for elements of grade n makes sense since (1.7) implies,

$$\xi = \frac{X}{a_1 \dots a_n} \,. \tag{1.8}$$

It is tempting to identify the basis element $a_1 \dots a_n$ with a definite number. However, this does not work because the transformation $f(a_1 \dots a_n)$ must be another element of grade n and so it must be proportional to $a_1 \dots a_n$. The numerical coefficient of proportionality must depend on the transformation f. It is $\det f^{-1}$. In fact, this is how the determinant of a transformation is defined in Whitehead's algebra[3].

$$f(a_1 \dots a_n) = fa_1 \dots fa_n = (\det f^{-1}) a_1 \dots a_n \tag{1.9}$$

The following diagram illustrates the difference between numbers which transform as scalars and the elements of grade n which get multiplied by the determinant.

$$\begin{array}{rcl} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ f(\xi) & & \xi \\ = \xi & & \\ f(a_1 \dots a_n) & & a_1 \dots a_n \\ = (\det f^{-1}) a_1 \dots a_n & & \end{array}$$

In the light of this difference, the elements of grade n are named *pseudonumbers*[4].

[3] The coefficient of proportionality is shown as the determinant of the inverse transformation f^{-1} and $\det(f^{-1}) = 1/\det(f)$. Whether or not the determinant or its inverse is used is just a matter of definition. As shown, equation (1.9) agrees with the definition of the matrix elements of f in section 3.11.1. Also see equation (4.47).

[4] The double algebra [5] avoids the issue of pseudonumbers by restricting the transformations to those which have unity determinant. Hence, the transformations of the double algebra are those of the special linear group.

Up to now, the weights multiplying the basis elements of the Grassmann algebra have been numbers. However, from now onwards, we allow the weights to be numbers or pseudonumbers. So, for example, if we have a grade-r basis element $a_1 \ldots a_r$, we can multiply it by a numerical weight ξ to form the element $\xi a_1 \ldots a_r$, and also by the pseudonumber weight[5] $\xi a_1 \ldots a_n$ to form the element $(\xi a_1 \ldots a_n) a_1 \ldots a_r$. The following diagram shows the difference between the way a basis element $X = a_1 \ldots a_r$ transforms when it is weighted by a number ξ as in ξX and when it is weighted by a pseudonumber $\omega = \xi a_1 \ldots a_n$ as in ωX.

$$
\begin{array}{rcl}
\text{Bob} & \xrightarrow{f} & \text{Alice} \\
f(\xi X) & & \xi X \\
= \xi f X & & \\
f(\omega X) & & \omega X \\
= (\det f^{-1}) \omega f X & &
\end{array}
$$

Now, the corollary mentioned at the end of section 1.5 argued that all the terms that are summed together to form an element of the algebra should all transform in the same way. It therefore follows, in the light of the above diagram showing that basis elements weighted by numbers and pseudonumbers transform differently, that in forming any element, the terms will always be weighted entirely with numbers or entirely with pseudonumbers and never with mixtures of the two types of weights.

Consider two basis elements X_1 and X_2 of the same grade. Form the linear combination in which these elements are weighted by pseudonumbers ω_1 and ω_2. The linear combination is the element $Y = \omega_1 X_1 + \omega_2 X_2$. Each pseudonumber is proportional to the element of grade n. So, $\omega_i = \xi_i a_1 \ldots a_n$ where the numbers ξ_i are the coefficients of proportionality. The element of grade n can be taken out as a common factor so that,

$$Y = \omega_1 X_1 + \omega_2 X_2 = (a_1 \ldots a_n)(\xi_1 X_1 + \xi_2 X_2) \; . \tag{1.10}$$

Now we know that two algebraic elements represent the same linear subspace if they only differ by some overall weight factor. Therefore, the linear subspace represented by equation (1.10) can be equally well represented by

[5] The observation that the grade n basis element $a_1 \ldots a_n$ evaluates to a weight is equivalent to saying that the basis elements are generated by products of anti-commuting reference points modulo the element $a_1 \ldots a_n$. For, as we generate the basis elements in the sequence $a_1, a_1 a_2, \ldots$ we eventually get to $a_1 \ldots a_n$. The attempt to generate a basis element of grade greater than n simply results in the cycle of basis elements being re-created. For example, $(a_1 \ldots a_n) a_1$ is proportional to a_1.

$\xi_1 X_1 + \xi_2 X_2$. This expression has numerical weights instead of pseudonumber weights. Therefore, in practical calculations, one can often ignore the distinction between numbers and pseudonumbers.

1.7 Partial Definition of the Product

An algebra always has a definition of the product XY of any pair of elements X and Y. However, we only need to define the product for basis elements because we can immediately extend it to apply to linear combinations of basis elements by the distributive rule,

$$(\xi X + \eta Y)Z = \xi(XY) + \eta(XZ) \tag{1.11}$$

where X, Y, Z are basis elements and ξ, η are weights.

In our case we have the antisymmetric product of points given by (1.3). Hitherto, this equation has been used to define the basis elements of the algebra, but it is not clear whether or not it determines the rule for the product XY for all X and Y.

Consider a basis element $a_1 \dots a_r a_{r+1} \dots a_{r+s}$ where $r + s \leq n$. The fact that we write such a basis element as a product of points without any brackets implies that the points associate freely. Hence, we can choose to evaluate the basis element as $(a_1 \dots a_r)(a_{r+1} \dots a_{r+s})$. So, for $X = a_1 \dots a_r$ and $Y = a_{r+1} \dots a_{r+s}$ with $r + s \leq n$ the product is,

$$XY = (a_1 \dots a_r)(a_{r+1} \dots a_{r+s}) = a_1 \dots a_r a_{r+1} \dots a_{r+s}\ . \tag{1.12}$$

Equation (1.12) can be extended by the distributive rule (1.11) to evaluate XY when X and Y are general elements of grades r and s respectively and $r + s \leq n$. By using the antisymmetry (1.3) we obtain the following useful formula that allows the order of factors in a product to be flipped in the case where $r + s \leq n$.

$$\begin{aligned} YX &= (a_{r+1} \dots a_{r+s})(a_1 \dots a_r) = a_{r+1} \dots a_{r+s} a_1 \dots a_r \\ &= (-1)^s a_1 a_{r+1} \dots a_{r+s} a_2 \dots a_r = (-1)^{rs} a_1 \dots a_r a_{r+1} \dots a_{r+s} \\ &= (-1)^{rs} XY \end{aligned} \tag{1.13}$$

Equation (1.12) only defines the product when the sum of the grades of X and Y do not exceed n. We are still left with the problem of what happens when the sum of the grades exceeds n? The answer is already known to us when the grade of X or the grade of Y is exactly n. For, suppose $X = a_1 \dots a_n$. Then X is a pseudonumber and by the argument of section 1.6 it evaluates to a weight and so the result is simply proportional to Y.

$$XY = (a_1 \ldots a_n)Y = Y(a_1 \ldots a_n) = YX \tag{1.14}$$

The remaining case of XY is when the sum of the grades of X and Y exceed n and neither element is of grade n. In order to appreciate the problem that this case causes, let us take $n = 4$ so that we are in 3-dimensional projective space. We consider the multiplication of the line $X = a_1a_2$ by the plane $Y = a_2a_3a_4$. The product is $XY = (a_1a_2)(a_2a_3a_4)$. Let's assume that the points are freely associative. This worked in equation (1.12). Firstly, we could try the association $XY = a_1(a_2a_2a_3a_4) = 0$. The result is zero by antisymmetry because the factor a_2 occurs twice in the factor in brackets. Alternatively, we could subject X to a permutation to give $X = -a_2a_1$ and associate as before. The result is $XY = -a_2(a_1a_2a_3a_4)$ which is proportional to a_2 since the factor in brackets is now a pseudonumber weight. These two results are mutually contradictory, so it is clear that the points do not freely associate in this case[6].

This is as far as we can go at this stage in the theory. The antisymmetric product of a pair of points (1.3) does not completely determine the algebraic product XY for all elements X, Y in the algebra. It only determines the product XY when the sum of the grades of X and Y do not exceed n.

1.8 Hyperplanes and Coordinates for Points

From (1.1), an arbitrary point can be expanded in terms of the reference points as,

$$p = \sum_{i=1}^{n} \xi_i a_i \tag{1.15}$$

[6] In his book [10], Whitehead's treatment has diverged from our own by this point. He assumes that products of points always freely associate. In our example of the product of the line and the plane, he would say that $XY = 0$ by antisymmetry. According to Whitehead, the product $XY = 0$ whenever the sum of the grades of X and Y exceeds n. However, this means his theory is rather impoverished because the product zeroes half of the things he would like to multiply. He then has to invent a second product to handle XY whenever the sum of the grades of X and Y exceeds n. He calls the two products *progressive* and *regressive* respectively in §98. However, at the end of §98, he assumes that pseudonumbers are the same as ordinary numbers and sets $a_1 \ldots a_n = 1$ in our notation. He ignores the fact that this assumption produces the contradiction that we have just found with the free association of the points in his progressive product. However, he seems to acknowledge that something is amiss, because in §98 he stipulates that our pseudonumbers are always to be enclosed in brackets.

where the ξ_i are numerical weights. We can find the weight ξ_j in (1.15) by multiplying both sides by $a_1 \dots \check{a}_j \dots a_n$ where the "ˇ" denotes a missing factor. By antisymmetry, this kills all the terms on the rhs of (1.15) except for the term $\xi_j a_j$ and we get,

$$pa_1 \dots \check{a}_j \dots a_n = \xi_j a_j a_1 \dots \check{a}_j \dots a_n = \xi_j (-1)^{j-1} a_1 \dots a_n \tag{1.16}$$

where the sign in the final expression appears because of a use of (1.4) or antisymmetry (1.3) to re-order the reference points in the basis element. Equation (1.16) is a relation between pseudonumbers. We can formally get the weight by an application of (1.8) to obtain[7],

$$\xi_j = \frac{(-1)^{j-1} pa_1 \dots \check{a}_j \dots a_n}{a_1 \dots a_n} \ . \tag{1.17}$$

From (1.17), let's define the *reference hyperplanes*,

$$A_j = \frac{(-1)^{j-1} a_1 \dots \check{a}_j \dots a_n}{a_1 \dots a_n} \ . \tag{1.18}$$

These n reference hyperplanes are dual to the n reference points because,

$$a_i A_j = \frac{(-1)^{j-1} a_i a_1 \dots \check{a}_j \dots a_n}{a_1 \dots a_n} = \frac{\delta_{ij} a_1 \dots a_n}{a_1 \dots a_n} = \delta_{ij} \tag{1.19}$$

where δ_{ij} is Kronecker's delta. The numerical weights are now $\xi_i = pA_i$.

As an example, consider the case $n = 3$. This is the projective plane with reference points a_1, a_2, a_3. In the plane, the hyperplanes are lines. Using equation (1.18) the dual lines are,

$$A_1 = \frac{a_2 a_3}{a_1 a_2 a_3} \ , \ A_2 = \frac{a_3 a_1}{a_1 a_2 a_3} \ , \ A_3 = \frac{a_1 a_2}{a_1 a_2 a_3} \ .$$

---

[7]If we substitute the weights (1.17) back into (1.15) we obtain the following identity.

$$p = \sum_{i=1}^{n} \frac{(-1)^{i-1} (pa_1 \dots \check{a}_i \dots a_n) a_i}{a_1 \dots a_n}$$

The numerator of each term in the above summation contains a product of the grade n element $(pa_1 \dots \check{a}_i \dots a_n)$ by the point a_i. Notice that the reference points must multiply modulo $a_1 \dots a_n$ as observed in the earlier footnote[5] and stated explicitly in equation (1.14), otherwise this product would be zero.

1.9 Dual Basis Elements

The principle of duality puts points and hyperplanes on the same footing in projective geometry. Therefore, instead of generating the basis elements of the Grassmann algebra from products of anti-commuting reference points, modulo $a_1 \dots a_n$, as we have done up to now, we can equally well repeat the theory using the reference hyperplanes in place of the reference points. In other words, the reference hyperplanes $A_1, A_2, \dots, A_n$ can also be taken as a set of generators, modulo $A_1 \dots A_n$ for the Grassmann algebra. These generators anti-commute,

$$A_i A_j = -A_j A_i \text{ and } (A_i)^2 = 0 \text{ for all } i,j. \tag{1.20}$$

and form a basis $A_i, A_i A_j, A_i A_j A_k, \dots$ for the Grassmann algebra. Notice that the reference points and the dual reference hyperplanes generate the same Grassmann algebra by virtue of equation (1.18) which shows that the A_i are elements of the Grassmann algebra generated by the a_i. This is contrary to the usual situation in mathematics where elements and their duals occupy different spaces. For example, in the theory of differential forms[11], vectors and 1-forms live in different spaces. Similarly, in Hilbert space, bras and kets live in different spaces.

The relation between the basis elements generated by products of hyperplanes and the basis elements generated by products of points can be found as follows. Consider the identity,

$$1 = \frac{a_1 \dots a_n}{a_1 \dots a_n} = \frac{(a_1 \dots a_r)(a_{r+1} \dots a_n)}{a_1 \dots a_n} \tag{1.21}$$

and define the basis element of grade r generated by a product of hyperplanes as,

$$A_1 \dots A_r = \frac{a_{r+1} \dots a_n}{a_1 \dots a_n} \ . \tag{1.22}$$

so that from (1.21),

$$(a_1 \dots a_r)(A_1 \dots A_r) = 1 \ . \tag{1.23}$$

Setting $r = n$ in (1.23) shows that the pseudonumbers are reciprocal,

$$A_1 \dots A_n = \frac{1}{a_1 \dots a_n} \ . \tag{1.24}$$

Substituting equation (1.24) into equation (1.22) gives the dual formula,

$$a_{r+1} \dots a_n = \frac{A_1 \dots A_r}{A_1 \dots A_n} \ . \tag{1.25}$$

An important special case of the formula (1.25) is obtained by setting $r = n - 1$ to obtain,

$$a_n = \frac{A_1 \dots A_{n-1}}{A_1 \dots A_n} \, .$$

Since the reference hyperplanes are a set of arbitrary hyperplanes, we can relabel the hyperplanes to obtain a formula that expresses a general reference point a_j in terms of the reference hyperplanes.

$$\begin{aligned} a_j &= \frac{A_1 \dots \check{A}_j \dots A_n}{A_1 \dots \check{A}_j \dots A_n A_j} = \frac{A_1 \dots \check{A}_j \dots A_n}{(-1)^{n-j} A_1 \dots A_n} \\ &= \frac{(-1)^{n-j} A_1 \dots \check{A}_j \dots A_n}{A_1 \dots A_n} \end{aligned} \tag{1.26}$$

This is the dual formula to equation (1.18).

In section 1.4 we took the case $n = 4$ as an example and set out the basis elements (linear subspaces) generated by the anti-commuting reference points. Dually, the basis elements generated by the anti-commuting hyperplanes (which are planes in the case $n = 4$) are as follows; there are 4 basis elements A_1, A_2, A_3, A_4 of hyperplane[8] grade 1 (planes), the 6 basis elements A_1A_2, A_1A_3, A_1A_4, A_2A_3, A_2A_4, A_3A_4 of hyperplane grade 2 (lines), the 4 basis elements $A_1A_2A_3$, $A_1A_2A_4$, $A_1A_3A_4$, $A_2A_3A_4$ of hyperplane grade 3 (points) and a single basis element $A_1A_2A_3A_4$ of hyperplane grade 4 (no physical interpretation). Using equation (1.22), we now show how these basis elements are related to the original ones in the example in section 1.4.

$$\begin{aligned} A_1 &= \frac{a_2a_3a_4}{a_1a_2a_3a_4} \\ A_2 &= \frac{a_1a_3a_4}{a_2a_1a_3a_4} = -\frac{a_1a_3a_4}{a_1a_2a_3a_4} \\ A_3 &= \frac{a_1a_2a_4}{a_3a_1a_2a_4} = \frac{a_1a_2a_4}{a_1a_2a_3a_4} \\ A_4 &= \frac{a_1a_2a_3}{a_4a_1a_2a_3} = -\frac{a_1a_2a_3}{a_1a_2a_3a_4} \\ A_1A_2 &= \frac{a_3a_4}{a_1a_2a_3a_4} \end{aligned}$$

[8]Notice that it now becomes necessary to distinguish between the *point grade* of an element and its *hyperplane grade*. For example, A_1 is of hyperplane grade 1, but when we write it as a product of points,

$$A_1 = \frac{a_2a_3a_4}{a_1a_2a_3a_4}$$

using (1.18) with $n = 4$, it is of point grade 3. In general, equation (1.22) shows that an element of hyperplane grade r can be written as a product of points of grade $n - r$.

$$
\begin{aligned}
A_1A_3 &= \frac{a_2a_4}{a_1a_3a_2a_4} = -\frac{a_2a_4}{a_1a_2a_3a_4} \\
A_1A_4 &= \frac{a_2a_3}{a_1a_4a_2a_3} = \frac{a_2a_3}{a_1a_2a_3a_4} \\
A_2A_3 &= \frac{a_1a_4}{a_2a_3a_1a_4} = \frac{a_1a_4}{a_1a_2a_3a_4} \\
A_2A_4 &= \frac{a_1a_3}{a_2a_4a_1a_3} = -\frac{a_1a_3}{a_1a_2a_3a_4} \\
A_3A_4 &= \frac{a_1a_2}{a_3a_4a_1a_2} = \frac{a_1a_2}{a_1a_2a_3a_4} \\
A_1A_2A_3 &= \frac{a_4}{a_1a_2a_3a_4} \\
A_1A_2A_4 &= \frac{a_3}{a_1a_2a_4a_3} = -\frac{a_3}{a_1a_2a_3a_4} \\
A_1A_3A_4 &= \frac{a_2}{a_1a_3a_4a_2} = \frac{a_2}{a_1a_2a_3a_4} \\
A_2A_3A_4 &= \frac{a_1}{a_2a_3a_4a_1} = -\frac{a_1}{a_1a_2a_3a_4} \\
A_1A_2A_3A_4 &= \frac{1}{a_1a_2a_3a_4}
\end{aligned}
$$

1.10 Another Type of Pseudonumber

From (1.24) and (1.9) it follows that[9],

$$
\begin{aligned}
1 &= f(a_1 \dots a_n . A_1 \dots A_n) = f(a_1 \dots a_n).f(A_1 \dots A_n) \\
&= \det(f^{-1}) a_1 \dots a_n . f(A_1 \dots A_n) = \frac{f(A_1 \dots A_n)}{(\det f) A_1 \dots A_n} \; .
\end{aligned}
$$

Therefore, the pseudonumber $A_1 \dots A_n$ transforms as,

$$
f(A_1 \dots A_n) = fA_1 \dots fA_n = (\det f) A_1 \dots A_n \; . \tag{1.27}
$$

This means that $a_1 \dots a_n$ and $A_1 \dots A_n$ pseudonumbers transform in different ways. Consequently, following the argument in section 1.5, this implies that $a_1 \dots a_n$ and $A_1 \dots A_n$ are different types of pseudonumber. Therefore, we can form linear combinations of basis elements weighted by pseudonumbers proportional to $A_1 \dots A_n$, but there will never be linear combinations weighted by mixtures of the two different types of pseudonumber.

[9]We use the notational conventions of section 1.1 freely from now on.

1.11 Full Definition of the Product

In section 1.7 we were able to define the product XY for the case when the sum of the point grades of X and Y does not exceed n. However, it was not possible to define the product when the sum of point grades exceeds n except when either X or Y is a pseudonumber. Now, let X be of point grade r and Y be of point grade s just as in section 1.7. Furthermore, assume $r + s > n$ so that we have the case for which we currently lack a rule. Note that $r + s > n$ implies that $(n - r) + (n - s) < n$. Now, if we express both X and Y in terms of hyperplanes by (1.22), X will be of hyperplane grade $n - r$ and Y will be of hyperplane grade $n - s$. In other words, the sum of the hyperplane grades of X and Y does not exceed n. Furthermore, by duality, the rule for multiplying X and Y will exist for precisely this case. An example will clarify the procedure.

In section 1.7 we took $n = 4$ and tried to multiply the line $X = a_1 a_2$ by the plane $Y = a_2 a_3 a_4$. The sum of the point grades is 5 so that there was no rule for this product. However, now we can write X and Y as products of hyperplanes. The examples at the end of section 1.9, or equation (1.25) can be used to obtain,

$$\begin{aligned} X &= a_1 a_2 = \frac{A_3 A_4}{A_3 A_4 A_1 A_2} = \frac{A_3 A_4}{A_1 A_2 A_3 A_4} \\ Y &= a_2 a_3 a_4 = \frac{A_1}{A_1 A_2 A_3 A_4} . \end{aligned}$$

The sum of the hyperplane grades is 3 and so we can evaluate the product by the dual version of (1.12). Thus,

$$\begin{aligned} XY &= \left(\frac{A_3 A_4}{A_1 A_2 A_3 A_4}\right)\left(\frac{A_1}{A_1 A_2 A_3 A_4}\right) = \frac{A_3 A_4 A_1}{(A_1 A_2 A_3 A_4)^2} = \frac{A_1 A_3 A_4}{(A_1 A_2 A_3 A_4)^2} \\ &= \frac{a_2}{A_1 A_2 A_3 A_4} = (a_1 a_2 a_3 a_4) a_2 \ . \end{aligned}$$

The algebraic product XY can now be defined for all basis elements X and Y. It can be extended to arbitrary elements formed as linear combinations of the basis elements by the distributive rule (1.11). There are three cases to consider. If one of the factors is a pseudonumber the product is proportional to the other factor. See equation (1.14). If the factors are written as products of points and the sum of the grades does not exceed n, the product is given by (1.12). Finally, if the factors are written as products of hyperplanes and the sum of the grades does not exceed n then we have the case $X = A_1 \dots A_r$ and $Y = A_{r+1} \dots A_{r+s}$ with $r + s \leq n$. The product is,

$$XY = (A_1 \dots A_r)(A_{r+1} \dots A_{r+s}) = A_1 \dots A_r A_{r+1} \dots A_{r+s} \ . \tag{1.28}$$

The dual formula to (1.13) is the useful result,

$$\begin{aligned} YX &= (A_{r+1}\ldots A_{r+s})(A_1\ldots A_r) = A_{r+1}\ldots A_{r+s}A_1\ldots A_r \\ &= (-1)^s A_1 a_{r+1}\ldots A_{r+s}A_2\ldots A_r = (-1)^{rs}A_1\ldots A_r a_{r+1}\ldots A_{r+s} \\ &= (-1)^{rs}XY\ . \end{aligned} \tag{1.29}$$

This formula used to swop the order of factors in a product when the product evaluates according to case (1.28)[10].

Let us now consider the meaning of the product. Let $X = a_1\ldots a_r$ and $Y = a_{r+1}\ldots a_{r+s}$ with $r+s \leq n$. The product is $XY = a_1\ldots a_{r+s}$ and it represents a joining together of the linear subspaces X and Y. For example, take the case $n = 4$ with the point $X = a_1$ and the line $Y = a_2a_3$. The product is the plane $XY = a_1a_2a_3$. Alternatively, suppose the factor subspaces are written as $X = A_1\ldots A_r$ and $Y = A_{r+1}\ldots A_{r+s}$ with $r+s \leq n$: the product is $XY = A_1\ldots A_{r+s}$ and it represents the intersection of the linear subspaces X and Y. In the example of the product of the line $X = a_1a_2$ and the plane $Y = a_2a_3a_4$ earlier in the current section, the result was proprtional to the point a_2. Clearly, the intersection of the line through points a_1 and a_2 with the plane through points a_2, a_3 and a_4 is at the point a_2. Notice that the basis element $a_1\ldots a_r$ represents the linear subspace throught the r points, whilst the basis element $A_1\ldots A_r$ represents the linear subspace at the intersection of the r hyperplanes. This can be seen from the examples at the end of section 1.9. For example, the point a_1 is at the common intersection of the three planes A_2, A_3 and A_4. In summary, when the factor subspaces X, Y intersect each other, the product

[10]This rule for swopping the order of factors in a product is consistent with the assumption that a pseudonumber $a_1\ldots a_n$ is a weight. A weight commutes with every element of the algebra. Let's check that $a_1\ldots a_n.X = X.a_1\ldots a_n$ for any element X. Since the pseudonumber is of point grade n, the sum of grades will automatically exceed n and so we have to evaluate $a_1\ldots a_n.X$ by writing both factors as products of hyperplanes. The commutation properties can then be investigated using (1.29). If X is of point grade r it is of grade $n-r$ when written as a product of hyperplanes. Similarly $a_1\ldots a_n$ is of grade $n-n=0$ when written as a product of hyperplanes. In fact this is because, from (1.24), we can write,

$$a_1\ldots a_n = \frac{1}{A_1\ldots A_n}$$

so that the numerator is a number which is of grade 0. Hence, the exponent of (-1) in (1.29) is zero and so the elements commute. In other words a pseudonumber commutes with everything. Furthermore, multiplication by a pseudonumber does not alter the grade; the grade of $a_1\ldots a_n.X$ being the same as the grade of X. Since a pseudonumber commutes with everything and does not alter the grade, it operates just like a number and so this is a further justification for allowing pseudonumbers as weights. The dual argument that $A_1\ldots A_n$ commutes with everything can also be made using (1.13).

XY is their intersection: when the factor subspaces do not intersect, the product represents their joining together. Intersection is made by products of hyperplanes: joining together is made by products of points.

1.12 Rule of the Middle Factor

Section 1.11 defined the algebraic product XY for all pairs of elements X, Y. The rule for calculating the product is a little cumbersome because it involves three cases given by equations (1.12,1.14, 1.28). The case chosen depends on the grades of the factors X, Y. We now derive the *rule of the middle factor* to evaluate the product in a more straightforward way[11].

Consider two linear subspaces X and Y. In order to get a neat formula for calculating the product XY we start by bringing the factors to a standard form by expressing X as a product of points and Y as a product of hyperplanes. If the points are more numerous than the hyperplanes then $X = a_1 \dots a_r a_{r+1} \dots a_{r+s}$ and $Y = A_1 \dots A_r$ and the product is the intersection of the subspaces[12]. The product is,

$$\begin{aligned} XY &= a_1 \dots a_r a_{r+1} \dots a_{r+s}.A_1 \dots A_r \\ &= \left(\frac{A_{r+s+1} \dots A_n}{A_{r+s+1} \dots A_n A_1 \dots A_{r+s}} \right) A_1 \dots A_r \\ &= \frac{A_{r+s+1} \dots A_n A_1 \dots A_r}{A_{r+s+1} \dots A_n A_1 \dots A_{r+s}} \\ &= a_{r+1} \dots a_{r+s} \end{aligned}$$

by two applications of equation (1.25). However, if the hyperplanes are more numerous than the points then $X = a_{r+1} \dots a_{r+s}$ and $Y = A_1 \dots A_r A_{r+1} \dots A_{r+s}$ and the product is the joining together of the subspaces[13]. The product is,

$$XY = a_{r+1} \dots a_{r+s}.A_1 \dots A_r A_{r+1} \dots A_{r+s}$$

[11] Appendix B covers the same material as the current chapter, but develops the theory in an axiomatic way by postulating the rule of the middle factor, instead of deriving it.

[12] The reason that the product is the intersection is explained as follows. Suppose we tried to evaluate XY as a product of points using (1.12). We would need to write Y as a product of $n - r$ points using (1.22). The sum of the point-grades of X and Y is then $r + s + (n - r) > n$. Hence the product is an intersection and has to be evaluated using the intersection of hyperplanes case (1.28) instead of the join of points case (1.12).

[13] The reason that the product is the joining together of the subspaces is explained as follows. We try to evaluate XY as a product of points using (1.12). We write Y as a product of $n - r - s$ points using (1.22). The sum of the point-grades of X and Y is then $s + (n - r - s) < n$ which means that we were correct in our choice of (1.12) as the rule for the product in this case. The product of points is the case of the joining together of subspaces.

$$
\begin{aligned}
&= a_{r+1}\dots a_{r+s}\left(\frac{a_{r+s+1}\dots a_n}{a_1\dots a_n}\right)\\
&= \frac{a_{r+1}\dots a_n}{a_1\dots a_n}\\
&= A_1\dots A_r
\end{aligned}
$$

by two applications of equation (1.22). The first of the above cases was calculated by writing XY as a product of hyperplanes, so in this case, XY is the intersection. In the second case, XY is the join of X and Y because it was written as a product of points. These results give the *rule of the middle factor*,

$$
\begin{aligned}
a_1\dots a_r a_{r+1}\dots a_{r+s}.A_1\dots A_r &= a_{r+1}\dots a_{r+s} && (1.30)\\
a_{r+1}\dots a_{r+s}.A_1\dots A_r A_{r+1}\dots A_{r+s} &= A_1\dots A_r && (1.31)
\end{aligned}
$$

for calculating the intersection or join of two regions. If the left and right factors have common labels, but there are no elements in the middle factor, then $XY = 1$ because the product reduces to the form of equation (1.23). The only other case that can occur is if the left and right factors have no common labels, then $XY = 0$ because when we convert to a product of elements of one type, there will be repeated elements which zero the product by antisymmetry. For example, in a 5-d space, so $n = 6$, we could have the product $a_1a_2a_3a_5.A_1A_2A_4$ which has no middle factor. We can evaluate this product by converting the first factor to a product of hyperplanes using (1.25). With neglect of an unimportant weight, the product becomes $A_4A_6.A_1A_2A_4 = A_4A_6A_1A_2A_4$ which is zero by antisymmetry because the hyperplane A_4 occurs twice in the product.

The rule of the middle factor is illustrated by the following two examples which have already been evaluated in section 1.11 using the original method of writing the factors as freely associative products of points (join) or hyperplanes (intersection). Both examples are for 3-dimensional projective space so that $n = 4$. The first example is the product of the point $X = a_1$ and the line $Y = a_2a_3$.

$$
\begin{aligned}
XY &= a_1.a_2a_3 = a_1.\frac{A_1A_4}{A_1A_4A_2A_3} = \frac{a_1.A_1A_4}{A_1A_2A_3A_4}\\
&= -\frac{a_1.A_4A_1}{A_1A_2A_3A_4} = -\frac{A_4}{A_1A_2A_3A_4} = a_1a_2a_3
\end{aligned}
$$

The second example is the product of the line $X = a_1a_2$ and the plane $Y = a_2a_3a_4$.

$$
XY = a_1a_2.a_2a_3a_4 = a_1a_2.\frac{A_1}{A_1A_2A_3A_4} = \frac{a_1a_2.A_1}{A_1A_2A_3A_4}
$$

$$= \quad \frac{a_2}{A_1A_2A_3A_4} = a_2(a_1a_2a_3a_4)$$

Finally, by a trivial substitution, we present the rule of the middle factor in a form which will be used in the derivation of certain formulae in section 1.14. By changing the definition of s and re-labelling the points of the middle factor in equation (1.30) we can just as well state this version of the rule as,

$$a_1 \ldots a_r a_{r+s+1} \ldots a_n.A_1 \ldots A_r = a_{r+s+1} \ldots a_n \ .$$

and then by substituting for the middle factor using equation (1.25) we arrive at,

$$(a_1 \ldots a_r.A_1 \ldots A_{r+s}).A_1 \ldots A_r = A_1 \ldots A_{r+s} \ . \tag{1.32}$$

Similarly, by substituting for the middle factor in equation (1.31) using (1.22) we obtain,

$$a_{r+1} \ldots a_{r+s}.(a_{r+1} \ldots a_n.A_{r+1} \ldots A_{r+s}) = a_{r+1} \ldots a_n \ .$$

By re-labelling and changing r and s we can just as well write this version of the rule as,

$$a_1 \ldots a_r.(a_1 \ldots a_{r+s}.A_1 \ldots A_r) = a_1 \ldots a_{r+s} \ . \tag{1.33}$$

The pair (1.32,1.33) constitute another form of the rule of the middle factor.

1.13 Coordinates for Linear Subspaces

An arbitrary element X of point grade r is given by equation (1.5) which is repeated as,

$$X = \sum_{j_1 \cdots j_r} \xi_{j_1 \ldots j_r} a_{j_1} \ldots a_{j_r}$$

for $1 \leq j_1 < \ldots < j_r \leq n$. The coefficients in the expansion are found as follows.

$$X.A_{k_1} \ldots A_{k_r} = \sum_{j_1 \cdots j_r} \xi_{j_1 \cdots j_r}(a_{j_1} \ldots a_{j_r}.A_{k_1} \ldots A_{k_r})$$

where $1 \leq k_1 < \ldots < k_r \leq n$. The labels in the term $a_{j_1} \ldots a_{j_r}.A_{k_1} \ldots A_{k_r}$ only match up when $k_1 = j_1, \ldots, k_r = j_r$. Hence, from the rule of the middle factor, the term is unity in this case and zero otherwise. The coefficients are therefore, $\xi_{j_1 \cdots j_r} = X.A_{j_1} \ldots A_{j_r}$. The expansion of an arbitrary element of point-grade r is therefore,

$$X = \sum_{j_1 \cdots j_r} (X.A_{j_1} \ldots A_{j_r}) a_{j_1} \ldots a_{j_r} \tag{1.34}$$

for $1 \leq j_1 < \ldots < j_r \leq n$. Similarly, the expansion of an arbitrary element of hyperplane grade s (point grade $r = n - s$) can be shown to be,

$$Y = \sum_{j_1 \cdots j_s} (a_{j_1} \ldots a_{j_s}.Y) A_{j_1} \ldots A_{j_s} \tag{1.35}$$

for $1 \leq j_1 < \ldots < j_s \leq n$.

As an example, consider the 3-d projective space with reference points a_1, a_2, a_3, a_4. The line l through the points $2a_1 + a_2$ and $a_3 - a_4$ is,

$$l = (2a_1 + a_2)(a_3 - 3a_4) = 2a_1a_3 - 6a_1a_4 + a_2a_3 - 3a_2a_4$$

and this expansion is of the form of equation (1.5). However, not all expansions yield linear subspaces. In order to see this, consider the linear subspace $X = a_1 \ldots a_r$ through the r points $a_1, \ldots a_r$. A general point p in this subspace is the linear combination,

$$p = \sum_{i=1}^{r} \xi_i a_i \ .$$

The product pX is most easily evaluated[14] as a case of the form (1.12),

$$pX = \sum_{i=1}^{r} \xi_i a_i.a_1 \ldots a_r = \sum_{i=1}^{r} \xi_i a_i a_1 \ldots a_r = 0$$

which is zero by antisymmetry because each point a_i in the summation also occurs in the product $X = a_1 \ldots a_r$. Thus, a linear subspace is the locus of points p which obey the equation $pX = 0$. Now the grade 2 element $X = a_1a_2 + a_3a_4$ does not represent a line. If it was a line, then the points p on the line would satisfy the equation $pX = 0$. However, if we write $p = \xi_1a_1 + \xi_2a_2 + \xi_3a_3 + \xi_4a_4$ and try to solve for the coefficients we find that,

$$0 = pX = \xi_1a_1a_3a_4 + \xi_2a_2a_3a_4 + \xi_3a_3a_1a_2 + \xi_4a_4a_1a_2 \ .$$

[14]We could also evaluate pX using the rule of the middle factor. From (1.25), X is represented as the following intersection of hyperplanes,

$$a_1 \ldots a_r = \frac{A_{r+1} \ldots A_n}{A_{r+1} \ldots A_n A_1 \ldots A_r} = \frac{(-1)^{r(n-r)} A_{r+1} \ldots A_n}{A_1 \ldots A_n} \ .$$

Now,

$$pX = \sum_{i=1}^{r} \xi_i a_i . \frac{(-1)^{r(n-r)} A_{r+1} \ldots A_n}{A_1 \ldots A_n} = \sum_{i=1}^{r} (-1)^{r(n-r)} \xi_i \frac{a_i.A_{r+1} \ldots A_n}{A_1 \ldots A_n} = 0$$

because there are no middle factors.

Each term in the above equation is independent and so $\xi_1 = \xi_2 = \xi_3 = \xi_4 = 0$. Therefore, there are no points which satisfy the equation of the line and so $X = a_1a_2 + a_3a_4$ is not a line. Instead, it is called a *compound element*. The elements which represent linear subspaces are called *simple elements* because they can always be written as a simple product of points for some choice of the points. It can be shown that all hyperplanes are simple elements. So, in 2-d projective space, where the hyperplanes are lines, all grade 2 elements are simple. However, in 3-d projective space, as we have seen, grade 2 elements can be compound.

1.14 Fundamental Formulae

The rule of the middle factor only applies to elements made from products of reference points and their dual hyperplanes. In this section we will find formulae that allow us to calculate the product XY of two arbitrary linear subspaces $X = a_1 \ldots a_r$ and $Y = B_1 \ldots B_s$ where the a_i are points and the B_i are hyperplanes.

Firstly, consider the case $r > s$ so the product is the intersection of the two subspaces. Expand Y in terms of products of the hyperplanes A_i which are dual to the points a_i using equation (1.35).

$$\begin{aligned} XY &= a_1 \ldots a_r . \sum_{j_1 \ldots j_s} (a_{j_1} \ldots a_{j_s} . Y) A_{j_1} \ldots A_{j_s} \\ &= \sum_{j_1 \ldots j_s} (a_{j_1} \ldots a_{j_s} . Y)(a_1 \ldots a_r . A_{j_1} \ldots A_{j_s}) \end{aligned} \tag{1.36}$$

The summation is over $1 \leq j_1 < \ldots < j_s \leq n$. In the term $a_1 \ldots a_r . A_{j_1} \ldots A_{j_s}$, the hyperplanes $A_{j_1} \ldots A_{j_s}$ must be taken from the set $A_1, \ldots, A_r$, otherwise the term $a_1 \ldots a_r . A_{j_1} \ldots A_{j_s}$ would be zero because there would be no middle factor. In this case, the rule of the middle factor, in the form of equation (1.33) gives,

$$X = a_1 \ldots a_r = a_{j_1} \ldots a_{j_s} . (a_1 \ldots a_r . A_{j_1} \ldots A_{j_s})$$

where the labels $1 \leq j_1 < \ldots < j_s \leq n$ are s labels taken from the set $1 \ldots r$. We can write this factoring of X as,

$$X = X_s^{(\lambda)} X_{r-s}^{(\lambda)} \tag{1.37}$$

where $X_s^{(\lambda)} = a_{j_1} \ldots a_{j_s}$ and λ ranges over the ${}^rC_s$ ways of taking s points out of the set $a_1, \ldots, a_r$. The product (1.36) XY is now,

$$XY = \sum_{\lambda} (X_s^{(\lambda)} Y) X_{r-s}^{(\lambda)} \ . \tag{1.38}$$

For example, let p, q, r be points and L be a hyperplane. The intersection of the plane pqr with the hyperplane is $pqr.L$. In this case $r = 3$ and $s = 1$. The ${}^3C_1 = 3$ ways of factoring pqr according to equation (1.37) are $pqr = p.qr = -q.pr = r.pq$. From (1.38) the product is,

$$pqr.L = (pL)qr - (qL)pr + (rL)pq \ .$$

This is a linear combination of lines qr, pr and pq with weights pL, qL and rL. Hence the intersection of a plane with a hyperplane is a line. In 3-d projective space, where $n = 4$, hyperplanes are planes, and so we have the formula for the intersection of two planes to give a line.

As another example, in 3-d projective space, consider the intersection of a plane pqr with a line st where p, q, r, s, t are all points. The intersection is the product $pqr.st$ and although the line st is written as a product of points, formula (1.38) requires that we need to know how many hyperplane factors are needed to write st. Since $n = 4$, the line st converts to a product of $4 - 2 = 2$ hyperplanes. We are not interested in the set of hyperplanes dual to the points s, t that would be needed to actually write st as a product of two hyperplanes using equation (1.25). Instead, we just think of st as a product of two hyperplanes. The ${}^3C_2 = 3$ ways of factoring pqr according to equation (1.37) are $pqr = pq.r = -pr.q = qr.p$. From (1.38) the product is,

$$pqr.st = (pq.st)r - (pr.st)q + (qr.st)p = (pqst)r - (prst)q + (qrst)p \ .$$

This is a linear combination of points r, q and p with pseudonumber weights $pqst$, $prst$ and $qrst$. Hence, in 3-d projective space, the intersection of a plane with a line is a point. The weights have turned out to be pseudonumbers instead of numbers because we did not bother to use equation (1.25) to convert st to a product of hyperplanes. If we had done this the weights would have been ordinary numbers. However, it makes no difference, because all pseudonumbers are proportional to some standard pseudonumber such as (say) $pqst$. If we take this pseudonumber out as a common factor, then,

$$pqr.st = (pqst)\left(r - \frac{prst}{pqst}q + \frac{qrst}{pqst}p\right)$$

where the fractions are both ordinary numerical weights. Since an overall weight has no physical significance, the formula for the intersection of a plane and a line is,

$$pqr.st = r - \frac{prst}{pqst}q + \frac{qrst}{pqst}p \ .$$

As well as giving another example of formula (1.38), this example has shown how we can treat pseudonumbers in the same way as ordinary numbers. In fact, in the example, we could have ignored the distinction between pseudonumbers and numbers by arbitrarily deciding that $pqst = 1$. This is what happens in the computer implementation of this scheme, where the pseudonumber $a_1 \ldots a_n = 1$. Of course, this is not strictly true because equation (1.9) has shown that numbers and pseudonumbers transform in different ways, so they are not the same kind of quantity.

Now, consider the case $r < s$ so the product is the join of the two subspaces. Expand X in terms of products of the points b_i which are dual to the hyperplanes B_i using equation (1.5).

$$\begin{aligned} XY &= \left(\sum_{j_1 \ldots j_r} (X.B_{j_1} \ldots B_{j_r}) b_{j_1} \ldots b_{j_r} \right) .B_1 \ldots B_s \\ &= \sum_{j_1 \ldots j_r} (X.B_{j_1} \ldots B_{j_r})(b_{j_1} \ldots b_{j_r}.B_1 \ldots B_s) \end{aligned} \tag{1.39}$$

The summation is over $1 \leq j_1 < \ldots < j_r \leq n$. In the term $b_{j_1} \ldots b_{j_r}.B_1 \ldots B_s$, the points $b_{j_1} \ldots b_{j_r}$ must be taken from the set $b_1, \ldots, b_s$, otherwise the term $b_{j_1} \ldots b_{j_r}.B_1 \ldots B_s$ would be zero because there would be no middle factor. In this case, the rule of the middle factor, in the form of equation (1.32) gives,

$$Y = B_1 \ldots B_s = (b_{j_1} \ldots b_{j_r}.B_1 \ldots B_s).B_{j_1} \ldots B_{j_r}.$$

where the labels $1 \leq j_1 < \ldots < j_r \leq n$ are r labels taken from the set $1 \ldots s$. We can write this factoring of Y as,

$$Y = Y^{(\lambda)}_{s-r} Y^{(\lambda)}_r \tag{1.40}$$

where $Y^{(\lambda)}_r = B_{j_1} \ldots B_{j_r}$ and λ ranges over the ${}^sC_r$ ways of taking r hyperplanes out of the set $B_1, \ldots, B_s$. The product (1.39) XY is now,

$$XY = \sum_\lambda (XY^{(\lambda)}_r) Y^{(\lambda)}_{s-r} \ . \tag{1.41}$$

This derivation is an example of the principle of duality because we have obtained equation (1.41) by following the derivation of equation (1.38) and just changing points to hyperplanes and adjusting a few words.

As an example of the use of formula (1.41) consider the product $p.LMN$ where p is a point and L, M, N are hyperplanes. In this case $r = 1$ and $s = 3$. The ${}^3C_1 = 3$ ways of factoring LMN according to equation (1.40) are $LMN = LM.N = -LN.M = MN.L$. From (1.41) the product is,

$$p.LMN = (pN)LM - (pM)LN + (pL)MN \ .$$

This is a linear combination of subspaces LM, LN and MN with weights pN, pM and pL. In 3-d projective space, where $n = 4$, hyperplanes are planes, the intersection of three planes LMN is a point. The example is a case of the join of two points to make a line.

We also need a formula to calculate the product when $r = s$ so that $XY = (a_1 \dots a_r)(B_1 \dots B_r)$. In this case we expand each hyperplane B_i according to $B_i = \sum_j (a_j B_i) A_j$. This expansion is a case of equation (1.35). Calculating,

$$\begin{aligned}(a_1 \dots a_r)(B_1 \dots B_r) &= a_1 \dots a_r . \sum_{j_1} (a_{j_1} B_1) A_{j_1} \dots \sum_{j_r} (a_{j_r} B_r) A_{j_r} \\ &= \sum_{j_1} \dots \sum_{j_r} (a_{j_1} B_1) \dots (a_{j_r} B_r)(a_1 \dots a_r . A_{j_1} \dots A_{j_r}) \ . \qquad (1.42)\end{aligned}$$

In this case each label in the sum goes from 1 to n. However, the term $a_1 \dots a_r . A_{j_1} \dots A_{j_r}$ is zero unless the labels $j_1, \dots, j_r$ are a permutation of $1, \dots, r$. This is apparent from the rule of the middle factor or from the argument concerning a similar term that occurred in equation (1.36). Let σ be a permutation of $1 \dots r$ so that,

$$\begin{aligned}a_1 \dots a_r . A_{j_1} \dots A_{j_r} &= a_1 \dots a_r . A_{\sigma(1)} \dots A_{\sigma(r)} \\ &= \mathrm{sgn}(\sigma) a_1 \dots a_r . A_1 \dots A_r = \mathrm{sgn}(\sigma) \ .\end{aligned}$$

Putting this result into equation (1.42) gives,

$$(a_1 \dots a_r)(B_1 \dots B_r) = \sum_\sigma \mathrm{sgn}(\sigma)(a_{\sigma(1)} B_1) \dots (a_{\sigma(r)} B_r)$$

We could also have expanded the points a_i in terms of points b_i dual to the hyperplanes B_i and this gives another version of the formula. The two versions are,

$$\begin{aligned}(a_1 \dots a_r)(B_1 \dots B_r) &= \sum_\sigma \mathrm{sgn}(\sigma)(a_{\sigma(1)} B_1) \dots (a_{\sigma(r)} B_r) \\ &= \sum_\sigma \mathrm{sgn}(\sigma)(a_1 B_{\sigma(1)}) \dots (a_r B_{\sigma(r)}) \qquad (1.43)\end{aligned}$$

Notice that (1.43) is how we actually evaluate the numbers $X_s^{(\lambda)} Y$ in formula (1.38) and $XY_r^{(\lambda)}$ in formula (1.41).

As an example, consider the number $pq.LM$ where p, q are points and L, M are hyperplanes. From (1.43),

$$pq.LM = (pL)(qM) - (pM)(qL) \ .$$

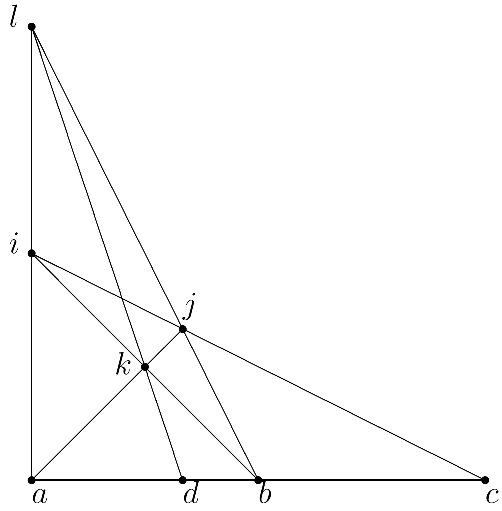

Figure 1.1: Construction of the Harmonic Section

It is also useful to note that the formula for a number such as $(a_1 \ldots a_r)(B_1 \ldots B_r)$ is associative and we can write,

$$(a_1 \ldots a_r)(B_1 \ldots B_r) = (a_1 \ldots a_{r-1}.B_1)(B_2 \ldots B_r) \ . \tag{1.44}$$

This follows because $a_1 \ldots a_r$ can be expressed as a product of $n-r$ hyperplanes by equation (1.25). The product of the r hyperplanes $B_1 \ldots B_r$ and the $n-r$ hyperplanes from the dual representation of $a_1 \ldots a_r$ is completely associative so we can bracket the expression as shown on the rhs of equation (1.44).

Equations (1.38), (1.41) and (1.43) are the fundamental formulae of the algebra. They specify how to calculate product XY of two linear subspaces X and Y. If the subspaces overlap the product is the intersection, whilst if they do not overlap, the product evaluates to the join of the subspaces.

1.15 The Harmonic Section Theorem

With the machinery of the algebra of projective geometry that we now have at our disposal, it would be perfectly possible to work through all of the material on projective geometry in (say) Coxeter's book [9]. However, this chapter does not aim to be a treatise on projective geometry, so we will just pause to prove the remarkable *harmonic section theorem*[15].

[15] Section 2.4 of Coxeter's book [9].

In figure 1.1, a, b and c are collinear points. Draw any line through c and pick any points i and j on this line. Join a and b in all possible ways to i and j and then join the resulting intersections to give the line lk. This line cuts line ab at point d. Point d is called the *harmonic conjugate* of c with respect to a and b. The remarkable fact, which we shall prove by a straightforward calculation, is that point d only depends on a, b and c. It does not depend on the construction line through c or on the points i and j. This is the harmonic section theorem.

In order to prove this result, take points a, b and i as reference points for the plane. Let $c = a + b$. With this formula, c can still be an arbitrary point on line ab if we absorb weights into the definitions of the points a and b. Let $j = c + i$. Point j can still be any point on the line through c if we absorb an arbitrary weight into the definition of point i. Absorbing weights in this way just simplifies the calculations. Substituting for c we get $j = a + b + i$. Now evaluate the intersections $k = aj.bi$ and $l = ai.bj$. Treat bi and bj as lines (hyperplanes) and use (1.38).

$$\begin{aligned}
k &= aj.bi = (abi)j - (jbi)a = (abi)(a + b + i) - ((a + b + i)bi)a \\
&= (abi)a + (abi)b + (abi)i - (abi)a = (abi)(b + i)\ .
\end{aligned}$$

The weight abi is not physically significant so that $k = b + i$. Similarly,

$$\begin{aligned}
l &= ai.bj = (abj)i - (ibj)a = (ab(a + b + i))i - (ib(a + b + i))a \\
&= (abi)i - (iba)a = (abi)i + (abi)a \\
l &= a + i\ .
\end{aligned}$$

Now calculate the point $d = kl.ab$.

$$\begin{aligned}
d &= kl.ab = ((b + i)(a + i)).ab = (ba + bi + ia).ab \\
&= bi.ab + ia.ab = -(iab)b + (iab)a \\
d &= a - b
\end{aligned}$$

The calculation shows that $d = a - b$ does not depend on the construction points i and j and this proves the theorem. This result also shows that, in Whitehead's algebra, the harmonic conjugate of $c = a + b$ with respect to a and b is written $d = a - b$.

Chapter 2

Computational Geometry on the Projective Plane

2.1 Introduction

We now use the algebra studied in chapter 1 to make a package for computational geometry on the projective plane. This will be a gentle introduction to the methods that we shall eventually employ for 3-d Euclidean computational geometry and it also serves to show how the theory of (n-1)-d projective space looks in the simple case of the projective plane where n=3. At the end of the chapter, we use the constructon of the function package to motivate the introduction of the important notion of a polarity.

2.2 Points and Lines

Consider the projective plane with reference points a_1, a_2, a_3. The pseudonumber $a_1a_2a_3 = 1$ in accordance with the explanation in section 1.14. A general point is,

$$x = x_1a_1 + x_2a_2 + x_3a_3 \tag{2.1}$$

and a general line[1] is,

$$y = y_{12}a_1a_2 + y_{13}a_1a_3 + y_{23}a_2a_3 \tag{2.2}$$

where the x_i and y_{ij} are weights[2]. In order to implement this scheme in a computer, it makes sense to represent the point x by the 1-d array of weights `x(i)` and the line y by the upper triangle part of the 2-d array `y(i,j)`.

[1] Recall that lines are the hyperplanes of the projective plane.

[2] Equations (2.1) and (2.2) are both special cases of the expansion (1.5).

The scientific interpreter *Yorick* [12] is ideally suited to handle arrays and consequently it is used for all the numerical implementations of Whitehead's algebra in this work. The following extract from a Yorick session shows how to define a point x, print out its representation and also see how it is stored as an array. The functions used for computational geometry on the projective plane are in the file `Whitehead2d.i`.

```
> #include "Whitehead2d.i"
> x=5.0*a1+6.0*a2+7.0*a3
> print_element, x
+5.000a1+6.000a2+7.000a3
> x
[5,6,7]
>
```

2.3 Dual Representations

From the example at the end of section 1.8, the lines dual to the reference points are

$$A_1 = a_2 a_3 \ , \ A_2 = -a_1 a_3 \ , \ A_3 = a_1 a_2$$

and from equation (1.26), the reference points are,

$$a_1 = A_2 A_3 \ , \ a_2 = -A_1 A_3 \ , \ a_3 = A_1 A_2$$

in terms of the dual lines. Consequently, we could also write the point x as

$$\begin{aligned} x &= x_1 a_1 + x_2 a_2 + x_3 a_3 = x_1 A_2 A_3 - x_2 A_1 A_3 + x_3 A_1 A_2 \\ &= X_{23} A_2 A_3 + X_{13} A_1 A_3 + X_{12} A_1 A_2 \end{aligned} \tag{2.3}$$

where the X_{ij} are weights. So, we could equally well represent the point x as an upper-triangle 2-d array `X(i,j)` related to the original point representation by `X(1,2)=x(3)`, `X(1,3)=-x(2)`, `X(2,3)=x(1)`. Similarly, the line y could be written,

$$\begin{aligned} y &= y_{12} a_1 a_2 + y_{13} a_1 a_3 + y_{23} a_2 a_3 = y_{12} A_3 - y_{13} A_2 + y_{23} A_1 \\ &= Y_3 A_3 + Y_2 A_2 + Y_1 A_1 \end{aligned} \tag{2.4}$$

where the Y_i are weights[3]. Therefore, the line y could be equally well represented by the 1-d array `Y(i)` related to the original 2-d array by `Y(1)=y(2,3)`, `Y(2)=-y(1,3)`, `Y(3)=y(1,2)`.

[3] Expansions (2.3) and (2.4) are both special cases of (1.35).

We choose to store points and lines in point-form as in equations (2.1,2.2), only converting to hyperplane-form as in equations (2.3, 2.4) when this is necessary for a calculation.

The function `I` is used to convert between the two ways of representing an element. In the following Yorick session, `x` and `X` are the two ways of representing the same point.

```
> x
[5,6,7]
> X=I(x)
> X
[[0,0,0],[7,0,0],[-6,5,0]]
>
```

2.4 The Antisymmetric Product

We now use the formulae of section 1.14 to calculate the various instances of the antisymmetric product of elements.

2.4.1 Point, Point

In order to calculate the product xy where x and y are both points, we use the Einstein summation convention[4] to write x in point-form as $x = x_i a_i$ and y in hyperplane-form as $y = Y_{jk} A_j A_k$. Then,

$$\begin{aligned} xy &= x_i a_i . Y_{jk} A_j A_k = x_i Y_{jk} (a_i . A_j A_k) \\ &= x_i Y_{jk} ((a_i A_k) A_j - (a_i A_j) A_k) = x_i Y_{jk} (\delta_{ik} A_j - \delta_{ij} A_k) \\ &= x_i Y_{ji} A_j - x_i Y_{ik} A_k = x_i Y_{ji} A_j - x_i Y_{ij} A_j \end{aligned}$$

where we used equations (1.41) and (1.19) and a change of dummy indices. Therefore, in hyperplane-form, the result of the product of $z = xy$ is the 1-d array of weights Z_j given by,

$$Z_j = x_i Y_{ji} - x_i Y_{ij} \tag{2.5}$$

which represents the line through the two points x and y.

[4]This means that there is an implied summation over any pair of repeated indices. So, $Y_{jk} A_j A_k$ means $\sum_{jk} Y_{jk} A_j A_k$.

2.4.2 Point, Line

In this case $z = xy$ where x is a point and y is a line. We write x in point-form and y in hyperplane-form.

$$z = xy = x_i a_i . Y_j A_j = x_i Y_j (a_i A_j) = x_i Y_j \delta_{ij} = x_i Y_i \tag{2.6}$$

The result is just a number.

2.4.3 Line, Point

The easiest way to evaluate $z = xy$ where x is a line and y is a point is to use equation (1.13) to swop the order of the factors to get $z = (-1)^{1.2} yx = yx$ and then use result (2.6).

2.4.4 Line, Line

In this case $z = xy$ where x and y are both lines. As usual we write x in point-form and y in hyperplane-form. Then,

$$\begin{aligned} z &= xy = x_{ij} a_i a_j . Y_k A_k = x_{ij} Y_k (a_i a_j . A_k) \\ &= x_{ij} Y_k ((a_i A_k) a_j - (a_j A_k) a_i) = x_{ij} Y_k (\delta_{ik} a_j - \delta_{jk} a_i) \\ &= x_{ij} Y_i a_j - x_{ij} Y_j a_i = x_{ij} Y_i a_j - x_{ji} Y_i a_j \end{aligned}$$

where we used equation (1.38) and a change of dummy indices. The result is a point given (in point-form) as the 1-d array,

$$z_j = x_{ij} Y_i - x_{ji} Y_i \tag{2.7}$$

which represents the point at the intersection of the two lines.

2.4.5 A Function for the Antisymmetric Product

The function[5] `wap` uses equations (2.5,2.6, 2.7) to implement all cases of the antisymmetric product of a pair of elements. Let us use the `wap` function to numerically check the validity of the harmonic section theorem of figure 1.1. The following program sets up fixed points a, b and c and then uses variable construction points i and j to calculate the point d which is the harmonic conjugate of a and b with respect to c. The point d is printed out with the coefficient of reference point a_1 normalized to unity.

[5]wap stands for *Whitehead Algebra Product*

```
//Yorick program hst.i which checks the validity
//of the harmonic section theorem
#include "Whitehead2d.i"
a=2.7*a1-6.9*a2+3.7*a3;//Fixed
b=3.5*a1+2.3*a2-4.6*a3;//Fixed
c=1.3*a-2.4*b;//Fixed
i=-2.3*a1-3.4*a2+4.5*a3;//Variable
j=10.4*i+1.3*c;//Variable
k=wap(wap(a,j),wap(b,i));
l=wap(wap(a,i),wap(b,j));
d=wap(wap(k,l),wap(a,b));
print_element, d/d(1);
```

The program produces the following output.

```
> #include "hst.i"
+1.000a1-0.290a2-0.523a3
>
```

The theorem is verified because point d is always the same point. It does not change if the construction points i and j are varied by editing their weights.

2.5 Yorick Package Whitehead2d

The Yorick package `Whitehead2d.i` for computational geometry in the projective plane is just comprised of the 3 functions, `I`, `wap` and `print_element` that have been used in the example Yorick sessions. The complete package is listed below.

```
/* Whitehead2d.i */

local Whitehead2d;
/* DOCUMENT Whitehead2d.i --- implementation of Whitehead's algebra
   for the 2-d elliptic space. All elements are stored on point format.
   Points are stored as 1-d arrays, lines (hyperplanes) as 2-d arrays.
   Each array dimension is of size 3. Pseudonumbers are treated as ordinary
   numbers.
   References.
   [1] Chaper 3 of "Computational Euclidean Geometry Modelled on the
   Absolute Quadric Hypersurface of Elliptic Space", S. Blake, December 2003.
   SEE ALSO: I, wap, print_element
*/
```

```
//Make some reference points for 2-d elliptic space.
a1=[1.,0.,0.];a2=[0.,1.,0.];a3=[0.,0.,1.];
//Sometimes it is useful to have the reference points as an array.
a=array(0.0,3,3);a(1,)=a1;a(2,)=a2;a(3,)=a3;

func I(x)
/* DOCUMENT I(x)
   returns the result of applying the absolute elliptic polarity to the
   element x. Also can be used to toggle the representation of an element.
   So, if (say) x is a line stored in point format as a 2-d array, then
   I(x) is the line stored in hyperplane format as a 1-d array.
   SEE ALSO: wap, print_element, Whitehead2d
*/
{
  //Get grade of element.
  nx=dimsof(x)(1);
  if (nx==1){
    X=array(double,3,3);
    X(1,2)=x(3);X(1,3)=-x(2);X(2,3)=x(1);
  } else if (nx==2){
    X=array(double,3);
    X(1)=x(2,3);X(2)=-x(1,3);X(3)=x(1,2);
  } else X=x;
  return X;
}

func wap(x,y)
/* DOCUMENT wap(x,y)
   returns the Whitehead algebra product xy.
   SEE ALSO: I, print_element, Whitehead2d
*/
{
  //Get grades of factors.
  nx=dimsof(x)(1);ny=dimsof(y)(1);
  if ((nx==0) || (ny==0)) {
    z=x*y;
  } else if ((nx==1) && (ny==1)){
    Y=I(y);
    Z=x(+)*Y(,+)-x(+)*Y(+,);
    z=I(Z);
```

```
  } else if ((nx==1) && (ny==2)){
    Y=I(y);
    Z=x(+)*Y(+);
    z=I(Z);
  } else if ((nx==2) && (ny==1)){
    z=wap(y,x);
  } else if ((nx==2) && (ny==2)){
    Y=I(y);
    z=x(+,)*Y(+)-x(,+)*Y(+);
  }
  return z;
}

func print_element(x)
/* DOCUMENT print_element, x
   prints out the Whitehead algebra element x to the terminal using the
   reference points a1,a2,a3.
   SEE ALSO: I, wap, Whitehead2d
*/
{
  if (dimsof(x)(1)>2){
    write,format="%s","Error -- not an element of Whitehead's algebra\n";
  } else if (dimsof(x)(1)==0) { write,format="%+.3f%s",x,"a1a2a3\n";
  } else {
    if (dimsof(x)(1)==1){
    list=[1,2,3];
    basis=["a1","a2","a3"];
    } else if (dimsof(x)(1)==2){
    list=[4,7,8];
    basis=["a1a2","a1a3","a2a3"];
    } for (i=1,str="";i<=numberof(list);++i){
      if (x(list)(i)) {
        str=str+swrite(format="%+.3f%s",x(list)(i),basis(i));
      }
    }
    if (strlen(str)) write,format="%s",str;else write,format="%s","0";
    write,format="%s","\n";
  }
}
```

2.6 Polarity

So far, the function `I` has just been used to convert a given element from point-form to hyperplane-form as in the example at the end of section 2.3. In this example, as in equation (2.3), the point x was initially written[6] as $x_i a_i$ in point-form and stored as the 1-d array x_i. This was converted into a 2-d array X_{ij} that represented the same point x in hyperplane-form as $X_{ij} A_i A_j$. The 1-d and 2-d arrays were related by $X_{12} = x_3$, $X_{13} = -x_2$, $X_{23} = x_1$. However, if we choose to interpret the 2-d array as a point-form, then it represents the line $X_{ij} a_i a_j$. This is what happens in the following Yorick session because all arrays are interpreted as point-forms by the function `print_element`.

```
> x=5.0*a1+6.0*a2+7.0*a3
> print_element, x
+5.000a1+6.000a2+7.000a3
> print_element, I(x)
+7.000a1a2-6.000a1a3+5.000a2a3
>
```

The conversion, which was introduced in order to calculate the antisymmetric product, defines a natural operator I which maps points to lines as,

$$\begin{aligned} Ix = I(x_1 a_1 + x_2 a_2 + x_3 a_3) &= X_{23} a_2 a_3 + X_{13} a_1 a_3 + X_{12} a_1 a_2 \\ &= x_1 a_2 a_3 - x_2 a_1 a_3 + x_3 a_1 a_2 = x_1 A_1 + x_2 A_2 + x_3 A_3 \ . \end{aligned} \tag{2.8}$$

From the above equation, the operator I is linear and acts on the reference points to give $I a_i = A_i$ for $i = 1, 2, 3$. A linear map from points to hyperplanes (in the general case of (n-1)-d projective space) is called a *correlation*. However, anticipating somewhat material from chapter 4, the correlation I is also involutary[7], as the following example calculation shows.

```
> print_element, I(I(a1))
+1.000a1
> print_element, I(I(wap(a1,a2)))
+1.000a1a2
>
```

An *involutary correlation* is called a *polarity*, so I is a polarity.

This section has just motivated the introduction of the polarity in the context of 2-d projective space. In chapter 4, the polarity is shown to be important for spaces which admit a notion of distance.

[6]The summation convention is in force in this section.

[7]In other words, $I^2 = 1$ up to a weight.

Chapter 3

Congruences

3.1 Introduction

This chapter treats the theory of congruences. In Whitehead's algebra, elliptic and hyperbolic geometries are introduced by picking a polarity in the projective space, and asserting that it appears invariant to all observers. The maps between the viewpoints of different observers are the congruences of the space[1]; they leave the polarity invariant. An arbitrary simple congruence is obtained by exponentiating an element of the Lie algebra of the group of congruences. A closed-form expression (3.26) for such a congruence in an (n-1)-dimensional space is given using the antisymmetric product of the algebra and the polarity. The theory of congruences is systematically exploited to define distance, angle, coordinates and volume.

3.2 Dual treatment of a collineation

Collineations[2] were introduced in section 1.2 as linear maps of points to points which are used to transform between the viewpoints of observers. Collineation is a synonym for linear operator. In equation (1.6), the action of a collineation on a grade r product of points was shown to obey the outermorphism property. This concentration on points is not in keeping with the principle of duality, so we have to put the transformation of points and hyperplanes on an equal footing.

Let $a_1, \ldots, a_n$ be reference points for the projective space. Suppose that a collineation takes points to points as $b_i = f(a_i)$ for $i = 1, 2, \ldots, n$. The

[1] In Euclidean space, the congruences are rotations and translations.
[2] The nomenclature here tends to follow Coxeter's book [9].

action of the collineation on the hyperplanes A_i dual to the reference points can be obtained from equation (1.18) and the outermorphism property.

$$f(A_i) = f\left(\frac{(-1)^{i-1}a_1 \ldots \check{a}_i \ldots a_n}{a_1 \ldots a_n}\right) = \frac{(-1)^{i-1}f(a_1)\ldots f(\check{a}_i)\ldots f(a_n)}{f(a_1)\ldots f(a_n)}$$
$$= \frac{(-1)^{i-1}b_1 \ldots \check{b}_j \ldots b_n}{b_1 \ldots b_n} = B_i$$

Here, the B_i are the hyperplanes dual to the points $b_1, \ldots, b_n$. Thus,

$$b_i = fa_i \text{ and } B_i = fA_i \tag{3.1}$$

and collineations have been treated dually.

Now consider the number U_rY_r formed from the product of a grade r product of points U_r and a grade r product of hyperplanes Y_r. Such products were considered in equation (1.43). Since U_rY_r is a number, it must transform as a scalar under the action of a collineation. Therefore $U_rY_r = f(U_rY_r) = f(U_r)f(Y_r)$. If we set $X_r = fU_r$ so that $U_r = f^{-1}X_r$ then,

$$f^{-1}(X_r)Y_r = X_rf(Y_r) \tag{3.2}$$

is a useful formula in applications of collineations.

Let us now consider how a collineation g transforms under a collineation f. We have already seen in section 1.2 that if Alice uses an element X of the algebra, then according to Bob, it appears as fX. Now, if Alice uses a collineation g, then according to Bob it appears as fgf^{-1}. To prove this, we follow the diagram below.

$$\begin{array}{rcl}
\text{Bob} & \xrightarrow{f} & \text{Alice} \\
fX & & X \\
fgX & & gX \\
g'fX = fgX & & \\
g' = fgf^{-1} & & g
\end{array}$$

Let Alice apply her collineation g to her element X to get a new element gX. This element appears to Bob as fgX. Alternatively, Alice's X appears to Bob as fX and Alice's g appears to Bob as the collineation g' (say). Now, for consistency, $g'fX = fgX$ for all X, so that,

$$g' = fgf^{-1} \tag{3.3}$$

is the way a collineation transforms.

3.3 Dual treatment of a correlation

As already introduced in section 2.6, a correlation is a linear map from points to hyperplanes. Consider two points a and b. A correlation F takes these to two hyperplanes $A = F(a)$ and $B = F(b)$. Consider a point $c = a + b$ on the line ab so that $abc = 0$. By linearity, the correlation takes the point c to the hyperplane $C = F(c) = A + B$. Now AB is the intersection of the hyperplanes A and B, and AB lies within the hyperplane C because $ABC = AB(A+B) = ABA+ABB = 0$ by antisymmetry. We therefore have the result that $0 = abc$ implies that $0 = ABC = F(a)F(b)F(c)$. There is then no inconsistency in assuming that a correlation obeys the outermorphism property $F(abc) = F(a)F(b)F(c)$. A similar argument shows that there is no inconsistency in assuming,

$$F(a_1 \dots a_r) = F(a_1) \dots F(a_r) \ . \tag{3.4}$$

Equation (3.4) shows that a correlation is similar to a collineation. This suggests another argument for the outermorphism property of a correlation. Let us generalize so that a correlation is understood as a transformation between more general observers. With the ubiquitous Alice and Bob as observers of this more general type, when Alice sees points a and b, these appear to Bob as hyperplanes $F(a)$ and $F(b)$. When Alice sees the line ab, Bob sees the subspace $F(a)F(b)$. However, Alice can also transform the line ab directly as $F(ab)$ so that consistency demands the outermorphism property for correlations $F(ab) = F(a)F(b)$. Clearly, this argument can be extended to give equation (3.4).

Let us consider the correlation dually. Let the correlation act on the reference points to give the hyperplanes $B_i = F(a_i)$ for $i = 1 \dots n$. The action of the correlation on the hyperplanes A_i, dual to the reference points, can be obtained from equation (1.18) and the outermorphism property.

$$F(A_i) = F\left(\frac{(-1)^{i-1}a_1 \dots \check{a}_i \dots a_n}{a_1 \dots a_n}\right) = \frac{(-1)^{i-1}F(a_1) \dots F(\check{a}_i) \dots F(a_n)}{F(a_1) \dots F(a_n)}$$
$$= \frac{(-1)^{i-1}B_1 \dots \check{B}_i \dots B_n}{B_1 \dots B_n} = \frac{(-1)^{n-1}(-1)^{i-1}B_n \dots \check{B}_i \dots B_1}{B_n \dots B_1} = (-1)^{n-1}b_i \ .$$

Formula (1.26) has been used to obtain the points b_i which are dual to the hyperplanes $B_1, \dots, B_n$. The index makes sense, because, from equation (1.19), $a_iA_j = \delta_{ij}$ is a number which transforms as a scalar. Therefore, $\delta_{ij} = F(a_iA_j) = (-1)^{n-1}B_ib_j = b_jB_i = \delta_{ij}$. Thus,

$$B_i = Fa_i \text{ and } b_i = (-1)^{n-1}FA_i \tag{3.5}$$

and correlations have been treated dually.

By analogous reasoning to that which led to equation (3.2), we can show that,

$$F^{-1}(U_r)X_r = U_r F(X_r) \tag{3.6}$$

where U_r and X_r are both grade r products of points, or are both grade r products of hyperplanes. Hence $U_r F(X_r)$ is always a number.

Consider a correlation $Y = Fx$ where x is a point and Y is a hyperplane. Clearly, $FFx = FY$ is a point. An involutary correlation is a correlation for which FFx is the same point as x. However, since points are only defined up to an arbitrary weight, the condition for an involutary correlation is $FFx = \xi x$. In order to find the weight, consider xY, which, being a number, transforms as a scalar. Hence,

$$xY = F(xY) = F(x)F(Y) = F(x)\xi x = \xi Yx = (-1)^{n-1}\xi xY$$

and so $FFx = (-1)^{n-1}x$ is the equation of an involutary correlation. As already mentioned in section 2.6, an involutary correlation is also called a polarity[3]. If F is a polarity, and X_r is a grade r product of points, then by the outermorphism property of a correlation, $FFX_r = (-1)^{(n-1)r}X_r = (-1)^{(n-r)r}X_r$. In section 1.9 it is shown that a grade r product of points can be written as a grade $n-r$ product of hyperplanes. Using this result, it is straightforward to show that the last result also holds if X_r is a grade r product of hyperplanes. Therefore, we have the following formulae for a polarity,

$$FFX_r = (-1)^{(n-r)r}X_r \text{ and } F^{-1}X_r = (-1)^{(n-r)r}FX_r \tag{3.7}$$

where X_r is a grade r product of points or a grade r product of hyperplanes.

Let U_r and X_r be grade r products of points (or hyperplanes) and let F be a polarity. Substituting equation (3.7) in equation (3.6) gives,

$$U_r F(X_r) = F^{-1}(U_r)X_r = (-1)^{(n-r)r}F(U_r)X_r = X_r F(U_r) \ . \tag{3.8}$$

The last expression in the above derivation uses the formula (1.13) for changing the order in a product and the fact that FU_r is a grade $n-r$ product of points.

[3]It is also possible to encounter a correlation which obeys $xFx = 0$ for all points x. This means that $(x+y)F(x+y) = 0$ which implies $xFy = -yFx$. If such a correlation is also involutary, so that $FFx = \xi x$, it is called a *null polarity*. In order to fix the constant ξ, note that xFy is a number, so it transforms as a scalar.

$$xFy = F(xFy) = Fx.FFy = \xi Fx.y = (-1)^{n-1}\xi y.Fx = (-1)^n \xi x.Fy$$

Hence, $FFx = (-1)^n x$ is the equation for a null polarity.

The argument leading to equation (3.3) is easily adapted to show that a correlation F transforms in the following way under the action of a collineation f.

$$F' = fFf^{-1} \tag{3.9}$$

3.4 The elliptic polarity

In section 2.6, it was shown how a natural polarity $Ia_1 = A_1$, $Ia_2 = A_2$, $Ia_3 = A_3$ arises in the projective plane. This argument generalises to (n-1)-d projective space as shown below.

Suppose that we represent an arbitrary element x as a product of points,

$$x = \sum_{1 \le j_1 < \ldots < j_r \le n} x_{j_1 \ldots j_r} a_{j_1} \ldots a_{j_r} \ .$$

We could store the array $x_{j_1 \ldots j_r}$ to represent x in a computer. Now, from equation (1.25) we can write,

$$a_{j_1} \ldots a_{j_r} = \frac{A_{j_{r+1}} \ldots A_{j_n}}{A_{j_{r+1}} \ldots A_{j_n} A_{j_1} \ldots A_{j_r}} = \frac{(-1)^{r(n-r)} A_{j_{r+1}} \ldots A_{j_n}}{A_{j_1} \ldots A_{j_n}}$$

and so we can equally well write x as an intersection of hyperplanes,

$$x = \sum \frac{x_{j_1 \ldots j_r} (-1)^{r(n-r)} A_{j_{r+1}} \ldots A_{j_n}}{A_{j_1} \ldots A_{j_n}} = \sum X_{j_{r+1} \ldots j_n} A_{j_{r+1}} \ldots A_{j_n} \ .$$

Therefore, x could also be represented in the computer by the array,

$$X_{j_{r+1} \ldots j_n} = \frac{(-1)^{r(n-r)} x_{j_1 \ldots j_r}}{A_{j_1} \ldots A_{j_n}}$$

where we have assumed that the pseudonumber $A_1 \ldots A_n = 1$ so $A_{j_1} \ldots A_{j_n}$ is a permutation factor equal to ± 1.

Now, if we have a function $x_{j_1 \ldots j_r} \mapsto X_{j_{r+1} \ldots j_n}$ that can map between representations, we can also consider,

$$x \mapsto Ix = \sum X_{j_{r+1} \ldots j_n} a_{j_{r+1}} \ldots a_{j_n} = \sum \frac{(-1)^{r(n-r)} x_{j_1 \ldots j_r}}{A_{j_1} \ldots A_{j_n}} \frac{A_{j_1} \ldots A_{j_r}}{A_{j_1} \ldots A_{j_n}} \ .$$

Comparing coefficients of $x_{j_1 \ldots j_r}$ shows that,

$$I(a_{j_1} \ldots a_{j_r}) = (-1)^{r(n-1)} A_{j_1} \ldots A_{j_r} \tag{3.10}$$

where we have used the fact that $(-1)^{r(n-1)} = (-1)^{r(n-r)}$. The action on points is found by setting $r = 1$ in equation(3.10),

$$Ia_j = (-1)^{n-1} A_j \tag{3.11}$$

so we have a natural correlation that takes points into hyperplanes[4]. Notice that equations (3.10,3.11) are consistent with the outermorphism property of a correlation as expressed by equation (3.4). The action of our natural I on pseudonumbers is found by setting $r = n$ in equation (3.10), so,

$$I(a_1 \ldots a_n) = (-1)^{n(n-1)} A_1 \ldots A_n = A_1 \ldots A_n \tag{3.12}$$

and there is no inconsistency between our assumption that the pseudonumber $a_1 \ldots a_n = A_1 \ldots A_n = 1$ and the fact that a correlation should treat numbers as scalars so that $I(\xi) = \xi$ where ξ is a number. Furthermore, our natural correlation I is involutary so that it is, in fact, a polarity. This can be shown by using equation (3.10) and the duality relations of section 1.9.

$$\begin{aligned} I^2(a_1 \ldots a_r) &= I((-1)^{r(n-1)} A_1 \ldots A_r) = (-1)^{r(n-1)} I(a_{r+1} \ldots a_n) \\ &= (-1)^{r(n-1)} (-1)^{(n-r)(n-1)} A_{r+1} \ldots A_n = (-1)^{r(n-1)} a_1 \ldots a_r \end{aligned}$$

which agrees with the formula (3.7) for a general polarity. The polarity defined by equation (3.11) is called an *elliptic polarity* [9]. By applying I to both sides of (3.11) it can be written more neatly as,

$$IA_j = a_j \ . \tag{3.13}$$

3.5 Congruences

According to Coxeter [9], elliptic geometry and hyperbolic geometry are obtained by starting from projective space and picking a polarity which appears invariant to all observers. Conventionally, an invariant polarity is given the symbol I and is called the absolute polarity. The type of the absolute polarity determines the type of the geometry.

$$\begin{array}{ccc} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ fIf^{-1} & & I \end{array}$$

[4]Setting $n = 3$ reproduces the results of section 2.6 for the case of the projective plane.

Using equation (3.9), the above diagram shows that Alice and Bob will see an invariant polarity I if their viewpoints are connected by a collineation f such that $I = fIf^{-1}$. In other words if $If = fI$ so that the collineation commutes with the absolute polarity. A collineation that commutes with the absolute polarity is called a *congruence* of the space. Congruences are the transformations that connect the viewpoints of observers who all agree on the geometry of the space[5].

The quickest route to the general formula for a congruence is via Lie group theory[6]. The congruences of a space form a group. If g is a congruence, it is regarded as a point in a group manifold. The identity is the special point 1 in the group manifold. In Whitehead's algebra, the group manifold of congruences is completely separate from the underlying space in which the congruences act. Consider a curve g_τ in the group manifold with parameter τ. A tangent vector,

$$\mathcal{A} = \frac{dg_\tau}{d\tau}$$

to the curve is called an element of the Lie algebra[7]. Right translation by a group element f takes an element g of the group to gf. Consider a curve which goes through the identity and has tangent vector $\mathcal{A}$. The curve is,

$$g_{d\tau} = 1 + d\tau\mathcal{A}$$

in the neighbourhood of the identity. Let's right translate this curve by f. Right translation produces a new curve $f + d\tau\mathcal{A}f$ in the neighbourhood of point f. The tangent to this curve is $\mathcal{A}f$. So, we can right translate a vector $\mathcal{A}$ at point 1 to get a new vector $\mathcal{A}f$ at point f. If we allow f to vary then we have a vector field $\mathcal{A}(f) = \mathcal{A}(1)f$, with $\mathcal{A}(1) = \mathcal{A}$, which is called a right invariant vector field. The field lines of a right invariant vector field are a set of curves which, loosely speaking, fill the group manifold. Let f_τ be a field

[5]One might think that the condition $If = \lambda fI$ where λ is a weight defines the most general congruence. However, this just corresponds to multiplying all points by a weight $\mu = \sqrt{\lambda}$. Let $fp = \mu p$ where p is a point. If H is a hyperplane, pH is a number and so $pH = f(pH) = fp.fH = \mu p.fH$ for all p. Hence, $fH = \mu^{-1}H$. Now, $Ifp = \mu Ip$, whilst $fIp = \mu^{-1}Ip$, so that $Ifp = \mu^2 fIp$. However, this congruence has no effect because p is the same point as μp. We can always multiply a congruence obeying $If = fI$ by such a nugatory congruence to get the overall result $If = \lambda fI$. Hence $If = fI$ is the condition for the most general congruence.

[6]The general formula (3.26) was first obtained from the fact that a simple congruence is a product of reflections in two hyperplanes. The connection between the Lie group derivation and the reflection derivation is established in section 3.9.

[7]Script letters are used for elements of the Lie algebra of the group of congruences until we can express them in terms of the elements of Whitehead's algebra.

line. The equation of the curve is,

$$\frac{df_\tau}{d\tau} = \mathcal{A} f_\tau \ . \tag{3.14}$$

The formal solution is $f_\tau = \exp(\tau\mathcal{A})f_0$. If we start from the identity, the curve (which can be found in any book on Lie groups) is,

$$f_\tau = e^{\tau\mathcal{A}}1 = e^{\tau\mathcal{A}} = 1 + \sum_{k=1}^{\infty} \frac{(\tau\mathcal{A})^k}{k!} \ . \tag{3.15}$$

Let us get the form of $\mathcal{A}$ in terms of elements of Whitehead's algebra. If p, q are two points in the underlying space,

$$qIf_\tau p = qf_\tau Ip = f_\tau^{-1}q.Ip$$

where we used equation (3.2) and the fact that the congruence commutes with the absolute polarity. Now substitute for the congruence using equation (3.15),

$$qIe^{\tau\mathcal{A}}p - e^{-\tau\mathcal{A}}q.Ip = 0 \ .$$

Differentiating by τ and setting τ to zero gives the following condition on $\mathcal{A}$.

$$qI\mathcal{A}p + \mathcal{A}q.Ip = 0 \tag{3.16}$$

This condition is satisfied if we set $\mathcal{A}p = ILM.Ip$ where L and M are hyperplanes. Let's check this by substitution. First, we use the outermorphism property of the polarity.

$$qI(ILM.Ip) + (ILM.Iq)Ip = qI^2(LM.p) + I((LM.q)p)$$

We want to reverse the order of terms like $LM.p$ by using the formula (1.13). $LM.p$ multiplies progressively, so we work in point-grades. LM, being the intersection of two hyperplanes, is of point-grade $n-2$, and p is of point-grade 1. Therefore, $LM.p = (-1)^{n-2}p.LM$. Continuing, and also employing equation (3.7) to handle the squared polarity.

$$qI(ILM.Ip) + (ILM.Iq)Ip = (-1)^{n-1}(-1)^{n-2}q(p.LM) + (-1)^{n-2}I((q.LM)p)$$

The terms like $p.LM$ can now be expanded into hyperplanes using equation (1.41).

$$\begin{aligned} qI(ILM.Ip) + (ILM.Iq)Ip &= -q((pM)L - (pL)M) - I(p(q.LM)) \\ &= -(pM)(qL) + (pL)(qM) - I(p((qM)L - (qL)M)) \\ &= -(pM)(qL) + (pL)(qM) - I((qM)(pL) - (qL)(pM)) \end{aligned}$$

The polarity in the last line acts on a number which is a scalar. Therefore, I does not change its argument and we get,

$$qI(ILM.Ip)+(ILM.Iq)Ip = -(pM)(qL)+(pL)(qM)-((qM)(pL)-(qL)(pM)) = 0$$

which shows that $\mathcal{A}p = ILM.Ip$ obeys condition (3.16).

Now L and M are hyperplanes and so IL and IM are points and $ILM = IL.IM$ is a line. Therefore, we have established[8] that an element of the Lie algebra of congruences is an operator, which, whenever it acts on a point, has the form,

$$\mathcal{A} = ILM.I \tag{3.17}$$

where ILM is a line. If we substitute $\mathcal{A}$ into equation (3.15) we get the following formal expression for the action of a simple congruence on a point in Whitehead's algebra.

$$f_\tau p = \exp{(\tau ILM.I)}p \tag{3.18}$$

It is called a simple congruence because products of simple congruences can be used to build more complicated congruences.

The two salient properties of a simple congruence follow from the power series expansion of equation (3.18). Firstly, if p lies in the subspace LM then $p.LM = 0$ and $ILM.Ip = I(LM.p) = 0$ and so every term in the power series is zero except for the leading term so that $f_\tau p = p$. In other words, the congruence acts like the identity on all points in the subspace LM. Secondly, consider a general point p no longer confined to the subspace LM. Since Ip is a hyperplane, equation (1.38) gives $ILM.Ip = (IL.Ip)IM - (IM.Ip)IL = (Lp)IM - (Mp)IL$. This is a linear combination of points IL and IM, so $ILM.Ip$ is always a point on the line ILM for all p. This means that all powers of $ILM.I$ in the expansion of the exponential just produce points on the line ILM. Only the zeroth term p in the exponential can be other than on the line ILM. Therefore, if p itself is on the line ILM, then $f_\tau p$ is also on the line ILM. The line ILM is therefore the *invariant line* of the simple congruence. This line is mapped into itself by the action of the congruence. This is the second salient property.

It is interesting to consider the eigenvalue equation $f_\tau p = \lambda p$ of a simple congruence. The space has n reference points[9] so there are n eigenpoints. Since the congruence acts like the identity in the subspace LM, there will be $n-2$ eigenpoints, each having eigenvalue $+1$ in LM. The remaining two

[8]Strictly speaking, we have not ruled out the possibility that there might be other forms for $\mathcal{A}$ that would obey condition (3.16).

[9]In other words it has dimension $n-1$.

eigenpoints will lie on the invariant line ILM. Their eigenvalues will either be both real or they will be a complex conjugate[10] pair[11].

3.6 Closed form for a simple congruence

We can obtain a closed form expression for a simple congruence by expanding the exponential in equation (3.18) in powers of the operator $ILM.I$. In obtaining the formulae for the powers, we use the expansion formulae of equations (1.38,(1.41), the commutation rule equation (1.13), the rule (3.7) for the square of the polarity and the outermorphism property (3.4) of the polarity . The first power of the operator is,

$$\begin{aligned}(ILM.I)p = ILM.Ip &= (IL.Ip)IM - (IM.Ip)IL \\ &= (-1)^{n-1}((pL)IM - (pM)IL)\ .\end{aligned}$$

The second power is as follows.

$$\begin{aligned}(ILM.I)^2p &= ILM.I[(-1)^{n-1}((pL)IM - (pM)IL)] \\ = (-1)^{n-1}[(pL)ILM.I^2M - (pM)ILM.I^2L] &= (pL)ILM.M - (pM)ILM.L \\ &= (pL)[(IL.M)IM - (IM.M)IL] - (pM)[(IL.L)IM - (IM.L)IL] \\ &= (pL)[(IL.M)IM - (IM.M)IL] - (pM)[(IL.L)IM - (IL.M)IL] \\ &= [(pL)(IL.M) - (pM)(IL.L)]IM + [(pM)(IL.M) - (pL)(IM.M)]IL\end{aligned}$$

The third power is as follows.

$$\begin{aligned}(ILM.I)^3p &= ILM.I\{[(pL)(IL.M) - (pM)(IL.L)]IM \\ &\quad +[(pM)(IL.M) - (pL)(IM.M)]IL\} \\ &= (-1)^{n-1}[(pL)(IL.M) - (pM)(IL.L)]ILM.M \\ &\quad +(-1)^{n-1}[(pM)(IL.M) - (pL)(IM.M)]ILM.L \\ &= (-1)^{n-1}[(pL)(IL.M) - (pM)(IL.L)][(IL.M)IM - (IM.M)IL] \\ &\quad +(-1)^{n-1}[(pM)(IL.M) - (pL)(IM.M)][(IL.L)IM - (IL.M)IL] \\ &= (-1)^{n-1}(pM)[(IL.L)(IM.M) - (IL.M)^2]IL \\ &\quad -(-1)^{n-1}(pL)[(IL.L)(IM.M) - (IL.M)^2]IM \\ &= (-1)^{n-1}(ILM.LM)[(pM)IL - (pL)IM] \\ &= (-1)^{n-1}(ILM.LM)I[(pM)L - (pL)M] \\ &= (-1)^{n-1}(ILM.LM)I(p.LM) = -(ILM.LM)I(LM.p) \\ &= -(ILM.LM)(ILM.I)p\end{aligned}$$

[10] conjugate!complex
[11] See equations (3.45,3.46).

The third power is therefore simply related to the first power by,

$$(ILM.I)^3 p = -(ILM.LM)(ILM.I)p = -(ILM.LM)(ILM.Ip) \ . \quad (3.19)$$

Before getting a similar relation between the fourth power and the second power, we digress to obtain a simple formula for the second power.

$$\begin{aligned}(ILM.I)^2 p &= ILM.I(ILM.Ip) = (-1)^{n-1} ILM.(LM.p) \\ &= (-1)^{n-1}(-1)^{n-2} ILM.(p.LM) = -ILM.(p.LM) \end{aligned} \quad (3.20)$$

Using this result and equation (3.19), we obtain the formula for the fourth power.

$$(ILM.I)^4 p = -(ILM.LM)(ILM.I)^2 p = (ILM.LM) ILM.(p.LM) \quad (3.21)$$

Equations (3.19) and (3.21) give a recursive way of writing down all the odd and even powers in the expansion of the exponential form of the congruence in equation (3.18).

$$\begin{aligned} f_\tau = \exp(\tau ILM.I) &= \sum_{k=0}^{\infty} \frac{\tau^k}{k!}(ILM.I)^k \\ &= 1 + \sum_{k=1}^{\infty} \frac{\tau^{2k}}{(2k)!}(-ILM.LM)^{k-1}(ILM.I)^2 \\ &+ \sum_{k=1}^{\infty} \frac{\tau^{2k-1}}{(2k-1)!}(-ILM.LM)^{k-1}(ILM.I) \\ &= 1 + \cosh(\tau\sqrt{-ILM.LM})\left(\frac{ILM.I}{\sqrt{-ILM.LM}}\right)^2 \\ - \left(\frac{ILM.I}{\sqrt{-ILM.LM}}\right)^2 &+ \sinh(\tau\sqrt{-ILM.LM})\left(\frac{ILM.I}{\sqrt{-ILM.LM}}\right) \end{aligned}$$

The way things stand, τ depends on the weights of L, M. However, if we make the replacement $\tau \to \tau/\sqrt{-ILM.LM}$ then τ is independent of these weights. We then have,

$$\begin{aligned} f_\tau = \exp\left(\frac{\tau ILM.I}{\sqrt{-ILM.LM}}\right) &= 1 + \cosh(\tau)\left(\frac{ILM.I}{\sqrt{-ILM.LM}}\right)^2 \\ - \left(\frac{ILM.I}{\sqrt{-ILM.LM}}\right)^2 &+ \sinh(\tau)\left(\frac{ILM.I}{\sqrt{-ILM.LM}}\right) \ . \end{aligned}$$

Now let the congruence act on a point p and substitute the formula (3.20) for the second power of the operator.

$$\begin{aligned} f_\tau p = \exp\left(\frac{\tau ILM.I}{\sqrt{-ILM.LM}}\right) p = \left(p - \frac{ILM.(p.LM)}{ILM.LM}\right) \\ + \cosh(\tau)\left(\frac{ILM.(p.LM)}{ILM.LM}\right) + \sinh(\tau)\left(\frac{ILM.Ip}{\sqrt{-ILM.LM}}\right) . \end{aligned} \tag{3.22}$$

Equation (3.22) is a closed-form for a simple congruence, so we could stop the derivation at this point. However, it is possible to get a neater form by using an equation which shows how an arbitrary point depends on the line ILM and the subspace LM. We therefore interrupt the derivation of the formula for a simple congruence in order to get this equation for a point.

Let c be any point in the subspace LM so that $c.LM = 0$. Since c lies in the intersection of hyperplanes L and M, then[12] c must lie on L and also on M so that $cL = 0$ and $cM = 0$. Let p be an arbitrary point. It can be written in the form $p = \lambda IL + \mu IM + \gamma c$ where λ, μ and γ are unknown weights. We now solve for the weights. Multiplying on the left by the point IL and then on the right by LM, we get,

$$\begin{aligned} (IL.p).LM = \mu ILM.LM + \gamma(IL.c).LM \\ = \mu ILM.LM + \gamma((IL.L)(cM) - (IL.M)(cL)) = \mu ILM.LM \end{aligned}$$

where $(IL.c).LM$ has been expanded using equation (1.43). This gives $\mu = (IL.p).LM/ILM.LM$. Similarly, we can get λ by multiplying on the right by IM and then on the right again by LM to give $\lambda = (p.IM).LM/ILM.LM$. Finally, by multiplying on the right by the hyperplane Ic and using outermorphism we find $\gamma = pIc/cIc$. A general point is therefore,

$$p = \left(\frac{(p.IM).LM}{ILM.LM}\right) IL - \left(\frac{(p.IL).LM}{ILM.LM}\right) IM + \left(\frac{pIc}{cIc}\right) c . \tag{3.23}$$

Using equation (3.23), we can get a useful formula for the contribution in the subspace LM as follows.

$$\begin{aligned} \left(\frac{pIc}{cIc}\right) c = p - \left(\frac{(p.IM).LM}{ILM.LM}\right) IL + \left(\frac{(p.IL).LM}{ILM.LM}\right) \\ = \frac{(ILM.LM)p - ((p.IM).LM)IL + ((p.IL).LM)IM}{ILM.LM} \end{aligned}$$

[12]More formally, using equation (1.41), $0 = c.LM = (cM)L - (cL)M$ so that $cM = 0$ and $cL = 0$ since L and M are independent.

The last line can be simplified by using equation (1.38) to give,

$$\left(\frac{pIc}{cIc}\right)c = \frac{(p.IL.IM).LM}{ILM.LM} = \frac{(p.ILM).LM}{ILM.LM} \ . \tag{3.24}$$

We can also get a formula for the contribution from the line ILM by expanding and re-arranging the first two terms on the RHS of equation (3.23).

$$\begin{aligned}
&[(p.IM).LM]IL - [(p.IL).LM]IM \\
&= [(pL)(IM.M) - (pM)(IL.M)]IL - [(pL)(IL.M) - (pM)(IL.L)]IM \\
&= (pM)[(IL.L)IM - (IL.M)IL] - (pL)[(IL.M)IM - (IM.M)IL] \\
&= (pM)(ILM.L) - (pL)(ILM.M) = ILM.((pM)L - (pL)M) \\
&= ILM.(p.LM)
\end{aligned}$$

On substituting this result for the first two terms on the RHS of equation (3.23) and replacing the third term using equation (3.24), we obtain[13],

$$p = \frac{ILM.(p.LM)}{ILM.LM} + \frac{(p.ILM).LM}{ILM.LM} \ . \tag{3.25}$$

The first term on the RHS of equation (3.25) is a point on the line ILM and the second term is a point in the subspace LM. If p is in the subspace LM, then $p.LM = 0$ and we are left with the second term. Conversely, if p is on the line ILM then $p.ILM = 0$ and we are left with the first term.

We now take the final step in the derivation of the formula for a simple congruence. Recall that the derivation was interrupted at equation (3.22) in order to obtain equation (3.25). We now simply replace the first term on the RHS of equation (3.22) using equation (3.25). This gives the following canonical formula[14] for a simple congruence in Whitehead's algebra.

$$\begin{aligned}
f_\tau p = \exp\left(\frac{\tau ILM.I}{\sqrt{-ILM.LM}}\right)p = \left(\frac{(p.ILM).LM}{ILM.LM}\right) \\
+ \cosh(\tau)\left(\frac{ILM.(p.LM)}{ILM.LM}\right) + \sinh(\tau)\left(\frac{ILM.Ip}{\sqrt{-ILM.LM}}\right)
\end{aligned} \tag{3.26}$$

[13] This formula is a special case of the general projection formula studied in section 6.4.

[14] Formula (3.26) is not in Whitehead's book [10] and does not appear to be known to the researchers active in this field in the late 19th century. Sir R. S. Ball provides bibliographical notes in his treatise on screw theory [15] (a screw is just a general congruence) which show that he had read widely the works of his contemporaries. For example, he had read Whitehead's book. Now, the puzzling thing about Ball's work on screws is that he restricts the study to infinitesimal motions of his bodies. It seems unlikely that he would have done this, if he had known of a closed-form of a simple congruence.

With this result, Whitehead's algebra is seen to be a strikingly austere structure, without any superflous constructs. The single antisymmetric product enables us to intersect and join subspaces, and by specifying an absolute polarity we can handle congruences of the resulting spaces. In contrast, one often sees the quaternion product introduced specifically to handle congruences, as for example in Coxeter's book [9].

Another form of the congruence, which is simpler to evaluate, is obtained by replacing the coefficient of $\cosh\tau$ in equation (3.26) with equation (3.25).

$$f_\tau p = \exp\left(\frac{\tau ILM.I}{\sqrt{-ILM.LM}}\right)p = (1-\cosh(\tau))\left(\frac{(p.ILM).LM}{ILM.LM}\right) + \cosh(\tau)p + \sinh(\tau)\left(\frac{ILM.Ip}{\sqrt{-ILM.LM}}\right) \quad (3.27)$$

Equations (3.26) and (3.27) blow up when $ILM.LM = 0$. This pathological case is considered later in section 3.10.

3.6.1 Example: Congruences of the elliptic plane

Let a_1, a_2, a_3 be reference points for the projective plane, with A_1, A_2, A_3 the dual hyperplanes[15] so that $a_jA_k = \delta_{jk}$. In the plane, the hyperplanes are lines. The projective plane is made into an elliptic space by introducing an absolute polarity of the form $Ia_1 = A_1$, $Ia_2 = A_2$ and $Ia_3 = A_3$. This is the natural polarity in the plane that was found in section 2.6. In the elliptic plane, let the invariant line of a congruence be $ILM = a_2a_3$. The subspace $LM = I^2LM = Ia_2a_3 = A_2A_3 = a_1/a_1a_2a_3$. Since $a_1a_2a_3$ is a pseudonumber which just weights the point, we see that the subspace LM is the point a_1. Hence a_1 is an eigenpoint of the congruence. We get the action of the congruence by evaluating formula (3.27).

$$f_\tau p = \exp(-i\tau a_2a_3.I)p = (1-\cosh(\tau))pa_2a_3.A_2A_3 + \cosh(\tau)p - i\sinh(\tau)a_2a_3.Ip$$

In the above equation $i = \sqrt{-1}$ and so we introduce the parameter $\tau = i\phi$ so that the congruence is explicitly real. Now we work out the action of the congruence on the reference points by replacing p by each reference point in turn.

$$fa_1 = \exp(\phi a_2a_3.I)a_1 = (1-\cos(\phi))(a_1a_2a_3)A_2A_3 + \cos(\phi)a_1 + \sin(\phi)a_2a_3.A_1 = (1-\cos(\phi))a_1 + \cos(\phi)a_1 + 0 = a_1$$

[15] See equation (1.19).

$$\begin{aligned} fa_2 = \exp(\phi a_2 a_3.I)a_2 = (1-\cos(\phi))(a_2 a_2 a_3)A_2 A_3 \\ +\cos(\phi)a_2 + \sin(\phi)a_2 a_3.A_2 = \cos(\phi)a_2 + \sin(\phi)a_3 \\ fa_3 = \exp(\phi a_2 a_3.I)a_3 = (1-\cos(\phi))(a_3 a_2 a_3)A_2 A_3 \\ +\cos(\phi)a_3 + \sin(\phi)a_2 a_3.A_3 = \cos(\phi)a_3 - \sin(\phi)a_2 \end{aligned}$$

This is a rotation about the eigenpoint a_1 through an angle[16] ϕ. This is the general form of all congruences in the elliptic plane. Although it looks as if we have chosen a particular case, the invariant line ILM could be an arbitrary line, and we just took the reference points a_2 and a_3 on this line.

3.6.2 Example: Natural coordinates for the elliptic plane

We can get natural coordinates for the elliptic plane by exponentiating a single reference point. For example, p, given by,

$$\begin{aligned} p &= \exp(\phi a_2 a_3.I)\exp(\theta a_1 a_2.I)a_1 = \exp(\phi a_2 a_3.I)(\cos(\theta)a_1 + \sin(\theta)a_2) \\ &= \cos(\theta)a_1 + \sin(\theta)\cos(\phi)a_2 + \sin(\theta)\sin(\phi)a_3 \end{aligned} \tag{3.28}$$

is an arbitrary point in the elliptic plane parameterized by the coordinates θ and ϕ.

3.7 Dual treatment of a simple congruence

Equation (3.26) gives the action of a congruence on a point p. The principle of duality demands that we also give the action of a congruence on a hyperplane. Write the hyperplane as $H = Ip$. Using equation (3.15) and the fact that a congruence commutes with the absolute polarity,

$$f_\tau H = f_\tau Ip = If_\tau p = Ie^{\tau\mathcal{A}}p = H + \sum_{k=1}^{\infty}\frac{\tau^k}{k!}I\mathcal{A}^k p$$

where the operator $\mathcal{A}$ is given, in un-normalised form, by equation (3.17). In order to get the final result, we have to remember to normalize the invariant line ILM by dividing it by $\sqrt{-ILM.LM}$. This simplifies the equations and is legitimate because $ILM.LM$ is not changed by the action of the polarity as it is a scalar number. Let's calculate the first few powers. The first power is $I\mathcal{A}p = I(ILM.Ip) = LM.IH$. The second power is,

$$\begin{aligned} I\mathcal{A}^2 p = I\mathcal{A}(\mathcal{A}p) = I(ILM.I\mathcal{A}p) = I(ILM.(LM.IH)) = LM.I(LM.IH) \\ = (LM.I)^2 H \end{aligned}$$

[16] Angle and distance are defined formally in section 3.8.

and so the pattern is clear. The formula for the action of a congruence on a hyperplane is therefore,

$$f_\tau H = H + \sum_{k=1}^{\infty} \frac{\tau^k}{k!}(LM.I)^k H = e^{\tau LM.I} H \ .$$

If we now put the normalisation back into $\mathcal{A}$, apply the polarity to the RHS of formula (3.26), use the outermorphism property and change the order of some factors, we obtain the following formula.

$$\begin{aligned} f_\tau H = \exp\left(\frac{\tau LM.I}{\sqrt{-ILM.LM}}\right) H = \left(\frac{ILM.(HLM)}{ILM.LM}\right) \\ + \cosh(\tau)\left(\frac{(ILM.H).LM}{ILM.LM}\right) + (-1)^{n-2}\sinh(\tau)\left(\frac{IH.LM}{\sqrt{-ILM.LM}}\right) \end{aligned} \tag{3.29}$$

This is the canonical formula for the action of a simple congruence on a hyperplane. It is the dual formula to equation (3.26).

3.8 Distance and angle

Let us define the distance between two points a and b. Set $a = IL$ and $b = IM$, so that the invariant line of a congruence is $ILM = ab$. The distance is defined as the value of the congruence parameter τ that takes a to b. In other words, set $p = IL = a$ and $f_\tau p = \mu IM = \mu b$. The unknown weight μ is used because the action of the congruence on $IL = a$ produces the point $b = IM$ with an unimportant weight. Using the congruence formula (3.26) we get the following equation for the distance τ.

$$\mu IM = \cosh(\tau)\left(\frac{ILM.(IL.LM)}{ILM.LM}\right) + (-1)^{n-1}\sinh(\tau)\left(\frac{ILM.L}{\sqrt{-ILM.LM}}\right)$$

The coefficient of IL on the RHS must be zero. Upon expanding the terms, we find,

$$0 = \cosh(\tau) - (-1)^{n-1}\sinh(\tau)\left(\frac{IL.M}{\sqrt{-ILM.LM}}\right) \ .$$

Putting $a = IL$ and $b = IM$ gives the distance from a to b as τ where,

$$\tanh(\tau) = \frac{\sqrt{-ILM.LM}}{(-1)^{n-1}IL.M} = \frac{\sqrt{-ab.Iab}}{aIb} \ . \tag{3.30}$$

In obtaining this result we used $LM = (-1)^{(n-1)2}LM = I^2LM = Iab$. By writing $\tanh\tau = (e^\tau - e^{-\tau})/(e^\tau + e^{-\tau})$ and solving for e^τ, we obtain the following alternate form[17] for the distance from a to b.

$$e^{2\tau} = \frac{aIb + \sqrt{-abIab}}{aIb - \sqrt{-abIab}} \tag{3.32}$$

We have just found the τ such that $f_\tau IL = \mu IM$. Acting with I on the left gives $f_\tau L = \mu M$ since I commutes with f_τ. Therefore, without any further calculation, the τ given by equation (3.30) takes hyperplane L to hyperplane M and is therefore also the angle between the hyperplanes. From (3.30) and the calculation leading to (3.32), the alternate form of the angle between the hyperplanes L and M is τ given by,

$$e^{2\tau} = \frac{IL.M + (-1)^{n-1}\sqrt{-ILM.LM}}{IL.M - (-1)^{n-1}\sqrt{-ILM.LM}} \tag{3.33}$$

We can usually ignore the sign $(-1)^{n-1}$ because it just changes the direction of motion $L \to M$ or $M \to L$.

3.8.1 Example: The line element

Equation (3.28) showed an example of coordinates for the elliptic plane. In the general case of a $(n-1)$-dimensional space, an arbitrary point $p(\theta^1, \ldots, \theta^{n-1})$ can be parameterized by $(n-1)$ coordinates labelled by superscripts. By varying the coordinates, a neighbouring point is,

$$p + dp = p + \frac{\partial p}{\partial \theta^k} d\theta^k \tag{3.34}$$

with the summation convention in force on repeated pairs of indices. Note that, in Whitehead's algebra, dp is *not* a displacement from p to $p + dp$. The infinitesimal dp is in fact just an ordinary point, with a very small intrinsic weight, which just draws the point $p + dp$ a little way towards itself and away from p as shown in figure 1.

The formula for the distance $d\tau$ from p to $p + dp$ can be found by writing $a = p$ and $b = p + dp$ in equation (3.30) and keeping only terms of the first order.

$$d\tau = \tanh(d\tau) = \frac{\sqrt{-p(p+dp).Ip(p+dp)}}{pI(p+dp)} = \frac{\sqrt{-pdp.Ipdp}}{pIp} \tag{3.35}$$

[17] Also note the following useful form of the distance which is easily obtained from equation (3.30).

$$\cosh^2(\tau) = \frac{(aIb)^2}{(aIa)(bIb)} \tag{3.31}$$

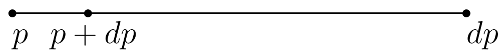

Figure 3.1: The meaning of an infinitesimal in Whitehead's algebra

As an example of this formula, using the natural coordinates for the elliptic plane from equation (3.28), the line element is found to be,

$$(d\sigma)^2 = (d\theta)^2 + \sin^2(\theta)(d\phi)^2 \tag{3.36}$$

where $\tau = i\sigma$ and $i = \sqrt{-1}$.

3.8.2 Example: The angle between two lines

Consider a triangle with vertices a, b, c. Let's obtain a formula for the angle $\angle bac$. Now, following section 3.8, $\angle bac$ can be found by finding a congruence that leaves point a invariant and rotates the line ab into the line ac. The value of the congruence parameter is defined as the angle. The simplest way to proceed is to assume that b and c lie on the hyperplane Ia so that we have the conditions $aIb = 0$ and $aIc = 0$. This does not change the angle because we can move points b and c along their respective lines from a until they hit the hyperplane Ia. Let the congruence be f and take its invariant line to be bc. The subspace in which the congruence acts like the identity is then Ibc. Now a lies in this subspace because $a.Ibc = (aIc)Ib - (aIb)Ic = 0$. Hence a is invariant so $fa = a$. Now take the congruence parameter to be such that $fb = c$. With this choice the congruence rotates line ab into line ac because $fab = fa.fb = ac$. The congruence parameter τ that takes b to c is, from equation (3.31),

$$\cosh^2(\tau) = \frac{(bIc)^2}{(bIb)(cIc)} \tag{3.37}$$

so that τ is the angle between the lines. Now, suppose that the points are general points and do not obey the conditions $aIb = aIc = 0$. Since we assumed that b and c moved along their respective lines until they hit Ia, we just need to move the points $b \to ab.Ia$ and $c \to ac.Ia$. Making these substitutions in equation (3.37) gives, after a little calculation,

$$\cosh^2(\tau) = \frac{(abIac)^2}{(abIab)(acIac)} \; . \tag{3.38}$$

This formula gives the angle between two lines ab and ac in any space with polarity I.

3.9 Products of reflections in two hyperplanes

Congruences were treated in section 3.5 from the stand-point of Lie group theory. However, a simple congruence can also be introduced as the result of reflections in two successive hyperplanes. We now establish the equivalence between these two approaches.

3.9.1 Reflections

A reflection of a point p in a hyperplane L is defined as follows. Consider the point IL. The line pIL intersects the hyperplane L at the point $pIL.L = (pL)IL - (IL.L)p$, and so we may write,

$$p = \frac{(pL)IL}{IL.L} - \frac{pIL.L}{IL.L} \ .$$

The harmonic conjugate[18] of p is defined as the reflection[19],

$$fp = \frac{(pL)IL}{IL.L} + \frac{pIL.L}{IL.L} = \frac{(pL)IL + (pL)IL - (IL.L)p}{IL.L} = \frac{2(pL)IL}{IL.L} - p \tag{3.39}$$

of point p in the hyperplane L.

The reflection is an involutary collineation,

$$f^2p = \frac{2(fp.L)IL}{IL.L} - fp = \frac{2(pL)IL}{IL.L} - \frac{2(pL)IL}{IL.L} + p = p$$

and so $f = f^{-1}$. The action of a reflection on a hyperplane H can be found with the aid of equation (3.2), which, in this context is $fp.H = f^{-1}p.H = pfH$. Upon substituting for the reflection using equation (3.39),

$$\frac{2(pL)(IL.H)}{IL.L} - pH = pfH$$

and since this holds for all p,

$$fH = \frac{2(IL.H)L}{IL.L} - H \ . \tag{3.40}$$

Reflections are important because they are congruences, since they commute with the absolute polarity. In order to see this, set $H = Ip$ in equation

[18] See the end of section 1.15.

[19] Coxeter [9] calls this transformation a *harmonic homology* instead of a reflection.

(3.40).

$$\begin{aligned} fIp &= \frac{2(IL.Ip)L}{IL.L} - Ip = \frac{(-1)^{n-1}2(pL)L}{IL.L} - Ip \\ &= \frac{2(pL)I^2L}{IL.L} - Ip = I\left(\frac{2(pL)IL}{IL.L} - p\right) = Ifp \end{aligned}$$

3.9.2 Double reflections

Consider the action of a reflection in hyperplane L, followed by a reflection in hyperplane M. Using equation (3.39), with an obvious notation,

$$\begin{aligned} f_M f_L p &= \frac{2(f_L p.M)IM}{IM.M} - f_L p \\ = \frac{2}{IM.M}\left(\frac{2(pL)IL.M}{IL.L} - pM\right)IM &- \frac{2(pL)IL}{IL.L} + p \\ = p - \frac{2(pL)IL}{IL.L} - \frac{2(pM)IM}{IM.M} &+ \frac{4(pL)(IL.M)IM}{(IL.L)(IM.M)} \ . \end{aligned} \tag{3.41}$$

We have to show that the canonical form (3.26) of a simple congruence is in fact the same as the double reflection formula (3.41).

Take equation (3.26) and substitute identity (3.25). This gives a third form of the simple congruence,

$$\begin{aligned} f_\tau p = \exp\left(\frac{\tau ILM.I}{\sqrt{-ILM.LM}}\right)p = p + (\cosh(\tau) - 1)\left(\frac{ILM.(p.LM)}{ILM.LM}\right) \\ + \sinh(\tau)\left(\frac{ILM.Ip}{\sqrt{-ILM.LM}}\right) \end{aligned} \tag{3.42}$$

which begins to look like the double reflection (3.41) because its leading term is p.

The angle between the reflection hyperplanes L, M will turn out to be half of the congruence parameter. Therefore, twice the angle between the reflection hyperplanes will be equal to the congruence parameter. Now, equation (3.33) directly gives a formula for twice the angle between the reflection hyperplanes. Therefore, the RHS of equation (3.33) is equal to e^τ where τ is the congruence parameter in equation (3.42). We are therefore going to use,

$$e^\tau = \frac{IL.M + (-1)^{n-1}\sqrt{-ILM.LM}}{IL.M - (-1)^{n-1}\sqrt{-ILM.LM}}$$

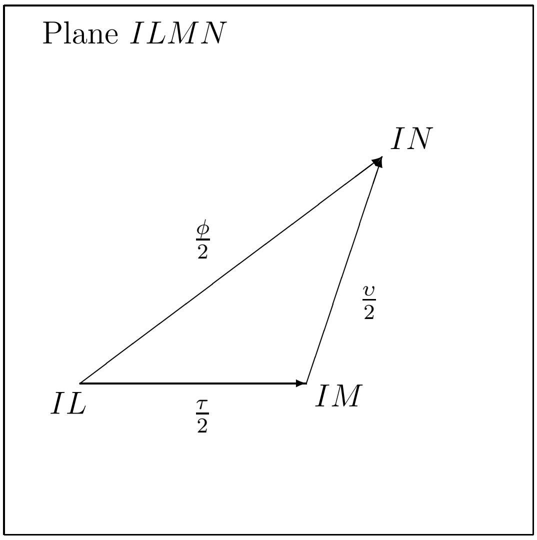

Figure 3.2: Donkin's theorem

to substitute for τ in equation (3.42). Solving for $\cosh(\tau)$ and $\sinh(\tau)$ we get,

$$\cosh(\tau) - 1 = -\left(\frac{2(ILM.LM)}{(IL.L)(IM.M)}\right)$$
$$\sinh(\tau) = (-1)^{n-1}\left(\frac{2(IL.M)\sqrt{-ILM.LM}}{(IL.L)(IM.M)}\right) .$$

Upon substituting these expressions for $\cosh(\tau)$ and $\sinh(\tau)$ into equation (3.42) and expanding the resulting expressions using the fundamental formulae in section 1.14, we recover, exactly, the double reflection formula (3.41). This shows that a double reflection is the same congruence that one gets from exponentiating an element of the Lie algebra, provided that the angle between the reflection hyperplanes is half the congruence parameter in the exponential.

3.9.3 Example: Donkin's theorem

The equivalence between a double reflection and a simple congruence can be used to prove Donkin's theorem [9] on the composition of two simple congruences. Consider three hyperplanes L, M, N. The reflections in the hyperplanes are denoted by f_L, f_M and f_N. The composite congruence $f_N f_M f_M f_L = f_N f_L$ because a reflection is involutary so $f_M f_M = 1$. The double reflection $f_M f_L$ is given by equation (3.41) and this has been shown to be a simple congruence of the form of (3.26) with the invariant line ILM

and congruence parameter τ equal to twice the angle between the hyperplanes L and M. In order to remind ourselves that $f_M f_L$ is a simple congruence let us write $f_\tau = f_M f_L$. Similarly, the simple congruence with invariant line IMN is $f_\upsilon = f_N f_M$ and υ is the congruence parameter which is twice the angle between hyperplanes M and N. Finally, the simple congruence with invariant line ILN is $f_\phi = f_N f_L$ and ϕ is the congruence parameter which is twice the angle between hyperplanes L and N. We therefore have the result $f_\phi = f_\upsilon f_\tau$. The result has a simple interpretation in terms of the triangle shown in figure 3.2 with vertices IL, IM and IN. From section 3.8, the angle $\tau/2$ between hyperplanes L and M is the same number as the distance between points IL and IM. Therefore, in figure 3.2 the sides of the triangle are the invariant lines of the simple congruences, and the lengths of the sides are each one half of the corresponding congruence parameter. Using Donkin's theorem, one can find the invariant line $ILIN$ and parameter ϕ of the result of combining two congruences $f_\phi = f_\upsilon f_\tau$ by constructing the triangle of figure 3.2. Of course, in order for Donkin's theorem to work, the invariant lines of the two congruences being combined must lie in a plane. In other words, the theorem does not work if the invariant lines are a pair of skew lines.

3.10 The parabolic congruence

It has already been observed that formula (3.26) for the simple congruence blows up when $ILM.LM = 0$. In order to get a formula which is valid in this case, we return to equation (3.18) and expand the exponential in a power series. Equations (3.19) and (3.21) show that the series stops after the quadratic term.

$$\begin{aligned} f_\tau p = \exp(\tau ILM.I)p &= p + \tau(ILM.Ip) + \frac{\tau^2}{2}(ILM.I(ILM.Ip)) \\ &= p + (-1)^{n-2}\tau I(p.LM) - \frac{\tau^2}{2}(ILM.(p.LM)) \end{aligned} \tag{3.43}$$

This formula can be written in a more illuminating way if we take a digression to look at the geometrical significance of the number $ILM.LM$.

From the polarity, we can construct a number pIp. If we constrain the point p to satisfy $pIp = 0$, it picks out a hypersurface called the *quadric locus* [9]. The quadric locus is drawn schematically as the ellipse in figure 3.3. The sign of $ILM.LM$ determines whether or not the invariant line ILM intersects the quadric. In order to see this, use the point $p = IL + \mu IM$ which moves along the line ILM by varying the weight μ. The points of

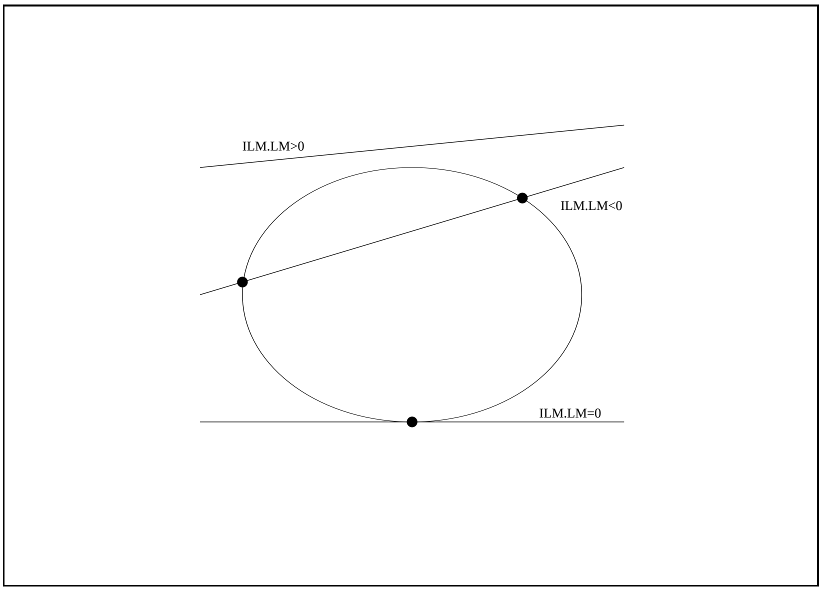

Figure 3.3: Elliptic, hyperbolic and parabolic invariant lines

intersection of the line ILM with the quadric are the solutions of $pIp = 0$. This condition gives the quadratic $\mu^2 IM.M + 2\mu IL.M + IL.L = 0$. The discriminant is $\sqrt{(IL.M)^2 - (IL.L)(IM.M)} = \sqrt{-ILM.LM}$. If $ILM.LM > 0$, the invariant line does not intersect the quadric in any real points and the line is called an *elliptic line*. If $ILM.LM < 0$, the invariant line intersects the quadric in two real points and the line is called a *hyperbolic line*. If $ILM.LM = 0$, there is only one point of intersersection and so the invariant line is a tangent to the quadric. This is the *parabolic* case[20].

Returning to formula (3.43) for the parabolic congruence, let us take IL as the single point at which the invariant line touches the quadric so that $IL.L = 0$. The hyperplane L is, in fact, the tangent hyperplane to the quadric at IL. In order to see this, assume that this is not the case so that L cuts the quadric. Let p be one of the other points where L cuts the quadric in addition to IL. The point p is on the quadric so $pIp = 0$ and it is also on the hyperplane L so $pL = 0$. The line pIL is now, by assumption, not a parabolic line, so,

$$0 \neq pIL.I(pIL) = (-1)^{n-1}pIL.(Ip.L) = (-1)^{n-1}((pIp)(IL.L) - (pL)(Lp)) = 0$$

which is a contradiction. Therefore, if IL lies on the quadric so that $IL.L = 0$, then the hyperplane L is the tangent hyperplane to the quadric at point IL. The invariant line ILM of the parabolic congruence, being a tangent to the quadric at IL, also lies in the tangent hyperplane L. Therefore, the point IM must lie on the tangent hyperplane so that $IL.M = 0$. We now expand the RHS of the parabolic congruence (3.43) using the conditions $IL.L = IL.M = 0$.

$$\begin{aligned} f_\tau p &= \exp(\tau ILM.I)p \\ &= p + (-1)^{n-2}\tau I((pM)L - (pL)M) - \frac{\tau^2}{2} ILM.((pM)L - (pL)M) \\ &= p + (-1)^{n-2}\tau((pM)IL - (pL)IM) - \frac{\tau^2}{2}(pL)(IM.M)IL \end{aligned} \tag{3.47}$$

[20]The points of intersection with the quadric are also the two eigenpoints of the congruence on the invariant line. To see this, take the case where the invariant line ILM of a congruence f_τ cuts the quadric locus $pIp = 0$. If p is a point on the quadric, then since pIp is a number it transforms as a scalar so,

$$0 = f_\tau(pIp) = f_\tau p.f_\tau Ip = f_\tau p.If_\tau p\ . \tag{3.44}$$

In other words if p is on the quadric so is $f_\tau p$. The action of a congruence is always to move a point within the quadric locus. Now, if p is also on the invariant line of the congruence, the action of the congruence is to move the point along the invariant line. These two actions of the congruence are only compatible if p is an eigenpoint so that $f_\tau p = \lambda p$. The formula (3.27) for f_τ is simplified if we take the points IL and IM that define the invariant line ILM as the two eigenpoints so that $IL.L = 0$ and $IM.M = 0$. Upon evaluating (3.27) we find,

$$f_\tau IL = \exp((-1)^n\tau)IL \tag{3.45}$$
$$f_\tau IM = \exp((-1)^{n-1}\tau)IM\ . \tag{3.46}$$

In the elliptic case, the two eigenpoints are a complex conjugate pair because the invariant line does not cut the quadric at any real points. In the parabolic case the two eigenpoints are not separated.

This is the canonical form of the parabolic congruence in Whitehead's algebra. Notice that, in this formula, we have not normalised the congruence parameter τ. Consequently, a change in the weight of point IM would change the parabolic congruence unless we made a compensating change in the value of τ. If we make the replacement $\tau \to \tau/\sqrt{M.IM}$ then τ is independent of the weight of M. We cannot do anything about the weight of L because $IL.L = 0$ and so it cannot be normalized.

3.10.1 Example: Flat space geometry on the quadric

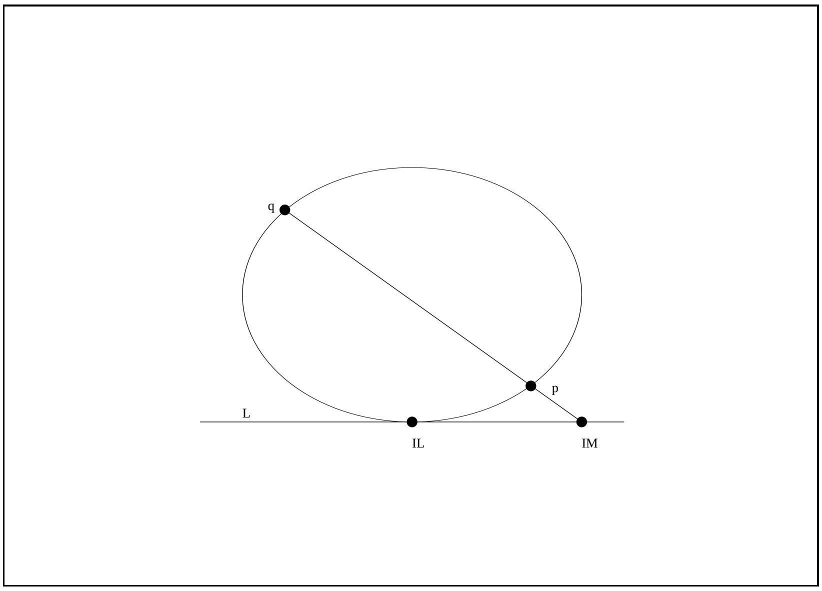

Figure 3.4: Construction for Euclidean distance on the quadric

Using the approach to geometric algrebra developed by David Hestenes, which is based on Clifford algebra [2, 3], references [22, 3] have developed a model of 3-d Euclidean geometry on a null surface in a 5-d space which has much to recommend it for computational work. We now provide an introduction to how this model looks in Whitehead's algebra, because it is an interesting application of the parabolic congruence. The model is based on the remarkable result due to Wachter[21] that the geometry on the absolute quadric of hyperbolic space is Euclidean. This fact is also known to general relativists under the guise of Lemaître's frame in de Sitter space-time[22]. In

[21]Wachter's result is theorem 11.31 in [9].

[22]In Lemaître's frame, space-like hypersurfaces of 4-d de Sitter space-time are flat 3-d

fact, it turns out that the intrinsic geometry on any quadric hypersurface is flat-space geometry, so we do not have to specify that the quadric is in elliptic or hyperbolic space. Later, in chapter 7, we will develop our own model for 3-d Euclidean computational geometry on the absolute quadric hypersurface of 4-d elliptic space. We will use this variant because, as section 3.4 has shown, the elliptic polarity is the most natural in Whitehead's algebra.

In figure 3.4, the ellipse represents the absolute quadric in an $(n-1)$-dimensional space. Let f be a congruence of the space. If p lies on the quadric so that $pIp = 0$ then, since f and I commute, $fp.Ifp = fp.fIp = f(pIp) = 0$ and so fp also lies on the quadric. The quadric is therefore an invariant hypersurface and we can study the intrinsic geometry of the quadric hypersurface.

Let us obtain the formula for the distance between two points p and q on the quadric hypersurface[23]. Pick a point IL on the quadric. It will turn out to represent infinity. Let IM be the point at which the line pq cuts the tangent hyperplane L. Thus, $IM = pq.L = (pL)q - (qL)p$. Now there is a natural parabolic congruence with invariant line ILM. We will adjust the parameter of this congruence so that it carries p to q. The value of the congruence parameter will be the distance from p to q in the quadric hypersurface. We require $q = f_\tau p$ up to some weight. Now, in figure 3.4, the point q is on the line pIM, so q cannot depend on the point IL. Therefore, the value of the parameter τ that carries p to q can be found from the requirement that the coefficient of IL is zero in equation (3.47).

$$0 = (-1)^{n-2}\tau(pM) - \frac{\tau^2}{2}(pL)(IM.M) \tag{3.48}$$

Evaluating pM and $IM.M$,

$$\begin{aligned} M &= (-1)^{n-1}I^2M = (-1)^{n-1}(pL)Iq - (-1)^{n-1}(qL)Ip \\ IM.M &= (-1)^{n-1}((pL)q - (qL)p)((pL)Iq - (qL)Ip) = 2(-1)^n(pL)(qL)(pIq) \\ pM &= (-1)^{n-1}(pL)(pIq) \end{aligned}$$

and substituting into condition (3.48), gives,

$$\tau = \frac{(-1)^{n-1}}{(pL)(qL)} \ .$$

Euclidean spaces. Lemaître's frame is described in section 6 of [18] or section 5.2 of [17]. Also, see equation (5.20) of the current text.

[23] A straightforward application of equation (3.30) gives the distance between the two points in the embedding space. This is ∞ because the quadric is the hypersurface at infinity of the space.

However, as it stands, this is not the distance formula because we have failed to take account of the fact that τ is not normalised in equation (3.47). When we make the replacement $\tau \to \tau/\sqrt{M.IM}$ and substitute for $M.IM = (-1)^{n-1}IM.M$ we get the formula for the distance between two points p and q on the absolute quadric.

$$\tau = (-1)^{n-1}\sqrt{\frac{-2(pIq)}{(pL)(qL)}} \ . \tag{3.49}$$

This formula makes sense because $\tau = 0$ if p and q are coincident. Furthermore, if either point is set to IL, the denominator blows up and $\tau = \infty$, which justifies our assertion that IL represents infinity.

It is not clear from equation (3.49) that the geometry on the quadric hypersurface is that of a flat space. In order to see this, consider the absolute quadric hypersurface $pIp = 0$. Take point IL as the point at infinity and a new point IN as the point representing the origin. Since both points are on the quadric we have $IL.L = 0$ and $IN.N = 0$. The hyperplanes L and N are tangent hyperplanes to the quadric at the points IL and IN respectively. The intersection LN is a subspace of grade $n-2$. Let IM be a point in LN so $0 = IM.LN = (IM.N)L - (IM.L)N$ and therefore $IM.L = 0$ and $IM.N = 0$ since L and N are independent hyperplanes[24] . We can *generate* an arbitrary point p on the quadric by moving the origin IN to p by the action of a parabolic congruence with invariant line ILM. From equation (3.47), with $\tau = 1$,

$$p = \exp(ILM.I)IN = IN + (-1)^{n-1}(IN.L)(IM - \frac{1}{2}(M.IM)IL) \ . \tag{3.50}$$

Notice that by setting $\tau = 1$ the size of the movement is controlled by the intrinsic weight of the point IM which we call the *generator* of point p. Equation (3.50) is a useful way of expressing a point on the quadric. Let's now work out the distance from the origin point IN to the point p using formula (3.49). Upon substituting for p using (3.50), we find,

$$\tau = (-1)^{n-1}\sqrt{\frac{-2(IN.Ip)}{(IN.L)(pL)}} = (-1)^{n-1}\sqrt{IM.M} \ . \tag{3.51}$$

As a concrete example, consider the 3-dimensional projective space with reference points $a_1, \ldots, a_4$ and dual hyperplanes (which are planes) $A_1, \ldots, A_4$ so that $a_jA_k = \delta_{jk}$. We make this into a 3-dimensional hyperbolic space by

[24]Notice that, by equation (3.8), these terms are symmetric, for example $IL.M = IM.L$.

introducing the absolute polarity $Ia_1 = A_1, Ia_2 = A_2, Ia_3 = A_3, Ia_4 = -A_4$. We take the point at infinity as $IL = a_3 + a_4$ and the origin as $IN = a_3 - a_4$. The planes $L = (-1)^{4-1} I^2 L = -A_3 + A_4$ and $N = -A_3 - A_4$. The point at infinity is on the quadric since $IL.L = (a_3 + a_4)(-A_3 + A_4) = -1 + 1 = 0$. The origin is also on the quadric since $IN.N = 0$. Now we take the generator as $IM = xa_1 + ya_2$. The generator lies in $LN = 2A_3A_4$ as required since $(xa_1 + ya_2).A_3A_4 = 0$. Upon substituting in equation (3.51), we obtain $\tau^2 = x^2 + y^2$. This is the elementary formula for the Euclidian distance on a plane. It shows that the intrinsic geometry on the quadric surface in 3-dimensional hyperbolic space is, in fact, Euclidean plane geometry.

In the general case, we can take $n - 2$ reference points a_i to span the subspace LN. The generator can then be written $IM = \sum x_i a_i$ where the x_i are numbers. Then $M = (-1)^{n-1} \sum x_i (Ia_i)$ and equation (3.51) gives the squared distance from the origin IN to point p as,

$$\tau^2 = M.IM = (-1)^{n-1} \sum x_i x_j Ia_j.a_i = \sum x_i x_j (a_i Ia_j) \; . \tag{3.52}$$

This shows that we can have any flat-space metric $\eta_{ij} = a_i Ia_j$ by choosing the form of the polarity I in the generating subspace LN.

3.11 The relation of a congruence to a general collineation

Recall from sections 3.2 and 3.5 that a collineation or linear operator is a linear map of points to points and a congruence is a collineation that commutes with the absolute polarity. The aim of the current section is to determine the relation between the two concepts. In order to do this we need to first explore the concept of the adjoint of an operator and then the eigenvalue problem for self-adjoint operators.

3.11.1 Matrix notation for a collineation

Suppose we have a collineation f which acts in an (n-1)-d projective space with a set of arbitrary reference points $a_1, \ldots, a_n$ and a set of dual hyperplanes $A_1, \ldots, A_n$ such that $a_i A_j = \delta_{ij}$. The matrix elements of the collineation are the numbers $a_i.fA_j$. This definition ensures that the matrix elements for the composition gf of two collineations obey the ordinary rules of matrices[25].

$$a_i.gfA_j = a_i.g\left(\sum_k A_k(a_k.fA_j)\right) = \sum_{k=1}^{n}(a_i.gA_k)(a_k.fA_j) \tag{3.53}$$

3.11.2 The adjoint of a collineation

Given a collineation f, the adjoint $\bar{f}$ is a collineation which obeys,

$$IX.\bar{f}Y = IY.fX \tag{3.54}$$

for all hyperplanes X and Y. By setting $X = A_i$ and $Y = A_j$ and assuming the elliptic polarity (3.13) so that $IX = IA_i = a_i$ and $IY = IA_j = a_j$, the matrix elements of the adjoint are,

$$a_i.\bar{f}A_j = a_j.fA_i \ . \tag{3.55}$$

In other words, for the elliptic polarity, the adjoint is the matrix transpose.

In a case in which the reference points and dual hyperplanes do not obey the relation (3.13), the adjoint is not the matrix transpose[26]. The reason for this is because when equation (3.13) is true, the matrix elements $IA_i.A_j = a_i.A_j = \delta_{ij}$ form the identity matrix. However, when (3.13) does not hold, all we can say is that the numbers $IA_i.A_j$ form a symmetric matrix since $IA_i.A_j = IA_j.A_i$ using property (3.8). The reference points a_i and their dual hyperplanes defined by (1.18) still obey $a_iA_j = \delta_{ij}$. Now a_iA_j is a number and so it remains invariant under the action of the polarity. Hence,

$$\delta_{ij} = a_iA_j = I(a_iA_j) = Ia_i.IA_j = \sum_k Ia_i.a_k(IA_j.A_k) = \sum_k (Ia_i.a_k)(IA_k.A_j) \tag{3.56}$$

so that the inverse of matrix $IA_i.A_j$ is given by the matrix $Ia_i.a_j$. We now calculate the matrix form of the adjoint.

$$\begin{aligned} IA_k.\bar{f}A_j &= IA_j.fA_k \\ \sum_l (IA_k.A_l)(a_l.\bar{f}A_j) &= \sum_l (IA_j.A_l)(a_l.fA_k) \end{aligned}$$

[25] If we had used the numbers $fa_i.A_j$ for the matrix elements of f then since $fa_i.A_j = a_i.f^{-1}A_j$ and $g^{-1}f^{-1} = (fg)^{-1}$ then we would have had a non-standard rule for matrix multiplication.

[26] There are two cases in which this can occur. Firstly, when the polarity is not elliptic. Section 5.4 gives an example in which the polarity is hyperbolic. Secondly, when the polarity is elliptic but the reference points and dual hyperplanes do not obey the neat relation (3.13). Equation (9.3) gives an example of this case.

$$\begin{aligned}
\sum_{kl}(Ia_i.a_k)(IA_k.A_l)(a_l.\bar{f}A_j) &= \sum_{kl}(Ia_i.a_k)(IA_j.A_l)(a_l.fA_k) \\
\sum_{l}\delta_{il}(a_l.\bar{f}A_j) &= a_i.\bar{f}A_j = \sum_{kl}(Ia_i.a_k)(IA_j.A_l)(a_l.fA_k) \\
a_i.\bar{f}A_j &= \sum_{kl}(Ia_i.a_k)(a_l.fA_k)(IA_l.A_j) \qquad (3.57)
\end{aligned}$$

In matrix notation, equation (3.57) can be written $\bar{f} = h^{-1}f^T h$ where h denotes the matrix of elements $IA_i.A_j$ and f^T is the transpose of the matrix of f. Notice that when (3.13) holds true, h is the identity matrix and we recover equation (3.55) which says that the adjoint is the same as the transpose $\bar{f} = f^T$.

We now obtain some properties of the adjoint. The adjoint of the adjoint is,

$$IX.\overline{(\bar{f})}Y = IY.\bar{f}X = IX.fY$$

so that since the above equation holds for all hyperplanes X, Y,

$$\overline{(\bar{f})} = f \ . \qquad (3.58)$$

Using (3.8) to justify swopping the order in expressions such as $IX.Y = IY.X$, we obtain the following result for the adjoint of the composition of a pair of collineations f, g.

$$IX.\overline{gf}Y = IY.gfX = IfX.\bar{g}Y = I\bar{g}Y.fX = IX.\bar{f}\bar{g}Y$$

Hence,

$$\overline{gf} = \bar{f}\bar{g} \ . \qquad (3.59)$$

Setting $g = f^{-1}$ in (3.59) we obtain $1 = \bar{f}\,\overline{f^{-1}}$ so that,

$$\overline{f^{-1}} = (\bar{f})^{-1} \ . \qquad (3.60)$$

If f is a congruence, it commutes with the absolute polarity so that,

$$IX.\bar{f}Y = IY.fX = f^{-1}IY.X = If^{-1}Y.X = IX.f^{-1}Y$$

and therefore,

$$\bar{f} = f^{-1} \ . \qquad (3.61)$$

The reason that the adjoint is a useful concept is because all observers are in agreement about it.

$$\begin{array}{rcc}
\text{Bob} & \xrightarrow{f} & \text{Alice} \\
fgf^{-1} & & g \\
f\bar{g}f^{-1} & & \bar{g} \\
\overline{fgf^{-1}} = f\bar{g}f^{-1} & &
\end{array}$$

For, following the above diagram, suppose observers Alice and Bob are connected by a congruence f so that if Alice uses a collineation g, it appears to Bob as fgf^{-1}. Now Alice takes her adjoint $\bar{g}$. Bob sees this as $f\bar{g}f^{-1}$. However, Bob could also take the adjoint of fgf^{-1} which is,

$$\overline{fgf^{-1}} = \overline{f^{-1}}\bar{g}\bar{f} = f\bar{g}f^{-1} \tag{3.62}$$

using equations (3.59), (3.60) and (3.61). Equation (3.62) shows that Bob's two versions of Alice's adjoint coincide so that the adjoint is a concept that all observers agree upon.

3.11.3 The eigenvalue problem for self-adjoint operators

$$\begin{array}{ccc} \text{Bob} & \xrightarrow{U} & \text{Alice} \\ UBU^{-1} = D & & B \end{array}$$

Suppose Alice has collineation/operator B and so Bob sees this as $UBU^{-1} = D$. Furthermore, suppose that the transformation U can be chosen so that D is diagonal. This means that in a set of reference points a_i with dual hyperplanes A_i such that $a_iA_j = \delta_{ij}$, Bob finds $DA_i = \lambda_i A_i$. This is the form of the eigenvalue problem for operator B because the matrix form of the eigenvalue equation,

$$BU^{-1} = U^{-1}D \tag{3.63}$$

is,

$$\begin{aligned} \sum_j (a_i.BA_j)(a_j.U^{-1}A_k) &= \sum_j (a_i.U^{-1}A_j)(a_j.DA_k) = \sum_j (a_i.U^{-1}A_j)\lambda_k(a_jA_k) \\ &= \sum_j (a_i.U^{-1}A_j)\lambda_k\delta_{jk} = (a_i.U^{-1}A_k)\lambda_k \ . \end{aligned} \tag{3.64}$$

The kth column of matrix $a_i.U^{-1}A_k$ is the eigenvector with eigenvalue λ_k of the matrix $a_i.BA_j$. The eigenvectors are mutually orthogonal because,

$$\begin{aligned} \sum_j (a_j.U^{-1}A_k)(a_j.U^{-1}A_l) &= \sum_j (IA_j.U^{-1}A_k)(a_j.U^{-1}A_l) \\ &= \sum_j (IU^{-1}A_k.A_j)(a_j.U^{-1}A_l) = IU^{-1}A_k.U^{-1}A_l \\ &= UIU^{-1}A_k.A_l = IA_k.A_l = a_kA_l = \delta_{kl} \ . \end{aligned}$$

In the above derivation of orthogonality, we used equations (3.13), (3.2) and the fact that congruence U commutes with the polarity I. Of course not all operators B can be diagonalised by a change of reference frame in this way. Since the diagonal operator is manifestly self-adjoint $D = \bar{D}$ then,

$$\begin{aligned} B &= U^{-1}DU \\ \bar{B} &= \overline{U^{-1}DU} = \bar{U}\bar{D}\overline{U^{-1}} = U^{-1}DU = B \end{aligned}$$

and so B itself must be self-adjoint.

3.11.4 Decomposition of a general collineation

Suppose we have a collineation B that *does not commute* with the absolute polarity I. Consider the operator $\bar{B}B$. This is self-adjoint because $\overline{\bar{B}B} = \bar{B}(\overline{\bar{B}}) = \bar{B}B$. Consequently, from section 3.11.3, we know that the self-adjoint operator $\bar{B}B$ can be seen as diagonal from Bob's frame.

$$\begin{array}{ccc} \text{Bob} & \xrightarrow{U} & \text{Alice} \\ U\bar{B}BU^{-1} = D & & \bar{B}B \end{array}$$

Since U is a congruence, it commutes with I and so from equation (3.61) $\bar{U} = U^{-1}$. Then, from the above diagram,

$$\bar{B}B = U^{-1}DU = \bar{U}D^{1/2}D^{1/2}U \tag{3.65}$$

where the square root of the diagonal operator $DA_i = \lambda_i A_i$ is defined by $D^{1/2}A_i = \lambda_i^{1/2}A_i$. It is easily checked that the solution to (3.65) is,

$$B = V^{-1}D^{1/2}U = \bar{V}D^{1/2}U \tag{3.66}$$

where V is a congruence, so it commutes with the polarity and hence $\bar{V} = V^{-1}$. This is the generalised *singular value decomposition* of an operator. In the case of an elliptic polarity, where the adjoint is the same as the transpose, it recovers the standard form of the singular value decomposition encountered in books on linear algebra [25].

As an example, take the operator B as the Lie algebra element $B = LM.I$ that appears in formula (3.29) for a congruence acting on a hyperplane. This operator does not have an inverse because it just projects hyperplanes into the subspace LM. The adjoint of this operator is easily shown to be $\bar{B} = -B$ from the definition (3.54) of the adjoint. It is also interesting to see this in another way. A congruence is of the form $\exp(B)$ and from the power series expansion of the exponential, the adjoint of a congruence is $\overline{\exp(B)} = \exp(\bar{B})$. Now, from (3.61), we know that the adjoint of the congruence is the inverse $\exp(-B)$ and therefore the adjoint of the Lie algebra element is $\bar{B} = -B$.

Now we can show that $B = LM.I$ can be decomposed into the form (3.66). Since only the complete subspace LM appears in the operator $B = LM.I$, there is no loss in generality in taking $IL.L = IM.M = 1$ and $IL.M = 0$. Set up the congruences U and V so that, $L = U^{-1}A_1 = (-1)^{n-2}V^{-1}A_2$ and $M = U^{-1}A_2 = (-1)^{n-1}V^{-1}A_1$. The diagonal operator $D^{1/2}$ is taken as $D^{1/2}A_1 = A_1$ and $D^{1/2}A_2 = A_2$ with $D^{1/2}A_i = 0$ for $i = 3, \ldots, n$. Now, $BL = LM.IL = (-1)^{n-1}M$ and $BM = (-1)^{n-2}L$ and for X in the dual subspace ILM then $BX = LM.IX = 0$. These facts are all reproduced by the form $V^{-1}D^{1/2}U$ as can be shown by straightforward calculation.

Chapter 4

Calculus in Whitehead's Algebra

4.1 Volume in Whitehead's algebra

Suppose that Alice assigns coordinates $\theta_1^A, \ldots, \theta_{n-1}^A$ to a point. She tentatively proposes to use the pseudonumber,

$$dV = (p^A I p^A)^{-n/2} p^A \frac{\partial p^A}{\partial \theta_1^A} \cdots \frac{\partial p^A}{\partial \theta_{n-1}^A} d\theta_1^A \ldots d\theta_{n-1}^A$$

as her volume element. In this expression p^A stands for $p(\theta_1^A, \ldots, \theta_{n-1}^A)$. The formula for dV makes sense because the number $(p^A I p^A)^{-n/2}$ removes any dependence on the weight of the point. However, if it is to have physical significance, it must be the same volume element used by Bob. Let's work out how Bob sees Alice's volume dV. Let f be the congruence connecting the view-points of Alice and Bob. Bob sees Alice's p^A as fp^A. Now, Bob has his own system of coordinates θ_j^B. Therefore, in his own system, fp^A is p^B. However, an unimportant weight λ might come into play, so, in general $fp^A = \lambda p^B$. The weight λ can be found by considering $p^A I p^A = fp^A . f I p^A = fp^A . I f p^A = \lambda^2 p^B I p^B$. Thus,

$$\lambda = \sqrt{\frac{p^A I p^A}{p^B I p^B}} \, .$$

Now let's start transforming Alice's dV. The partial derivatives commute with f because if δ denotes a small change, $\delta fp = f(p + \delta p) - fp = f\delta p$.

$$f dV = \lambda^{-n} (p^B I p^B)^{-n/2} \lambda p^B \frac{\partial \lambda p^B}{\partial \theta_1^A} \cdots \frac{\partial \lambda p^B}{\partial \theta_{n-1}^A} d\theta_1^A \ldots d\theta_{n-1}^A$$

Antisymmetry kills off all the derivatives of λ so,

$$fdV = (p^B I p^B)^{-n/2} p^B \frac{\partial p^B}{\partial \theta_1^A} \cdots \frac{\partial p^B}{\partial \theta_{n-1}^A} d\theta_1^A \ldots d\theta_{n-1}^A \ .$$

Now use the chain rule to change to derivatives with respect to Bob's coordinates. The result is,

$$fdV = (p^B I p^B)^{-n/2} p^B \frac{\partial p^B}{\partial \theta_1^B} \cdots \frac{\partial p^B}{\partial \theta_{n-1}^B} J d\theta_1^A \ldots d\theta_{n-1}^A$$

where the J is the Jacobian. Since $d\theta_1^B \ldots d\theta_{n-1}^B = J d\theta_1^A \ldots d\theta_{n-1}^A$, we get,

$$fdV = (p^B I p^B)^{-n/2} p^B \frac{\partial p^B}{\partial \theta_1^B} \cdots \frac{\partial p^B}{\partial \theta_{n-1}^B} d\theta_1^B \ldots d\theta_{n-1}^B$$

and so Bob sees Alice's volume dV as fdV and fdV turns out to be exactly the same volume element that he himself would use. Hence both observers find that their definitions of volume are mutually consistent. However, it does not mean that they necessarily agree on the value of the volume. Since dV is a pseudonumber, $fdV = \det(f^{-1})dV$. Both observers agree on how to measure volume, but space might be expanded or contracted for one of them. For the congruences given by equation (3.26), ILM and LM together span the projective space, and one can show that the product of the eigenvalues for the two eigenpoints on the invariant line ILM is unity. Since the congruence acts like the identity in LM, it follows that the determinant is always unity. Hence, $fdV = dV$ and both observers agree on the definition of volume, and on the volume computed for a region. Therefore, in Whitehead's algebra, the volume element is defined as,

$$dV = (pIp)^{-n/2} p \frac{\partial p}{\partial \theta_1} \cdots \frac{\partial p}{\partial \theta_{n-1}} d\theta_1 \ldots d\theta_{n-1} \ . \tag{4.1}$$

The notation,

$$pd^{n-1}p = p \frac{\partial p}{\partial \theta_1} \cdots \frac{\partial p}{\partial \theta_{n-1}} d\theta_1 \ldots d\theta_{n-1} \tag{4.2}$$

will often be used to simplify the writing of the un-normalised part of a volume element.

Finally, using antisymmetry of pp, we can also write the volume element (4.1) in normalised form,

$$dV = \frac{p}{\sqrt{pIp}} \frac{\partial}{\partial \theta_1} \left(\frac{p}{\sqrt{pIp}} \right) \cdots \frac{\partial}{\partial \theta_{n-1}} \left(\frac{p}{\sqrt{pIp}} \right) d\theta_1 \ldots d\theta_{n-1} \ . \tag{4.3}$$

4.1.1 Example: Area of the elliptic plane

Using the expression for a point given by equation (3.28), we find that the area of the elliptic plane is,

$$\int dS = \int (pIp)^{-3/2} p \frac{\partial p}{\partial \theta} \frac{\partial p}{\partial \phi} d\theta d\phi = a_1 a_2 a_3 \int_0^\pi \sin\theta d\theta \int_0^\pi d\phi = 2\pi a_1 a_2 a_3 \ . \tag{4.4}$$

4.2 The derivative with respect to a point

4.2.1 The equation of a smooth hypersurface

We motivate the introduction of the derivative with respect to a point by considering the equation of a hypersurface. We have already encountered the quadric hypersurface $pIp = 0$. In general, a hypersurface is given by the equation $\phi(p) = 0$ where $\phi(p)$ is a number-valued function of a point p. Now p and λp, where λ is a numerical weight, represent the same point. Therefore if p is on the hypersurface so that $\phi(p) = 0$, then λp must obey $\phi(\lambda p) = 0$ because λp is also on the hypersurface. The only way this can happen is if ϕ is a homogeneous function of p so that $\phi(\lambda p) = \lambda^m \phi(p)$ where m is the degree of the homogeneous function. Suppose we are in a projective space with reference points $a_1, \ldots, a_n$ and dual hyperplanes $A_1, \ldots, A_n$ so that $a_i A_j = \delta_{ij}$. A general point can be written $p = \sum \xi_i a_i$ where the ξ_i are numbers. Therefore we can write the function as $\phi(p) = \phi(\xi_1, \ldots, \xi_n)$ and by Euler's theorem on homogeneous functions,

$$m\phi(\xi_1, \ldots, \xi_n) = \sum_{i=1}^{n} \xi_i \frac{\partial \phi(\xi_1, \ldots, \xi_n)}{\partial \xi_i} \ . \tag{4.5}$$

Now, if we have a hyperplane $H = \sum \eta_i A_i$, we can form the number $pH = \sum \xi_i \eta_i$. This suggests that the derivatives in equation (4.5) are, in fact, components of a hyperplane that we can denote by $\partial\phi/\partial p$. We therefore tentatively define,

$$\frac{\partial \phi(p)}{\partial p} = \sum_{i=1}^{n} \frac{\partial \phi(\xi_1, \ldots, \xi_n)}{\partial \xi_i} A_i \ .$$

and since ϕ is an arbitrary number-valued function, the *derivative with respect to a point* is,

$$\frac{\partial}{\partial p} = \sum_{i=1}^{n} A_i \frac{\partial}{\partial \xi_i} = \sum_{i=1}^{n} A_i \frac{\partial}{\partial (pA_i)} \ . \tag{4.6}$$

Before accepting this definition we have to check that it does not depend on the particular choice of reference hyperplanes on the RHS. Introducing a new set of reference points b_i and dual hyperplanes B_i with $b_iB_j = \delta_{ij}$, we check this as follows.

$$\begin{aligned}
\frac{\partial}{\partial p} &= \sum_i A_i \frac{\partial}{\partial(pA_i)} = \sum_{ij} A_i \frac{\partial(pB_j)}{\partial(pA_i)} \frac{\partial}{\partial(pB_j)} = \sum_{ijk} A_i \frac{\partial(pA_k)(a_kB_j)}{\partial(pA_i)} \frac{\partial}{\partial(pB_j)} \\
&= \sum_{ijk} A_i \delta_{ki}(a_kB_j) \frac{\partial}{\partial(pB_j)} = \sum_{ij} A_i(a_iB_j) \frac{\partial}{\partial(pB_j)} = \sum_j B_j \frac{\partial}{\partial(pB_j)}
\end{aligned}$$

4.2.2 How does the derivative transform?

The following diagram indicates how a point p, a hyperplane H and the derivative $\partial/\partial p$ transform under a collineation f.

$$\begin{array}{ccc}
\text{Bob} & \xrightarrow{f} & \text{Alice} \\
fp & & p \\
fH & & H \\
\dfrac{\partial}{\partial fp} & & \dfrac{\partial}{\partial p}
\end{array}$$

Bob's derivative can be written,

$$\frac{\partial}{\partial fp} = \sum_i A_i \frac{\partial}{\partial(fp.A_i)} = \sum_i A_i \frac{\partial}{\partial(p.f^{-1}A_i)} = f\sum_i (f^{-1}A_i) \frac{\partial}{\partial(p.f^{-1}A_i)} = f\frac{\partial}{\partial p}$$

because the hyperplanes $f^{-1}A_i$ are just another set of hyperplanes, an alternative set to the A_i, and we know from the end of section 4.2.1 that the derivative does not depend on the particular set of hyperplanes used in its definition. Hence we find that the derivative transforms like a hyperplane.

$$\frac{\partial}{\partial fp} = f\frac{\partial}{\partial p} \tag{4.7}$$

In order to show that (4.7) makes sense, consider the action of the derivative on a number-valued function of a point $\phi(p)$.

$$\begin{array}{rcl}
\text{Bob} & \xrightarrow{f} & \text{Alice} \\
\phi(f^{-1}.) & & \phi(.) \\
\phi(f^{-1}fp) = \phi(p) & & \phi(p)
\end{array}$$

The first line of the above diagram shows that the scalar field $\phi(.)$ transforms as $\phi(f^{-1}.)$ where the dot indicates a slot for a point. The second line shows that this transformation law for the scalar field does in fact mean that the value of the field at a point is a scalar.

$$\begin{array}{ccc} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ f\dfrac{\partial}{\partial p} & & \dfrac{\partial}{\partial p} \\ f\dfrac{\partial\phi(p)}{\partial p} & & \dfrac{\partial\phi(p)}{\partial p} \end{array}$$

The above diagram shows that the hyperplane $\partial\phi(p)/\partial p$ transforms correctly as a hyperplane should when we use rule (4.7) and the fact that $\phi(p)$ transforms as a scalar.

4.2.3 The derivative for arbitrary coordinates

In the calculation of the volume element in section 4.1, a point in (n-1)-d projective space was written as a general function of $n-1$ coordinates in the form $p = p(\theta_1, \dots, \theta_{n-1})$. Let us now obtain the form of the derivative for these general coordinates. In order to make things compatible with definition (4.6) we need n coordinates instead of $n-1$. So, we actually write,

$$\begin{aligned} p &= \theta_0 e_0(\theta_1, \dots, \theta_{n-1}) \\ e_0 &= \frac{\partial p}{\partial \theta_0}, \dots, e_i = \frac{\partial p}{\partial \theta_i}, \dots, e_{n-1} = \frac{\partial p}{\partial \theta_{n-1}} \ . \end{aligned} \tag{4.8}$$

The coordinate θ_0 is just the weight attached to the point, so it has no physical significance. The $e_0, \dots, e_{n-1}$ form a set of n reference points for the (n-1)-d projective space. Consequently, using formula (1.18), we can find a set of dual hyperplanes $E_0, \dots, E_{n-1}$ with the property $e_i E_j = \delta_{ij}$. Calculating,

$$\sum_i E_i \frac{\partial}{\partial \theta_i} = \sum_{ij} E_i \frac{\partial(pA_j)}{\partial \theta_i} \frac{\partial}{\partial(pA_j)} = \sum_{ij} E_i (e_i A_j) \frac{\partial}{\partial(pA_j)} = \sum_j A_j \frac{\partial}{\partial(pA_j)} = \frac{\partial}{\partial p}$$

we find that the derivative is given by,

$$\frac{\partial}{\partial p} = \sum_i E_i \frac{\partial}{\partial \theta_i} \ . \tag{4.9}$$

Therefore, equations (4.9,4.6) are compatible versions of the derivative. Using (4.9), we can get an explicit formula for the dual hyperplanes.

$$\frac{\partial \theta_j}{\partial p} = \sum_i E_i \frac{\partial \theta_j}{\partial \theta_i} = \sum_i E_i \delta_{ij} = E_j \tag{4.10}$$

For applications, it is useful to keep in mind the result,

$$p\frac{\partial}{\partial p} = \theta_0 e_0.\sum_i E_i \frac{\partial}{\partial \theta_i} = \theta_0 \sum_i \delta_{0i}\frac{\partial}{\partial \theta_i} = \theta_0 \frac{\partial}{\partial \theta_0} \ . \tag{4.11}$$

This result may look a little strange at first, but any suspicion is put to rest by the following derivation of Euler's theorem on homogeneous functions[1].

$$\begin{aligned} p\frac{\partial}{\partial p}\phi(p) &= \theta_0 \frac{\partial}{\partial \theta_0}\phi(\theta_0 e_0) = \theta_0 \frac{\partial}{\partial \theta_0}\theta_0^m \phi(e_0) \\ &= m\theta_0^m \phi(e_0) = m\phi(\theta_0 e_0) = m\phi(p) \end{aligned} \tag{4.12}$$

We can now perform two consistency checks on the formula (4.10) for the dual hyperplanes. Firstly, we know that the coordinate $\theta_0 = \theta_0(p)$ must be a homogeneous function of degree 1. This follows by multiplying p in (4.8) by a weight λ. It is clear that the coordinate θ_0 gets changed according to $\theta_0 \mapsto \lambda\theta_0$. Hence, from Euler's theorem on homogeneous functions (4.12),

$$p\frac{\partial \theta_0}{\partial p} = \theta_0$$

and this says $pE_0 = \theta_0$ which is correct because $p = \theta_0 e_0$ and $e_0 E_0 = 1$ by definition of the dual hyperplanes. The second consistency check follows from noting that the coordinates θ_i for $i = 1, \ldots, n-1$ must be homogeneous functions of p of degree zero. This is because the weight of a point has no physical significance so that the coordinates - with the exception of θ_0 - cannot depend on the weight of a point. Consequently, from (4.12),

$$p\frac{\partial \theta_i}{\partial p} = 0 \text{ for } i \neq 0 \tag{4.13}$$

and this says that $e_0 E_i = 0$ for $i \neq 0$. This is already known to be correct and so gives further confidence that $\partial/\partial p$ makes sense.

Another useful formula is,

$$\frac{\partial p}{\partial \theta_j}\frac{\partial}{\partial p} = e_j.\sum_i E_i \frac{\partial}{\partial \theta_i} = \sum_i \delta_{ji}\frac{\partial}{\partial \theta_i} = \frac{\partial}{\partial \theta_j} \ . \tag{4.14}$$

which also shows that equation (4.11) is just (4.14) for the index $j = 0$. Upon multiplying (4.14) by a small change $\delta\theta_j$ in the coordinates and summing, we find the following formula for the operator that produces a small change.

$$\delta p.\frac{\partial}{\partial p} = \sum_j \delta\theta_j \frac{\partial p}{\partial \theta_j}.\frac{\partial}{\partial p} = \sum_j \delta\theta_j \frac{\partial}{\partial \theta_j} = \delta \tag{4.15}$$

[1] Compare equation (4.5).

Equation (4.15) furnishes an easy way to evaluate derivatives[2]. For example, to evaluate the derivative of pH where H is a constant hyperplane, we vary pH as follows,

$$\delta p.\frac{\partial(pH)}{\partial p} = \delta(pH) = \delta p.H$$

and note that this equation holds for all variations δp so that we must have,

$$\frac{\partial(pH)}{\partial p} = H \ . \tag{4.16}$$

We shall also need the derivative of pIp where I is the absolute polarity.

$$\delta p.\frac{\partial(pIp)}{\partial p} = \delta(pIp) = \delta p.Ip + p.I\delta p = 2\delta p.Ip$$

If this is to hold for all variations δp then,

$$\frac{\partial(pIp)}{\partial p} = 2Ip \ . \tag{4.17}$$

Finally, we note an amusing notational device that works for the derivative with respect to a point. Suppose we expand a hyperplane X in terms of the reference hyperplanes E_i. The formula is $X = \sum_i E_i(e_i X)$ and so we can regard each term $E_i e_i(.)$ as an operator with a slot for a hyperplane. Thus $E_i e_i$ acts on a hyperplane X to give the hyperplane $E_i(e_i X)$. From this point of view, which is, of course, nothing other that Dirac's bra and ket notation, the sum $\sum_i E_i e_i$ is the identity operator because $\sum_i E_i(e_i X) = X$. So we can write $\sum_i E_i e_i = 1$ and we get the following rather pleasing result[3].

$$\frac{\partial}{\partial p}p = \sum_i E_i \frac{\partial}{\partial \theta_i} p = \sum_i E_i \frac{\partial p}{\partial \theta_i} = \sum_i E_i e_i = 1 \tag{4.18}$$

4.3 Stokes's Theorem

The general form of an integral[4] is,

$$\int \mathcal{L}(p, pd^{n-1}p) = \int \mathcal{L}(p, e_0 e_1 \dots e_{n-1} \theta_0 d\theta_1 \dots d\theta_{n-1})$$

[2]The chain rule (9.35) for the derivative of a function of a function is also derived from (4.15).

[3]This result can also be obtained by letting (4.15) act on p to the right.

[4]This section and the example on parallel transport which follows can be omitted in a first reading.

where $\mathcal{L}$ is some function which is linear in the second slot which contains the un-normalised volume element[5]. It also seems to be the case that, in physical applications, $\mathcal{L}$ has no overall dependence on the weight of the variable point p. This is another way of saying that the point p participates in the function in a normalized form, usually as $p/\surd(pIp) = e_0/\surd(e_0Ie_0)$. This seems to be necessary, because otherwise the value of the integral would change by simply applying a physically meaningless re-weighting to the points. So, let us assume that $\mathcal{L}(p, pd^{n-2}p)$ has no overall dependence on the weight θ_0.

The form of Stokes's theorem in Whitehead's algebra is,

$$\int_V \mathcal{L}\left(p, pd^{n-1}p.\frac{\partial}{\partial p}\right) = -\int_S \mathcal{L}(p, pd^{n-2}p) \tag{4.19}$$

where V is a region (generalized volume) and S is the boundary of V. It is a straightforward calculation. Starting on the LHS of (4.19), we substitute for $\partial/\partial p$ using (4.9) and then use the linearity of the second slot in $\mathcal{L}$ and the fact (already argued) that $\mathcal{L}$ has no overall dependence on θ_0.

$$\begin{aligned}
\int_V \mathcal{L}\left(p, pd^{n-1}p.\frac{\partial}{\partial p}\right) &= \sum_{\text{cells}} \mathcal{L}\left(p, e_0e_1\ldots e_{n-1}\theta_0 d\theta_1 \ldots d\theta_{n-1}.\sum_{i=0}^{n-1} E_i \frac{\partial}{\partial \theta_i}\right) \\
&= \sum_{\text{cells}} \sum_{i=0}^{n-1} (-1)^i \frac{\partial}{\partial \theta_i} \mathcal{L}\left(p, e_0e_1\ldots \check{e}_i \ldots e_{n-1}\theta_0 d\theta_1 \ldots d\theta_{n-1}\right) \\
&= -\sum_{\text{cells}} \sum_{i=1}^{n-1} (-1)^{i-1} d\theta_i \frac{\partial}{\partial \theta_i} \mathcal{L}(p, e_0e_1\ldots \check{e}_i \ldots e_{n-1}\theta_0 d\theta_1 \ldots \check{d\theta}_i \ldots d\theta_{n-1}) \\
&= -\sum_{\text{cells}} \sum_{i=1}^{n-1} (-1)^{i-1} (\mathcal{L}(p, e_0e_1\ldots \check{e}_i \ldots e_{n-1}\theta_0 d\theta_1 \ldots \check{d\theta}_i \ldots d\theta_{n-1})|_{\theta_i + d\theta_i/2} \\
&\qquad - \mathcal{L}(p, e_0e_1\ldots \check{e}_i \ldots e_{n-1}\theta_0 d\theta_1 \ldots \check{d\theta}_i \ldots d\theta_{n-1})|_{\theta_i - d\theta_i/2})
\end{aligned}$$

In the last line, we have a sum over the $2(n-1)$ faces of each cell of the volume. Furthermore, on each face, the second slot of $\mathcal{L}$ contains the correct un-normalised boundary element $pd^{n-2}p$. The sums over the interior faces of the cells cancel each other out, so, in the limit, we obtain Stokes's theorem[6] in the form of (4.19).

---

[5]The notation is that of equations (4.2,4.8).

[6]In the light of equation (4.11) and the fact that $\mathcal{L}(p, pd^{n-2}p)$ is assumed, on physical grounds, to have no overall dependence on the weight θ_0, then an alternative form of Stokes's theorem is,

$$\int_V \mathcal{L}\left(p, d^{n-1}p.\frac{\partial}{\partial p}\right) = \int_S \mathcal{L}(p, pd^{n-2}p)\ . \tag{4.20}$$

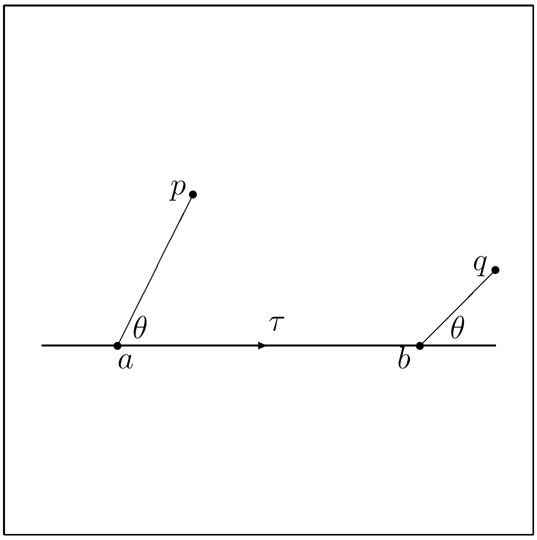

Figure 4.1: Parallel transport along the line ab

4.3.1 Example: Parallel transport around a loop

As an example, we use Stokes's theorem equation (4.19) to obtain a formula for parallel transport around a small loop and then show how this result can be applied to obtain the Gauss-Bonnet theorem for the elliptic plane. In the course of these derivations we will put in place some results that will be needed later in the theory of a general hypersurface embedded in Euclidean space.

In figure 4.1 we want to transport the line ap parallel to itself by a parameter distance τ along the line ab. The result is that line ap is parallel-transported[7] to bq. In Whitehead's algebra, the parallel-transporter is simply a congruence f with invariant line ab and parameter τ chosen as the distance from a to b. This means that a is transported to b by $b = fa$ and a general point p is transported to $q = fp$. We can show that this makes sense by calculating the angle $\angle bap$ and showing that it is invariant under parallel-transport. The angle between the lines is parameter θ which is given by equation (3.38). In the current instance this is,

$$\cosh^2(\theta) = \frac{(abIap)^2}{(abIab)(apIap)} \ .$$

[7]In figure 4.1 it does not look as though the lines ap and bq are parallel because in projective geometry any pair of lines in a plane intersect at some point. However, if the intersection point is *at infinity*, then the lines can be said to be parallel.

If we now parallel-transport everything $p \to fp$, the angle is invariant,

$$\frac{(fafbIfafp)^2}{(fafbIfafb)(fafpIfafp)} = \frac{(fabIfap)^2}{(fabIfab)(fapIfap)}$$
$$= \frac{(fab.fIap)^2}{(fab.fIab)(fap.fIap)} = \frac{(abIap)^2}{(abIab)(apIap)} = \cosh^2(\theta)$$

which shows our parallel-transporter works correctly.

We actually need to be able to parallel-transport along a curve. We can do this by assuming that the path is made up of many short straight segments, which, in the limit, become a curve. The congruence which is the transporter for the segment $a \to a + da$ is, from equation (3.26),

$$f = \exp\left(\frac{d\tau a(a+da).I}{\sqrt{-a(a+da)Ia(a+da)}}\right) = \exp\left(\frac{d\tau ada.I}{\sqrt{-adaIada}}\right) .$$

Now the distance $d\tau$ is given by equation (3.35) so that,

$$d\tau = \frac{\sqrt{-ada.Iada}}{aIa}$$

and by substituting this into the formula for the infinitesimal transporter we get,

$$f = \exp\left(\frac{ada.I}{aIa}\right) . \tag{4.21}$$

The parallel-transporter along a finite curve $a(\tau)$ is,

$$f = \prod_0^\tau \exp\left(\frac{d\tau a\frac{da}{d\tau}.I}{aIa}\right) . \tag{4.22}$$

It is not possible to write the exponential as,

$$\exp\left(\int_0^\tau \frac{ada.I}{aIa}\right)$$

because the Lie-algebra elements $ada.I$ do not commute at different points on the curve[8].

[8]For reference, we give the formula for the Lie-bracket/commutator,

$$[ab.I, cd.I] = ab.I(cd.I) - cd.I(ab.I) = (abd.Ic - abc.Id).I \tag{4.23}$$

where a, b, c, d are points. The formula is a straightforward calculation using the rules of section 1.14 after letting the commutator act on a point.

The multiplicative integral of equation (4.22) will be treated along the lines of section 4.2 of [13]. We start by writing it in general form as,

$$f_\tau = \prod_0^\tau e^{d\tau\mathcal{A}(\tau)} = \prod_0^\tau [1 + d\tau\mathcal{A}(\tau)] \tag{4.24}$$

where $\mathcal{A}$ stands for the Lie-algebra element. Now,

$$f_{\tau+d\tau} = \prod_0^{\tau+d\tau} e^{d\tau\mathcal{A}(\tau)} = e^{d\tau\mathcal{A}(\tau)} f_\tau$$

so that,

$$\frac{df_\tau}{d\tau} = \frac{f_{\tau+d\tau} - f_\tau}{d\tau} = \frac{(1 + d\tau\mathcal{A}(\tau))f_\tau - f_\tau}{d\tau} = \mathcal{A}(\tau)f_\tau \tag{4.25}$$

and we note that this equation is similiar to (3.14), and could have been obtained by just thinking about what happens when the Lie-algebra element depends on the parameter τ.

Now consider,

$$\int_0^\tau d\tau' \frac{df_{\tau'}}{d\tau'} = f_\tau - 1 = \int_0^\tau d\tau' \mathcal{A}(\tau') f_{\tau'}$$

so that,

$$f_\tau = 1 + \int_0^\tau d\tau' \mathcal{A}(\tau') f_{\tau'} \ . \tag{4.26}$$

Iterating,

$$\begin{aligned} f_\tau &= 1 + \int_0^\tau d\tau' \mathcal{A}(\tau') \left(1 + \int_0^{\tau'} d\tau'' \mathcal{A}(\tau'') f_{\tau''}\right) \\ &= 1 + \int_0^\tau d\tau' \mathcal{A}(\tau') + \int_0^\tau d\tau' \mathcal{A}(\tau') \int_0^{\tau'} d\tau'' \mathcal{A}(\tau'') f_{\tau''} \\ &= 1 + \int_0^\tau d\tau' \mathcal{A}(\tau') + \int_0^\tau d\tau' \mathcal{A}(\tau') \int_0^{\tau'} d\tau'' \mathcal{A}(\tau'') + \ldots \end{aligned} \tag{4.27}$$

gives the formal expansion of a multiplicative integral.

Let's evaluate the first term in the series for parallel transport around a closed loop. The integral is,

$$\int_0^\tau d\tau' \mathcal{A}(\tau') = \oint \frac{a da.I}{aIa} = -\int_S d\theta_1 d\theta_2 \left(\frac{a\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}\cdot\frac{\partial}{\partial a}}{aIa}\right).I$$

and Stokes's theorem (4.19) has been used to convert the line integral around the loop to an integral over the 2-d surface of which the loop is the boundary.

Expanding the part which will turn out to be the invariant line of the overall congruence for the loop gives,

$$\frac{a\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}.\frac{\partial}{\partial a}}{aIa} = \frac{\left(a.\frac{\partial}{\partial a}\right)\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2} - \frac{\partial}{\partial\theta_1}a\frac{\partial a}{\partial\theta_2} + \frac{\partial}{\partial\theta_2}a\frac{\partial a}{\partial\theta_1}}{aIa} = \frac{-\frac{\partial}{\partial\theta_1}a\frac{\partial a}{\partial\theta_2} + \frac{\partial}{\partial\theta_2}a\frac{\partial a}{\partial\theta_1}}{aIa}$$

where the term $a.\partial/\partial a$ vanishes[9] because of equation (4.11). Notice that the partial derivatives act on everything in sight, in particular the term aIa in the denominator. This is clear from the derivation of Stokes's theorem in section 4.3 where the linearity of the second slot of $\mathcal{L}(.,.)$ enables the partial derivative operators $\partial/\partial\theta_i$ to be brought out onto the LHS of $\mathcal{L}$ where their action looks conventional. Continuing, using antisymmetry of aa, we can write,

$$\frac{a\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}.\frac{\partial}{\partial a}}{aIa} = -\frac{\partial}{\partial\theta_1}\left(\frac{a}{\sqrt{aIa}}\frac{\partial}{\partial\theta_2}\left(\frac{a}{\sqrt{aIa}}\right)\right) + \frac{\partial}{\partial\theta_2}\left(\frac{a}{\sqrt{aIa}}\frac{\partial}{\partial\theta_1}\left(\frac{a}{\sqrt{aIa}}\right)\right)$$

and then, by equality of mixed partial derivatives,

$$\frac{a\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}.\frac{\partial}{\partial a}}{aIa} = -2\frac{\partial}{\partial\theta_1}\left(\frac{a}{\sqrt{aIa}}\right)\frac{\partial}{\partial\theta_2}\left(\frac{a}{\sqrt{aIa}}\right) .$$

Finally, we can make the expression look like an infinitesimal surface element of the form of equation (4.3),

$$\frac{a\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}.\frac{\partial}{\partial a}}{aIa} = -2\frac{a}{\sqrt{aIa}}\frac{\partial}{\partial\theta_1}\left(\frac{a}{\sqrt{aIa}}\right)\frac{\partial}{\partial\theta_2}\left(\frac{a}{\sqrt{aIa}}\right).\frac{Ia}{\sqrt{aIa}} = \frac{-2dS.Ia}{d\theta_1 d\theta_2\sqrt{aIa}}$$

by using the fact that,

$$2\frac{\partial}{\partial\theta_i}\left(\frac{a}{\sqrt{aIa}}\right).\frac{Ia}{\sqrt{aIa}} = \frac{\partial}{\partial\theta_i}\left(\frac{aIa}{\sqrt{aIa}\sqrt{aIa}}\right) = 0 . \qquad (4.28)$$

After these manipulations, the first term of the series for the multiplicative integral for parallel transport around a loop is,

$$\int_0^\tau d\tau'\mathcal{A}(\tau') = \oint\frac{ada.I}{aIa} = 2\int_S\frac{(dS.Ia).I}{\sqrt{aIa}} . \qquad (4.29)$$

We also have to evaluate the quadratic term in the series (4.27) because it turns out to contain a term proportional to the area of the loop and so,

[9] The same result could have been obtained immediately by using the alternative form of Stokes's theorem given by equation (4.20).

even for a small loop, it is of the same order as the linear term (4.29). Firstly, we split the term into symmetric and antisymmetric parts.

$$\begin{aligned}\int_0^\tau d\tau'\mathcal{A}(\tau')\int_0^{\tau'} d\tau''\mathcal{A}(\tau'') = \\ &\frac{1}{2}\int_0^\tau d\tau'\int_0^{\tau'} d\tau''(\mathcal{A}(\tau')\mathcal{A}(\tau'')+\mathcal{A}(\tau'')\mathcal{A}(\tau')) \\ + \ &\frac{1}{2}\int_0^\tau d\tau'\int_0^{\tau'} d\tau''(\mathcal{A}(\tau')\mathcal{A}(\tau'')-\mathcal{A}(\tau'')\mathcal{A}(\tau'))\end{aligned} \tag{4.30}$$

The symmetric integral can be handled by swopping dummy integration variables in its second term.

$$\int_0^\tau d\tau'\int_0^{\tau'} d\tau''(\mathcal{A}'\mathcal{A}''+\mathcal{A}''\mathcal{A}') = \int_0^\tau d\tau'\int_0^{\tau'} d\tau''\mathcal{A}'\mathcal{A}'' + \int_0^\tau d\tau''\int_0^{\tau''} d\tau'\mathcal{A}'\mathcal{A}''$$

The first integral is over the lower right triangle of the square in the (τ', τ'') plane whilst the second integral is over the upper left triangle of the square. Hence the complete symmetric integral is over the complete square and decomposes into the square of the first integral (4.29).

$$\int_0^\tau d\tau'\int_0^{\tau'} d\tau''(\mathcal{A}'\mathcal{A}''+\mathcal{A}''\mathcal{A}') = \int_0^\tau d\tau'\int_0^{\tau} d\tau''\mathcal{A}'\mathcal{A}'' = \left(\int_0^\tau d\tau'\mathcal{A}'\right)^2$$

Consequently, for a small loop, the symmetric integral only contributes a term proprtional to the square of the loop area, and can be neglected in comparison with the term (4.29), which is proportional to the loop area.

Now we evaluate the antisymmetric integral from equation (4.30). The antisymmetric combination of Lie algebra elements can immediately be put into the form of a commutator and evaluated using equation (4.23).

$$\begin{aligned}\frac{1}{2}\int_0^\tau d\tau'\int_0^{\tau'} d\tau''(\mathcal{A}'\mathcal{A}''-\mathcal{A}''\mathcal{A}') &= \frac{1}{2}\int_0^\tau d\tau'\int_0^{\tau'} d\tau''[\mathcal{A}',\mathcal{A}''] \\ &= \frac{1}{2}\int_0^\tau\int_0^{\tau'}(a'da'da''.Ia''-a'da'a''.Ida'').I\end{aligned} \tag{4.31}$$

In writing this equation, we have shortened the formulae by writing a for $a/\sqrt{aIa}$ or, alternatively, just assuming the points are normalised so $aIa = 1$. The full expressions can easily be restored at the end of the calculation. We can evaluate (4.31) for a small loop by neglecting the variation of the points as they traverse the paths. Thus, we write $a' \approx a$ and $a'' \approx a$ where a stands for the point at the start of the loop.

$$\begin{aligned}\frac{1}{2}\int_0^\tau\int_0^{\tau'}(a'da'da''.Ia''-a'da'a''.Ida'').I &\approx \frac{1}{2}\int_0^\tau\int_0^{\tau'}(ada'da''.Ia-ada'a.Ida'').I \\ = \frac{1}{2}\int_0^\tau\int_0^{\tau'}(ada'da''.Ia).I &= \frac{1}{2}\int_0^\tau(ada'(a'-a).Ia).I = -\frac{1}{2}\oint(aa'da'.Ia).I\end{aligned}$$

The integrand can be written,

$$\begin{aligned} aa'da'.Ia &= a'da' - (a'Ia)ada' + (da'Ia)aa' = a'da' - a((a'Ia)da' - (da'Ia)a') \\ &= a'da' - a(a'da'.Ia) \ . \end{aligned}$$

Now we already know how to integrate the $a'da'$ term from equation (4.29). Therefore, the antisymmetric integral for a small loop is,

$$\begin{aligned} -\frac{1}{2}\oint(aa'da'.Ia).I &= -\frac{1}{2}\oint a'da'.I + \frac{1}{2}\oint(a(a'da'.Ia)).I \\ &= -\int_S (dS.Ia).I + \int_S (a((dS.Ia).Ia)).I \\ &= -\int_S (dS.Ia).I + \int_S (a(dS.Ia.Ia)).I \\ &= -\int_S (dS.Ia).I \ . \end{aligned}$$

Upon restoring the explicit normalization, we have the following result for the antisymmetric integral for a small loop.

$$\frac{1}{2}\int_0^\tau d\tau' \int_0^{\tau'} d\tau''[\mathcal{A}', \mathcal{A}''] = -\int_S \frac{(dS.Ia).I}{\sqrt{aIa}} \tag{4.32}$$

Adding equations (4.29) and (4.32) gives the parallel transporter (4.22) for a small loop to the first order in the area of the loop.

$$f = \prod \exp\left(\frac{ada.I}{aIa}\right) = 1 + \int_S \frac{(dS.Ia).I}{\sqrt{aIa}} + \dots \tag{4.33}$$

So, the effect of parallel transport around a small loop which starts from point a is the same[10] as a rotation with invariant line,

$$dS.Ia = d\theta_1 d\theta_2 a \frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}.Ia = d\theta_1 d\theta_2 \frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2} \ . \tag{4.34}$$

The surface element dS is an element of point-grade 3. The magnitude of the area is found by extracting a number from dS in the most obvious way,

$$\begin{aligned} \sqrt{dS.IdS} &= d\theta_1 d\theta_2 \sqrt{a\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}.Ia\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}} \\ &= d\theta_1 d\theta_2 \sqrt{\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}.I\frac{\partial a}{\partial\theta_1}\frac{\partial a}{\partial\theta_2}} = \sqrt{(dS.Ia).I(dS.Ia)} \ . \end{aligned} \tag{4.35}$$

Equations (4.34) and (4.35) show that the numerical value of the area of the small loop is equal to the magnitude of the invariant line. In other words,

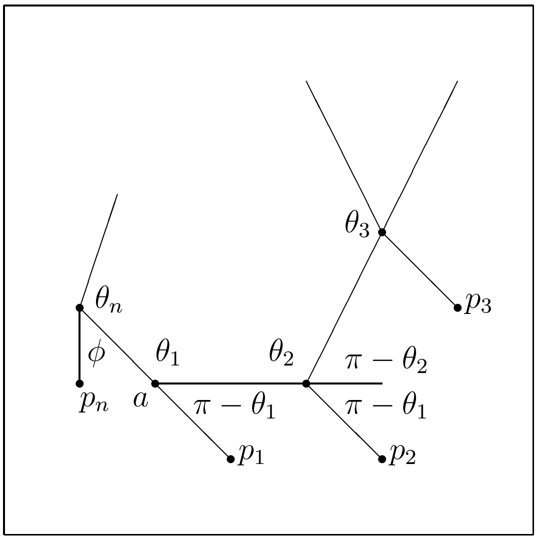

Figure 4.2: Parallel transport around a planar polygon

the rotation angle produced by parallel transport around a small loop is equal to the area of the loop.

Let us now consider parallel transport around a planar polygon as shown in figure 4.2. The interior angles of the polygon are $\theta_1, \ldots, \theta_n$. The start of the path is point a and the point being transported is p_1. Under parallel transport along the sides of the polygon, the point goes through the sequence $p_1 \to p_n$. Figure 4.1 has already shown that the angle between the transported line and the polygon side remains constant under parallel transport. Of course, this angle changes discontinuously at each vertex of the polygon when the direction of parallel transport changes. For example, at the start of the path in figure 4.2, the angle is $\pi - \theta_1$. Along the second side of the polygon it is $2\pi - \theta_1 - \theta_2$. Along the final side of the polygon the angle is,

$$\phi = n\pi - \sum_{i=1}^{n} \theta_i = \sum_{i=1}^{n} (\pi - \theta_i) \ .$$

The angle ϕ is how much extra rotation needs to be applied to the line in order to make a full rotation of 2π after passage around the polygon. Therefore, if we denote by Φ the actual rotation of the line under parallel transport around the planar polygon, then $\Phi + \phi = 2\pi$. Hence,

$$\Phi = 2\pi - \sum_{i=1}^{n} (\pi - \theta_i) \tag{4.36}$$

[10]In order to simplify the formulae, we revert to the normalisation $aIa = 1$.

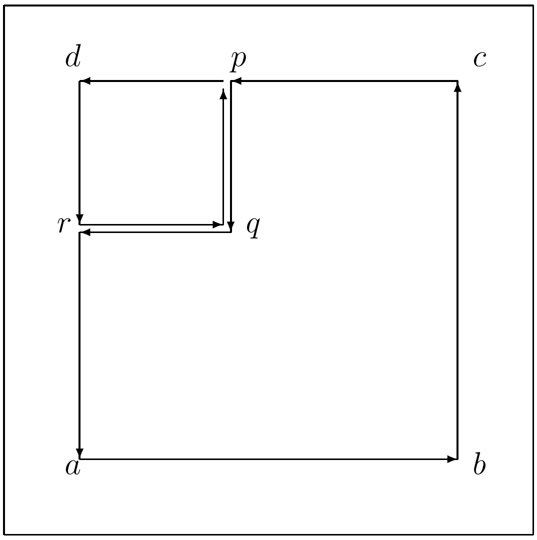

Figure 4.3: Building a loop from small loops

is the formula for the rotation angle for parallel transport around a polygon in the elliptic plane.

Now let's find the relation between the two ways of thinking about parallel transport around a loop, which are summarised by equations (4.36) and (4.33). However, since (4.36) refers to a finite loop, whilst (4.33) refers to an infinitesimal loop, we need to make a finite loop out of many small loops. Figure 4.3 shows how this is done. The parallel-transporter or congruence will be denoted by f with the vertices of the path as a subscript. So, the transporter for the small loop in the top right hand corner of figure 4.3 is f_{pdrqp}. The transporter for the L-shaped loop is $f_{abcpqra}$. Let's start with the L-shaped loop and add the small loop to make the transporter for the square loop f_{abcda}.

$$f_{abcda} = f_{pqra} f_{pdrqp} f_{abcp}$$

This is more informatively written as $f_{\text{square}} = f_{\text{back}} f_{\text{loop}} f_{\text{out}}$. Now, this can be written,

$$f_{\text{square}} = f_{\text{back}} f_{\text{loop}} f_{\text{out}} = (f_{\text{back}} f_{\text{loop}} f_{\text{back}}^{-1})(f_{\text{back}} f_{\text{out}}) \ . \qquad (4.37)$$

Now, $f_{\text{back}} f_{\text{out}}$ is the original L-shaped loop. So, we have succeeded in making the square loop out of the L-shaped loop by adding the small loop and arranging things so that the new transporter is made by just acting on the LHS of the original transporter with a transformed version of the small loop. Equation (4.37) is general and is not restricted to planar loops. It shows how we could make a finite loop by continually applying infinitesimal loops on the LHS as in equation (4.37). Each infinitesimal transporter, the f_{loop} part of equation (4.37), would be given by equation (4.33).

In order to make further progress, we need to work on the transformed transporter $f_{\text{back}} f_{\text{loop}} f_{\text{back}}^{-1}$. From equation (3.3), this is just the formula for the transformation of a congruence. Let's work on this in the general case of the transformation $g' = fgf^{-1}$ of equation (3.3). From equation (3.15), we write $g = \exp(\mathcal{A})$ where $\mathcal{A}$ is an element of the Lie algebra.

$$g' = fe^{\mathcal{A}}f^{-1} = f\left(1 + \sum_{k=1}^{\infty} \frac{(\mathcal{A})^k}{k!}\right) f^{-1} = 1 + \sum_{k=1}^{\infty} \frac{(f\mathcal{A}f^{-1})^k}{k!}$$

For g a simple congruence, equation (3.17) gives $\mathcal{A} = l.I$ where $l = ILM$ is the invariant line of the congruence. Now,

$$f\mathcal{A}f^{-1} = f(l.I(f^{-1}(.))) = fl.fIf^{-1}(.) = fl.ff^{-1}I(.) = fl.I$$

so that in order to transform a simple congruence, it is only necessary to transform its invariant line as in,

$$g' = fgf^{-1} = fe^{l.I}f^{-1} = e^{fl.I} \ . \tag{4.38}$$

The invariant line of f_{loop} is given by equation (4.34) as $pd^2p.Ip$ since the loop is based at p. The action of f_{back} is to move the loop so that it is based at a.

$$f_{\text{back}}\left(d\theta_1 d\theta_2 p \frac{\partial p}{\partial \theta_1}\frac{\partial p}{\partial \theta_2}.Ip\right) = d\theta_1 d\theta_2 a \frac{\partial a}{\partial \theta_1}\frac{\partial a}{\partial \theta_2}.Ia = d\theta_1 d\theta_2 \frac{\partial a}{\partial \theta_1}\frac{\partial a}{\partial \theta_2} \ .$$

In this general case, the effect of the combination of all these small loops based at a is still intractable because the invariant lines will not commute. However, for a finite planar loop, the invariant lines of all the infinitesimal loops based at a will be same line, only their weights will differ because each weight is the area of the small loop. To see this, consider the plane $a_1a_2a_3$ with the start of the planar loop at a_1. The area elements for the infinitesimal loops based at a_1 are all of the form $|dS|a_1a_2a_3.Ia_1 = |dS|a_2a_3$ where $|dS|$ is the magnitude of the area. Each infinitesimal loop has invariant line a_2a_3 so they all commute. Therefore, a planar loop can be built out of the joint effect of commuting infinitesimal loops based at a as,

$$\prod \exp(|dS|l.I) = \exp\left(\int |dS|l.I\right) \tag{4.39}$$

where l is the constant normalized invariant line for the loops based at a and the integral is over the area of the finite planar loop. Furthermore, this result shows that the rotation angle for parallel transport around a finite planar loop is $\Phi = \int |dS|$. Combining this with equation (4.36), gives the Gauss-Bonnet theorem for polygon figures in the elliptic plane.

$$\int_S |dS| + \sum_{i=1}^{n}(\pi - \theta_i) = 2\pi \tag{4.40}$$

For a triangle in the elliptic plane, we find that the area is given by the sum of the interior angles minus π.

4.4 The generalised derivative

4.4.1 Definition

Let X be an element of point grade r. If we have reference points $a_1, \ldots, a_n$ with dual hyperplanes $A_1, \ldots, A_n$ so that $a_i A_j = \delta_{ij}$ then X may be expanded as (1.5). Note that the $X.A_{j_1} \ldots A_{j_r}$ are numerical coefficients. By analogy with the definition (4.6) of the derivative with respect to a point, we can define a generalised derivative with respect to an element of point grade r as,

$$\frac{\partial}{\partial X} = \sum_{1 \le j_1 < \ldots < j_r \le n} A_{j_1} \ldots A_{j_r} \frac{\partial}{\partial (X.A_{j_1} \ldots A_{j_r})} \ . \tag{4.41}$$

Note that the derivatives under the summation sign are just ordinary partial derivatives with respect to the scalar variables $X.A_{j_1} \ldots A_{j_r}$. This derivative is meaningful because it is independent of the reference hyperplanes. In order to see this we generalise the argument at the end of section 4.2.1. Start with a set of alternative hyperplanes B_i and their dual points b_i so that $b_i B_j = \delta_{ij}$.

$$\begin{aligned} & \sum_{j_1,\ldots,j_r} B_{j_1} \ldots B_{j_r} \frac{\partial}{\partial (X.B_{j_1} \ldots B_{j_r})} \\ = & \sum_{j_1,\ldots,j_r} \sum_{k_1,\ldots,k_r} B_{j_1} \ldots B_{j_r} \frac{\partial (X.A_{k_1} \ldots A_{k_r})}{\partial (X.B_{j_1} \ldots B_{j_r})} \frac{\partial}{\partial (X.A_{k_1} \ldots A_{k_r})} \end{aligned} \tag{4.42}$$

Expanding X in terms of the $B_{j_1} \ldots B_{j_r}$, we find,

$$(X.A_{k_1} \ldots A_{k_r}) = \sum_{1 \le j_1 < \ldots < j_r \le n} (X.B_{j_1} \ldots B_{j_r})(b_{j_1} \ldots b_{j_r}.A_{k_1} \ldots A_{k_r})$$

so that,

$$\frac{\partial (X.A_{k_1} \ldots A_{k_r})}{\partial (X.B_{j_1} \ldots B_{j_r})} = (b_{j_1} \ldots b_{j_r}.A_{k_1} \ldots A_{k_r}) \ .$$

Substituting in (4.42) we get,

$$\begin{aligned}
&\sum_{j_1,\ldots,j_r} B_{j_1}\ldots B_{j_r}\frac{\partial}{\partial(X.B_{j_1}\ldots B_{j_r})}\\
=&\sum_{j_1,\ldots,j_r}\sum_{k_1,\ldots,k_r} B_{j_1}\ldots B_{j_r}(b_{j_1}\ldots b_{j_r}.A_{k_1}\ldots A_{k_r})\frac{\partial}{\partial(X.A_{k_1}\ldots A_{k_r})}\\
=&\sum_{k_1,\ldots,k_r} A_{k_1}\ldots A_{k_r}\frac{\partial}{\partial(X.A_{k_1}\ldots A_{k_r})}=\frac{\partial}{\partial X}
\end{aligned}$$

which shows that the derivative defined by (4.41) does not depend on the particular set of hyperplanes used in the definition.

The properties of the derivative with respect to a point also apply to the generalised derivative. For example, a straightforward generalisation of the proof of equation (4.7) shows that the generalised derivative transforms as,

$$\frac{\partial}{\partial fX}=f\frac{\partial}{\partial X} \tag{4.43}$$

under a collineation f. Furthermore, the operation of a small change is,

$$\delta=\delta X.\frac{\partial}{\partial X}\ . \tag{4.44}$$

4.4.2 Duality

Let X be a hyperplane. From (4.41), the derivative with respect to a hyperplane is obtained by setting $r=n-1$. The summation is over the labels $1\leq j_1<\ldots<j_{n-1}\leq n$. Hence, the labels in the sum will go from $2,3,\ldots,n$ up to $1,2,\ldots,n-1$. In other words, we can get the labels correctly by just noticing which single number is missing from the sequence $1,2,\ldots,n$. From (4.41), but writing the labels in the new way, the derivative with respect to a hyperplane is,

$$\frac{\partial}{\partial X}=\sum_{j=1}^{n}A_1\ldots\check{A}_j\ldots A_n\frac{\partial}{\partial(X.A_1\ldots\check{A}_j\ldots A_n)}$$

where the "ˇ" denotes a missing hyperplane. According to equation (1.26), the hyperplane products are proportional to a_j, so that the hyperplane derivative becomes,

$$\frac{\partial}{\partial X}=\sum_{j=1}^{n}a_j\frac{\partial}{\partial(Xa_j)}=(-1)^{n-1}\sum_{j=1}^{n}a_j\frac{\partial}{\partial(a_jX)}\ . \tag{4.45}$$

This is the same as the definition (4.6) of the derivative with respect to a point, except that hyperplanes have been replaced by points in accordance with the principle of duality.

4.4.3 Invariants of a collineation

Let X be of point-grade r and let g be a collineation. Let's evaluate the expression $X.g(\partial/\partial X)$ where the derivative acts on the variable X to the left.

$$
\begin{aligned}
X.g\left(\frac{\partial}{\partial X}\right) &= X.g\left(\sum_{j_1,\ldots,j_r} A_{j_1}\ldots A_{j_r}\frac{\partial}{\partial(X.A_{j_1}\ldots A_{j_r})}\right) \\
&= \sum_{j_1,\ldots,j_r} \frac{\partial X}{\partial(X.A_{j_1}\ldots A_{j_r})}.g(A_{j_1}\ldots A_{j_r})
\end{aligned}
$$

The derivative can be evaluated by differentiating (1.5). The result is,

$$
X.g\left(\frac{\partial}{\partial X}\right) = \sum_{1\le j_1<\ldots<j_r\le n} (a_{j_1}\ldots a_{j_r}).g(A_{j_1}\ldots A_{j_r})\ . \tag{4.46}
$$

$X.g(\partial/\partial X)$ is a number which, as we know, transforms as a scalar in Whitehead's algebra. The rhs of (4.46) shows how to evaluate the expression in terms of a set of reference points and dual hyperplanes. However, the lhs shows that the value of the expression does not depend on the particular set of reference points used.

The cases $r=1$ and $r=n$ give familiar invariants. For $r=1$ the variable X is a point and from (4.46),

$$
X.g\left(\frac{\partial}{\partial X}\right) = \sum_{j=1}^{n} a_j.g(A_j) = \mathrm{tr}(g)\ .
$$

This is the trace of g because the sum is over the diagonal elements of the matrix[11] of g. For $r=n$, the variable X is a pseudonumber. Equation (4.46) gives,

$$
\begin{aligned}
X.g\left(\frac{\partial}{\partial X}\right) &= a_1\ldots a_n.g(A_1\ldots A_n) = g^{-1}(a_1\ldots a_n).A_1\ldots A_n \\
&= \det(g)a_1\ldots a_n.A_1\ldots A_n = \det(g)
\end{aligned}
$$

where we used (1.9)[12]. Although the trace $\mathrm{tr}(g)$ and the determinant $\det(g)$ are the only invariants constructed from a collineation in everyday use, there

[11] See section 3.11.1.

[12] Note that the inverse of the determinant appeared in equation (1.9) because we have effectively defined the determinant by,

$$
g(A_1\ldots A_n) = \det(g)A_1\ldots A_n\ . \tag{4.47}
$$

are n invariants[13] of g that can be obtained from (4.46) by setting $r = 1, 2, \ldots, n$.

Let us clarify the invariant character of these numbers that can be obtained from a collineation.

$$\begin{array}{rcl} & \text{Bob} \xrightarrow{f} & \text{Alice} \\ & fX & X \\ & fgf^{-1} & g \\ & f\dfrac{\partial}{\partial X} & \dfrac{\partial}{\partial X} \\ fX.fgf^{-1}\left(f\dfrac{\partial}{\partial X}\right) & = fX.fg\left(\dfrac{\partial}{\partial X}\right) & X.g\left(\dfrac{\partial}{\partial X}\right) \\ = f\left[X.g\left(\dfrac{\partial}{\partial X}\right)\right] & = X.g\left(\dfrac{\partial}{\partial X}\right) & \end{array}$$

In the diagram above, Alice evaluates $X.g(\partial/\partial X)$. The diagram shows how the constituents X, g and $\partial/\partial X$ transform so that the result is that Bob sees the same number as Alice. There is nothing unexpected here because $X.g(\partial/\partial X)$ is a number and a number transforms as a scalar. This is not what we mean by saying that $X.g(\partial/\partial X)$ is an invariant. Instead, $X.g(\partial/\partial X)$ is really a functional - a number-valued function of a function g. In the case of $\mathrm{tr}(g)$ and $\det(g)$ this is clear from the notation. The expression $X.g(\partial/\partial X)$ is just a prescription which tells the observers how to get a number out of a collineation. Let's see what happens when the observers apply the fixed prescription to their own versions of the collineation.

$$\begin{array}{rcl} & \text{Bob} \xrightarrow{f} & \text{Alice} \\ & fgf^{-1} & g \\ & X.fgf^{-1}\left(\dfrac{\partial}{\partial X}\right) & X.g\left(\dfrac{\partial}{\partial X}\right) \\ & = f^{-1}X.g\left(\dfrac{\partial}{\partial f^{-1}X}\right) & \\ = Y.g\left(\dfrac{\partial}{\partial Y}\right) & = X.g\left(\dfrac{\partial}{\partial X}\right) & \end{array}$$

In the above diagram we used equation(4.43) and then defined a new variable $Y = f^{-1}X$. However, Y is just a dummy variable in $Y.g(\partial/\partial Y)$ and so Y can

[13] These infrequently-used invariants are mentioned in a corollary to theorem 11 of section 10.4 on page 334 of [28].

be replaced by X. The result is that $X.fgf^{-1}(\partial/\partial X) = X.g(\partial/\partial X)$ so all observers agree on the number assigned to a collineation or operator. This is the sense in which $X.g(\partial/\partial X)$ is an invariant.

Finally, suppose we use a set of hyperplanes in which g is diagonal so that $gA_i = \lambda_i A_i$. Then, from (4.46),

$$\begin{aligned} X.g\left(\frac{\partial}{\partial X}\right) &= \sum_{1\leq j_1<\ldots<j_r\leq n} \lambda_{j_1}\ldots\lambda_{j_r} a_{j_1}\ldots a_{j_r}.A_{j_1}\ldots A_{j_r} \\ &= \sum_{1\leq j_1<\ldots<j_r\leq n} \lambda_{j_1}\ldots\lambda_{j_r}\ . \end{aligned} \tag{4.48}$$

This expression reduces to the familiar formulae for the trace $\mathrm{tr}(g) = \sum_i \lambda_i$ and determinant $\det(g) = \prod_i \lambda_i$ in terms of the eigenvalues when $r = 1$ and $r = n$ respectively.

Chapter 5

Space-Time

5.1 Introduction

This chapter considers space-time and consequently it may be omitted by those readers who are primarily interested in 3-d computational geometry. We make our argument qualitatively here, filling in the mathematical details in the bulk of the chapter.

> The two chief ways ways of approaching non-Euclidean geometry are that of Gauss, Lobatschewsky, Bolyai, and Riemann, who began with Euclidean geometry and modified the postulates, and that of Cayley and Klein, who began with projective geometry and singled out a polarity. *Coxeter [9]*

It is clear from section and the above quotation that Whitehead's algebra is firmly in the camp of Cayley and Klein . As has already been explained in section , they saw the different types of geometry descending from projective geometry by picking invariant structures.

Projective The parent geometry.

- **Elliptic** Pick an elliptic polarity.
- **Affine** Pick a hyperplane.
  - **Euclidean** Make the hyperplane an elliptic space.
  - **Minkowskian** Make the hyperplane a hyperbolic space.
- **Hyperbolic** Pick a hyperbolic polarity.

In this hierarchy, affine geometry incorporates the notion of parallel lines by the device of picking an invariant hyperplane which is called the hyperplane at infinity. Lines which meet at a point in this hyperplane are said to be parallel because they meet at infinity. Since Minkowski space-time also has the notion of parallelism, it has to be an affine geometry. Pure affine geometry has no light cone structure. In figure 5.1 we add the light cone structure and get Minkowski space-time by making the hyperplane at infinity into a hyperbolic space by picking an invariant hyperbolic polarity within the hyperplane. If instead, the hyperplane at infinity is an elliptic space, then we get Euclidean space.

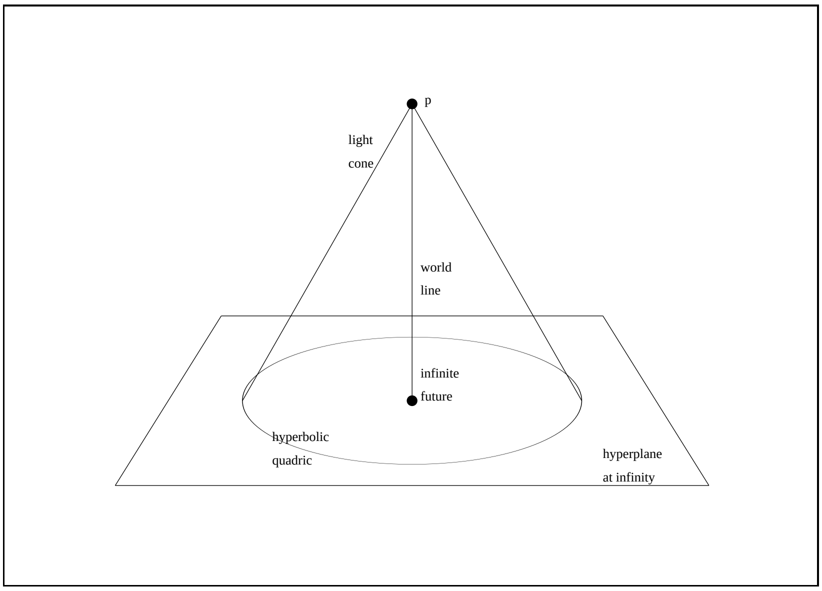

Figure 5.1: Minkowski space-time

A general congruence of an affine space can be split into a product of a congruence that takes the hyperplane at infinity into itself, and a dilatation. In 4-d Minkowski space-time, the group of congruences of the hyperbolic space at infinity is the 6 parameter Lorentz group consisting of 3 spacelike rotations and 3 timelike rotations or boosts. To this must be added the group of dilatations. A dilatation is an expansion or contraction of the space which peters out to the identity transformation in the hyperplane at infinity [14]. A translation is a limiting case of a dilatation, in which the point at the centre of the expansion is moved onto the hyperplane at infinity. The problem is that the general dilatations are never experienced in our world. We habitually experience spacelike rotations and spacelike and timelike translations, and in accelerator labs we see the effects of timelike rotations, but we never experience a general dilatation in which space expands or contracts.

The same problem occurs with Euclidean space. A general congruence of 3-d Euclidean space consists of a product of a congruence which takes the elliptic plane at infinity into itself and a dilatation. The congruence of the elliptic plane at infinity is a rotation. The dilatation is never experienced, except in its limiting form of a translation.

This problem seems to be either ignored or not taken seriously by those who have adopted the Cayley-Klein view of geometries. Sir Robert Ball ignores the problem in his treatise on screw theory [15]. In his introduction, he derives the screw as a general congruence of 3-d Euclidean space by considering the possibilities for its eigenpoints. There are four eigenpoints. Three of them lie on the elliptic plane at infinity. Of these three, one is real and two are a complex conjugate pair. The real one is the centre of a rotation in the plane at infinity. The complex conjugate pair form the invariant line of the rotation. Sir Robert Ball argues that the fourth eigenpoint must lie on the plane at infinity and be coincident with the real eigenpoint of the rotation. The fourth eigenpoint is in fact the centre of the dilatation. By putting the fourth eigenpoint on the plane at infinity, he forces the dilatation into its limiting form of a translation. The natural thing to do would be to leave the fourth eigenpoint at a finite position. He justifies putting the fourth eigenpoint at infinity because the congruence "does not alter distances", but this puts the cart before the horse.

Another example is the treatment of dilatations by Semple and Kneebone in their book on projective geometry [16]. They define the 2-d Euclidean plane as an affine plane with an elliptic line at infinity. In theorem 17 of chapter IV, they obtain the general congruence as a rotation and a dilatation. They define the distance between two points (X_1, X_2) and (Y_1, Y_2) as,

$$\rho = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2}$$

and remark that the distance transforms as $\rho \to c\rho$ under the action of the general congruence. Here, the number c is the expansion factor of the dilatation. They then make the following statement.

> It (ρ) is not a euclidean invariant, but any ratio of distances is invariant. For this reason the euclidean plane, as we have defined it, is sometimes referred to as the similarity euclidean plane.

This makes it clear that their Euclidian plane admits congruences which are not seen in the real world. However, a little later, they refuse to take the dilatations seriously by adding the following remark.

> The constant c corresponds to the arbitrariness of the unit of length.

In contrast to Semple and Kneebone, we argue that a dilatation cannot be dismissed as a change in the unit of length, and, in fact, is just as real as a rotation, translation or boost. Once this point is accepted, it then follows, from the fact that we do not experience dilatations, that space-time cannot be Minkowskian.

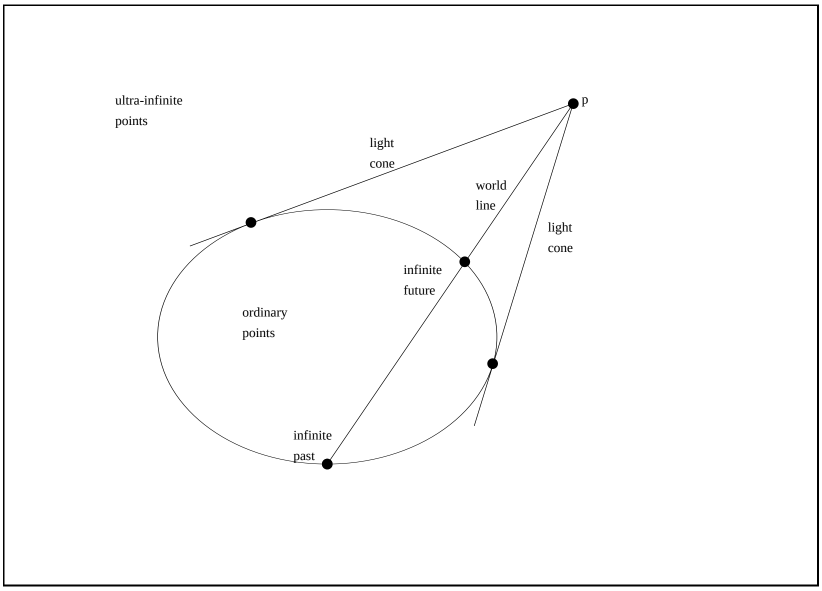

Figure 5.2: The natural light cone structure in hyperbolic geometry

Given that space-time is not Minkowskian, we start with a clean sheet of paper and try to construct the simplest space-time in the Cayley-Klein hierarchy of geometries. The obvious thing to do is to look for an invariant light cone structure in either elliptic or hyperbolic geometry. The idea is to use the absolute quadric locus[1] of the geometry to define the light cone structure. In elliptic geometry, the quadric is imaginary so it is unlikely that it could be used to make a real light cone. However, in hyperbolic geometry, the quadric locus is real and there is a natural light cone structure which is shown in figure 5.2.

In figure 5.2 the real quadric locus of hyperbolic space is drawn schematically as an ellipse. The point p outside the quadric is a space-time event. The light cone at p is the set of straight lines which pass through p and which are tangents to the quadric. The worldline of an observer at p is a line within the light cone of p. The infinite past and the infinite future of the observer are the points at which the observer's worldline intersects the quadric.

[1]See section 3.10.

In hyperbolic geometry, the ordinary points are those inside the quadric and the points outside the quadric are called *ultra-infinite* [9]. In figure 5.2, the physical events in space-time are modelled as the ultra-infinite points of hyperbolic geometry. The natural space-time formed by the set of ultra-infinite points in hyperbolic geometry is called *de Sitter space* in General Relativity [17, 18][2]. However, unlike the situation in General Relativity where de Sitter space is a solution of the gravitational field equations, here, de Sitter space emerges as the natural space-time from the Cayley-Klein point of view before any consideration of the laws governing forces. De Sitter space has a 10 parameter group of congruences. It has 1 timelike translation, 3 spacelike translations, 3 spacelike rotations and 3 timelike rotations or boosts. There are no dilatations and so the de Sitter group is fully in accord with our experience of the real world, which is unlike the situation we encountered in Minkowski space. We therefore propose that the *background for Special Relativity should be de Sitter space*[3] instead of Minkowski space. De Sitter space, being the ultra-infinite points of hyperbolic geometry, is curved. Therefore, we are proposing that *there is curvature without matter.*

The remainder of this chapter fills in the mathematical details of the foregoing argument, and also obtains some of the properties of de Sitter space from the Cayley-Klein perspective.

5.2 Affine space

In projective space, any two hyperplanes L and M always meet in the subspace LM. In contrast, an affine space has the concept of parallel hyperplanes. Two parallel hyperplanes do not meet. Furthermore, this notion is absolute, for if two hyperplanes are parallel according to affine observer Alice, they are also parallel according to affine observer Bob.

In Whitehead's algebra, two hyperplanes will always meet in the subspace LM. The notion of parallel hyperplanes is introduced by the device of picking a hyperplane A_0 and calling it *the hyperplane at infinity.* Then, two hyperplanes are said to be parallel if they meet at infinity, in other words, if LM lies in A_0, so that $LMA_0 = 0$. Parallelism is made an absolute notion by requiring that all affine observers agree on the hyperplane at infinity. So,

[2]In order to avoid confusion, it should be noted that there are two topologically different types of de Sitter space. The de Sitter space in [17] has spacelike slices that are 3-spheres, whilst in [18] the spacelike slices are 3-d elliptic spaces. The de Sitter space which emerges naturally in Whitehead's algebra has elliptic spacelike slices. This version is known as *elliptic de Sitter space* in the modern literature [19].

[3]This point of view has also been proposed by the authors of [20], although their motivation is not the fact that we do not experience dilatations.

if f is an affine congruence,

$$fA_0 = \alpha^{-1}A_0 \tag{5.1}$$

where α is a weight. To see that parallelism is absolute, suppose that Alice says that L and M are parallel so that $LMA_0 = 0$. Bob sees L as fL and M as fM. By the outermorphism property of the collineation[4], $0 = f(LMA_0) = fL.fM.fA_0 = \alpha^{-1}fL.fM.A_0$ so that $fL.fM.A_0 = 0$ which shows that the hyperplanes are also parallel according to Bob's point of view.

5.2.1 The form of the general affine congruence

Equation (5.1) governing a general affine congruence can be satisfied by writing it as a product $f = f^{(2)}f^{(1)}$ of a pair of simpler affine congruences. In order to see this, take an n-dimensional affine space and pick reference points $a_1, \ldots, a_n$ for the hyperplane A_0 at infinity and a reference point a_0 in the finite space. We can then make a set of hyperplanes $A_0, A_1, \ldots, A_n$ dual to the reference points[5] with,

$$A_0 = \frac{a_1 \ldots a_n}{a_0 a_1 \ldots a_n} \tag{5.2}$$

as the hyperplane at infinity. Note that $a_0A_0 = 1$. When restricted to act on the points of A_0, $f^{(1)}$ is a general collineation of the hyperplane into itself. In the finite space, $f^{(1)}a_0 = a_0$. Conversely, $f^{(2)}$ acts like the identity on all the points in A_0 so $f^{(2)}a_i = a_i$ for $i = 1, \ldots, n$. In the finite space, there is no loss of generality in choosing a_0 as the eigenpoint of $f^{(2)}$ with eigenvalue α so that $f^{(2)}a_0 = \alpha a_0$. The $f^{(2)}$ part is called a *dilatation*. This product automatically satisfies equation (5.1).

$$\begin{aligned} fA_0 &= \frac{f^{(2)}f^{(1)}a_1 \ldots a_n}{f^{(2)}f^{(1)}a_0.f^{(2)}f^{(1)}a_1 \ldots a_n} = \frac{f^{(1)}a_1 \ldots a_n}{\alpha a_0.f^{(1)}a_1 \ldots a_n} \\ &= \frac{\det f^{(1)}a_1 \ldots a_n}{\alpha a_0.\det f^{(1)}a_1 \ldots a_n} = \frac{a_1 \ldots a_n}{\alpha a_0 a_1 \ldots a_n} = \frac{A_0}{\alpha}\ . \end{aligned}$$

5.2.2 Dilatations

In order to simplify the notation, we drop the superscript and write a pure dilatation as f. Consider two points p and q in the finite space. The line pq intersects the hyperplane at infinity at the point $pq.A_0$. Since a dilatation acts

[4]See equation (1.6).

[5]See equation (1.18).

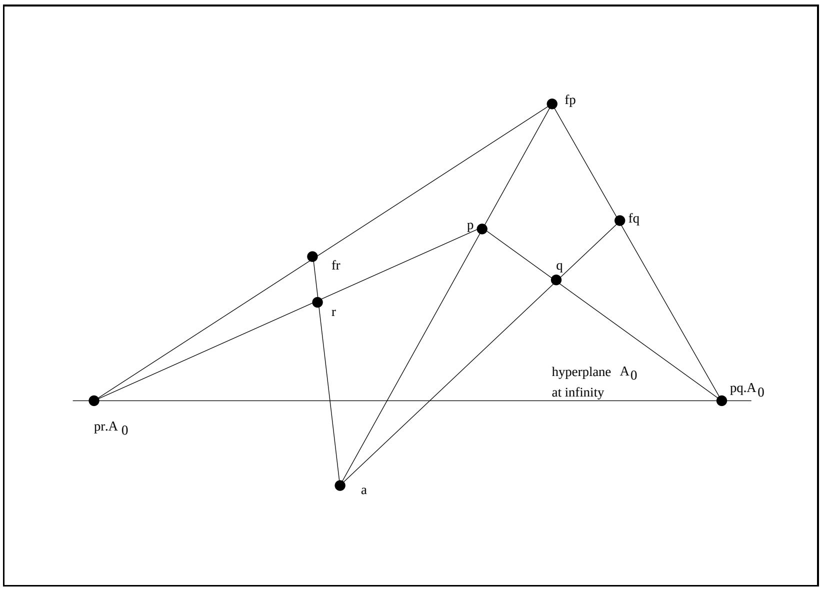

Figure 5.3: A dilatation

like the identity in A_0, $pq.A_0 = f(pq.A_0) = (fp.fq).fA_0 = \alpha^{-1}(fp.fq).A_0$. Therefore, the line pq and the line $fp.fq$ both intersect A_0 at the same point. This is shown in figure 5.3. Notice that the lines pq and $fp.fq$ are parallel because they meet at infinity.

Now consider the point a shown in figure 5.3. The line pa intersects the hyperplane at $pa.A_0$. From the previous result, the line $fp.fa$ also intersects the hyperplane at the same point $pa.A_0$. Therefore, fa must lie on the line pa. Similarly, by considering the line qa, we find that fa also lies on qa. Therefore a and fa are coincident, and so a is an eigenpoint of the dilatation. Figure 5.3 shows that a is the centre of the dilatation because a general point, such as p, moves to fp on the radial line pa. There are no more eigenpoints in the finite space because the dilatation acts as the identity when restricted to A_0.

Using figure 5.3, we obtain a formula for a general dilatation given the pair of points p, q and their images fp, fq. In order to simplify the notation, for any point p, we write p' instead of fp. The eigenpoint a is the intersection of lines pp' and qq'.

$$a = pp'.qq' = (pqq')p' - (p'qq')p \tag{5.3}$$

Even though the affine space is n-dimensional, the points p, p', q, q' all lie in a plane and so the terms pqq' and $p'qq'$ in equation (5.3) are to be regarded as pseudonumbers for the plane under consideration. Let r be a new point[6] and let us construct $r' = fr$. The point r' will lie at the intersection of the radial line ar and the line $p'(pr.A_0)$.

$$\begin{aligned}
r' &= ar.[p'(pr.A_0)] = (pp'.qq')r.[p'((pA_0)r - (rA_0)p)] \\
&= [(pqq')p'r - (p'qq')pr][(pA_0)p'r - (rA_0)p'p] \\
&= -(pqq')(rA_0)p'r.p'p - (p'qq')(pA_0)pr.p'r + (p'qq')(rA_0)pr.p'p \\
&= (pqq')(rA_0)(rp'p)p' - (p'qq')(pA_0)(pp'r)r - (p'qq')(rA_0)(rp'p)p \\
&= (rp'p)[(pqq')(rA_0)p' + (p'qq')(pA_0)r - (p'qq')(rA_0)p]
\end{aligned}$$

Re-weighting,

$$\begin{aligned}
r' &= r + \frac{(rA_0)(pqq')p'}{(pA_0)(p'qq')} - \frac{(rA_0)p}{(pA_0)} = r + \frac{(rA_0)}{(pA_0)}\left(\frac{(pqq')p'}{(p'qq')} - p\right) \\
&= r + \frac{(rA_0)}{(pA_0)}\left(\frac{(pqq')p' - (p'qq')p}{(p'qq')}\right) = r + \frac{(rA_0)(pp'.qq')}{(pA_0)(p'qq')} \ . \qquad (5.4)
\end{aligned}$$

Equation (5.4) is the formula for a general dilatation in Whitehead's algebra. If r is on the hyperplane so that $rA_0 = 0$ then it gives $r' = r$, which is the identity transformation. If r is the eigenpoint a given by equation (5.3), then equation (5.4) confirms that $a' = fa = \alpha a$ with the eigenvalue,

$$\alpha = \frac{(p'A_0)(pqq')}{(pA_0)(p'qq')} \ . \qquad (5.5)$$

Another form of the dilatation is,

$$r' = r + \frac{(\alpha - 1)(rA_0)a}{(aA_0)} \qquad (5.6)$$

which is obtained by substituting equations (5.3) and (5.5) into (5.4).

5.2.3 Translations and vectors

A translation is a special case of a dilatation which is obtained when the eigenpoint a in figure 5.3 lies in the hyperplane at infinity so that $aA_0 = 0$. Substituting this condition in equation (5.3) for the eigenpoint, we obtain,

$$\frac{(pqq')}{(p'qq')} = \frac{(pA_0)}{(p'A_0)} \ .$$

[6]Note that, in contrast to the way it is shown in figure 5.3, r need not be in the plane pqp'.

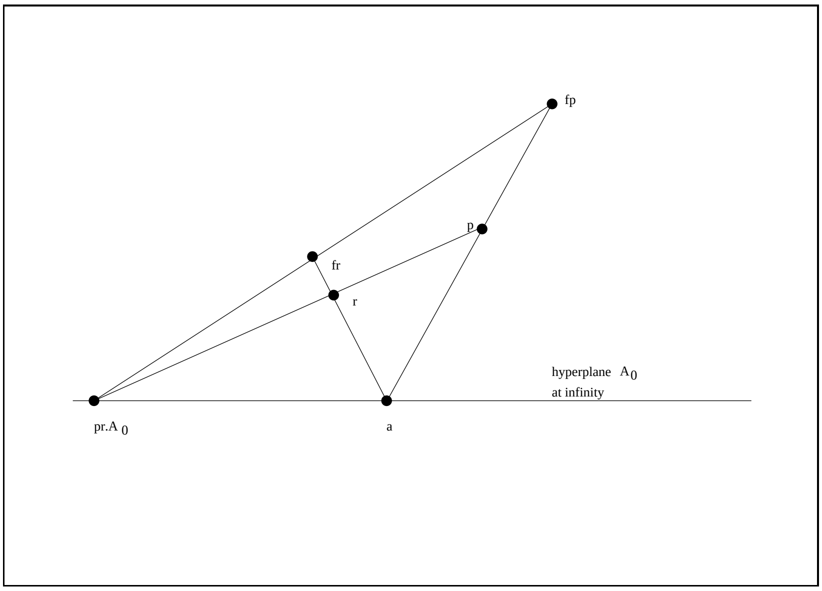

Figure 5.4: A translation

Substituting this result in equation (5.4) for the general dilatation gives,

$$r' = r + \frac{(rA_0)}{(pA_0)}\left(\frac{(pA_0)p'}{(p'A_0)} - p\right) = r + (rA_0)\left(\frac{p'}{(p'A_0)} - \frac{p}{(pA_0)}\right) . \qquad (5.7)$$

This is the formula for a general translation in Whitehead's algebra. Figure 5.4 shows the translation by simply re-drawing figure 5.3 with the eigenpoint of the dilatation moved onto the hyperplane at infinity. Notice that opposite sides of the quadrilateral with points p, p', r, r' in figure 5.4 are parallel because they meet at infinity. Therefore, the quadrilateral is in fact a parallelogram.

If we adjust the weights of all finite points p so that $pA_0 = 1$, the translation of equation (5.7) takes the simple form,

$$r' = r + p' - p\ . \qquad (5.8)$$

With this normalization, we have arrived at an equation of vector algebra. The point r is moved to r' by the addition of a "displacement vector" $(p'-p)$. If we re-arrange equation (5.8) to read $(r'-r) = (p'-p)$ then we see the origin of the assertion in vector algebra that "vectors" can be "moved around". However, in Whitehead's algebra, the vector $(p'-p)$ is not an arrow extending from p to p' which can be moved around the space. Instead, the vector $(p'-p)$ is a point on the hyperplane at infinity because $(p' - p)A_0 = p'A_0 - pA_0 =$

$1 - 1 = 0$. The vector $(p' - p)$ is the eigenpoint a in figure 5.4. Most mathematics that is applied to physics starts with a vector space. This seems a little bizarre from the Cayley-Klein view-point because, as we have seen, vectors only arise naturally by studying affine space and then artificially restricting the natural group of congruences - the dilatations - to the subgroup of translations. The current work is an attempt to set out a more natural geometric algebra for mathematical physics.

5.3 Minkowski space-time

Minkowski space-time is a 4-d affine space in which the hyperplane at infinity is a hyperbolic space. The light cone structure is provided by the absolute hyperbolic quadric at infinity as indicated in figure 5.1. The hyperplane at infinity is a 3-d hyperbolic space which can be handled using the formulae of Whitehead's algebra set out in chapter 1.

5.3.1 The hyperbolic space at infinity

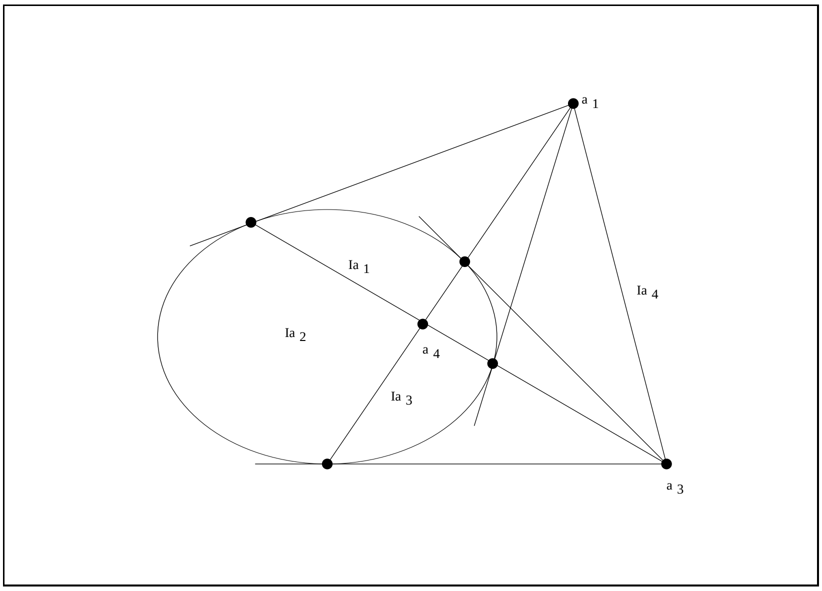

Figure 5.5: Hyperbolic space at infinity

Let I be the polarity of the 3-d hyperbolic space at infinity. This polarity defines the real quadric surface $pIp = 0$ where p is a point. The sign of pIp changes as p crosses the quadric surface. The sign of the polarity is chosen so that $pIp < 0$ for p inside the quadric surface. The reference points for the hyperbolic space are chosen as follows. Start with an arbitrary point a_4 inside the quadric. Pick any plane through a_4 and call it Ia_1 so that $0 = a_4 I a_1 = a_1 I a_4$. Now pick a_2 on the intersection of planes Ia_1 and Ia_4 so that $0 = a_2 I a_1$ and $0 = a_2 I a_4$. Finally, a_3 is the unique point at the intersection of planes Ia_1, Ia_2 and Ia_4. Therefore $a_j I a_k = 0$ for $j \neq k$. Point a_4 is inside the quadric surface so that $a_4 I a_4 < 0$ and by adjusting its intrinsic weight we can set $a_4 I a_4 = -1$. The other points are all outside the quadric, and by adjusting their intrinsic weights we set $a_j I a_j = 1$ for $j = 1, 2, 3$.

Figure 5.5 shows a 2-d slice through the 3-d hyperbolic space. The plane of the figure is Ia_2 so that the point a_2 is out of the page. In the figure, notice that for any point a outside the quadric, a tangent line ap to the quadric, which touches the quadric at p, has the property that p is also on Ia. The reason for this is because the tangent line[7] has $0 = ap.Iap = (aIa)(pIp) - (aIp)^2$. Point p is on the quadric so $pIp = 0$ and hence $0 = aIp = pIa$ and so p is on Ia.

In order to make the light cone an invariant structure, the polarity must be invariant to all observers. This means that the $f^{(1)}$ part of the general affinity $f = f^{(2)} f^{(1)}$ in section 5.2.1 must commute[8] with the polarity I. In other words, $f^{(1)}$ must be a congruence of the 3-d hyperbolic space at infinity. We have already seen in section 5.2.3 that the $f^{(2)}$ part can represent translations. Consequently, the $f^{(2)}$ part of the general affinity produces spacelike and timelike translations, and so the $f^{(1)}$ part must produce the spacelike rotations and the boosts.

5.3.2 Natural coordinates for Minkowski space

Natural coordinates can be obtained by applying a sequence of congruences to a reference point[9]. In this case, if we first apply a timelike translation to a reference point a_0 in the finite space. The translation is in the direction of the point a_4. From equation (5.8), the result is $f_t a_0 = a_0 + t a_4$ where t is the weight of the displacement in time. The next congruence is a spacelike translation in the direction of the point a_1 with a weight x. The result is $g_x f_t a_0 = a_0 + t a_4 + x a_1$. Continuing in this way, we obtain a general space-time point,

$$p(x, y, z, t) = f a_0 = a_0 + x a_1 + y a_2 + z a_3 + t a_4 \tag{5.9}$$

by applying translations in the direction of points a_2 and a_3.

[7] See figure 3.3.
[8] See section 3.5.
[9] See section 3.6.2.

5.3.3 The Lorentz boost

Observer Alice sees a space-time point p given by equation (5.9). Bob, who is boosted along Alice's z-axis, sees her point p as $f_\psi p$ where f_ψ is a congruence of the hyperbolic space at infinity with invariant line $ILM = a_4 a_3$. Let's show that we recover the Lorentz transformation by applying the general formula for a congruence in Whitehead's algebra given by equation (3.27). The subspace $LM = I^2 LM = Ia_4.Ia_3$ and $ILM.LM = a_4a_3.Ia_4a_3 = -1$.

$$
\begin{aligned}
f_\psi p = \exp\left(\frac{\psi ILM.I}{\sqrt{-ILM.LM}}\right) p = (1-\cosh(\psi))\left(\frac{(p.ILM).LM}{ILM.LM}\right) \\
+\cosh(\psi)p + \sinh(\psi)\left(\frac{ILM.Ip}{\sqrt{-ILM.LM}}\right) = a_0 + xa_1 + ya_2 \\
+z(\cosh(\psi)a_3 + \sinh(\psi)a_4a_3.Ia_3) + t(\cosh(\psi)a_4 + \sinh(\psi)a_4a_3.Ia_4) \\
= a_0 + xa_1 + ya_2 + (z\cosh\psi - t\sinh\psi)a_3 + (t\cosh\psi - z\sinh\psi)a_4
\end{aligned}
\tag{5.10}
$$

Bob measures coordinates x', y', z', t' for the point $f_\psi p$ by the assignment $p(x', y', z', t') = f_\psi p(x, y, z, t)$. In this way we get the standard Lorentz transformation,

$$
x' = x \ , \ y' = y \ , \ z' = z\cosh\psi - t\sinh\psi \ , \ t' = t\cosh\psi - z\sinh\psi \ .
$$

It is easy to show that Bob sees Alice moving along his z-axis with velocity $dz'/dt' = -\tanh\psi$. This confirms that a congruence with invariant line $ILM = a_4a_3$ and parameter ψ boosts Bob along Alice's positive z-axis.

5.3.4 Dilatations of affine space-time

Let us now repeat the argument of section 5.3.3 for a dilatation. Take the centre of the dilatation as the general space-time point,

$$
a = a_0 + x_a a_1 + y_a a_2 + z_a a_3 + t_a a_4 \ .
$$

From equation (5.6), the dilatation is,

$$
\begin{aligned}
fp &= p + (\alpha - 1)a \\
&= \alpha(a_0 + a_1(\alpha^{-1}(x - x_a) + x_a) + a_2(\alpha^{-1}(y - y_a) + y_a) \\
&\quad + a_3(\alpha^{-1}(z - z_a) + z_a) + a_4(\alpha^{-1}(t - t_a) + t_a) \ .
\end{aligned}
$$

So, Alice sees a point $p(x, y, z, t)$ and Bob, who is dilatated with respect to Alice, sees her point p as fp. Bob measures coordinates x', y', z', t' for the point fp by the assignment $p(x', y', z', t') = \lambda fp(x, y, z, t)$ where λ is an unimportant weight. In this way we get the following transformation

$$\begin{aligned} x' = \alpha^{-1}(x - x_a) + x_a \ , \ y' = \alpha^{-1}(y - y_a) + y_a \\ z' = \alpha^{-1}(z - z_a) + z_a \ , \ y' = \alpha^{-1}(t - t_a) + t_a \ . \end{aligned} \tag{5.11}$$

In order to simplify the argument, suppose that the centre of the dilatation is a_0 and $\alpha \gg 1$. The transformation now simplifies to,

$$x' = \alpha^{-1}x \ , \ y' = \alpha^{-1}y \ , \ z' = \alpha^{-1}z \ , \ t' = \alpha^{-1}t$$

which means that Bob sees Alice's measuring rods contracted and her clocks appear to run fast. Spectral lines measured by Bob are blue-shifted with respect to those measured by Alice, but *they both agree on the speed of light.* The situation is not symmetric. Alice sees spectral lines red-shifted with respect to those measured by Bob.

5.3.5 Distance in affine space-time

We define the distance between two finite space-time points p and q as,

$$s = \frac{\sqrt{(pq.A_0)I(pq.A_0)}}{(pA_0)(qA_0)} \ . \tag{5.12}$$

In coordinates, this evaluates to the standard formula.

$$s^2 = (x_q - x_p)^2 + (y_q - y_p)^2 + (z_q - z_p)^2 - (t_q - t_p)^2$$

Suppose that Alice evaluates the distance s between p and q according to equation (5.12). Bob sees the points as fp and fq and he evaluates the distance between them as,

$$\begin{aligned} \frac{\sqrt{(fpfq.A_0)I(fpfq.A_0)}}{(fpA_0)(fqA_0)} &= \frac{\sqrt{(fpfq.\alpha fA_0)I(fpfq.\alpha fA_0)}}{(fp.\alpha fA_0)(fq.\alpha fA_0)} \\ = \frac{\sqrt{f^{(2)}f^{(1)}(pq.A_0)If^{(2)}f^{(1)}(pq.A_0)}}{\alpha(pA_0)(qA_0)} &= \frac{\sqrt{f^{(1)}(pq.A_0)If^{(1)}(pq.A_0)}}{\alpha(pA_0)(qA_0)} \\ = \frac{\sqrt{f^{(1)}(pq.A_0).f^{(1)}I(pq.A_0)}}{\alpha(pA_0)(qA_0)} &= \frac{\sqrt{(pq.A_0)I(pq.A_0)}}{\alpha(pA_0)(qA_0)} = \frac{s}{\alpha} \end{aligned}$$

where we used equation (5.1), the outermorphism property, the form $f = f^{(2)}f^{(1)}$ of the general affinity from section 5.2.1, and the fact that $f^{(1)}$ commutes with I. The distance is not invariant because it transforms as $s \to \alpha^{-1}s$. Only when the dilatation part of the affinity is restricted to be a translation, so that $\alpha = 1$, is the distance an invariant.

5.3.6 The problem of dilatations of affine space-time

In the real world we experience spacelike rotations, timelike rotations (boosts), spacelike translations and timelike translations. Nevertheless, sections 5.3.4 and 5.3.5 have shown that the group of congruences of affine (Minkowskian) space-time also contains the dilatations[10]. The dilatations are an expansion or contraction of space and time by the same factor α so that velocities, such as the speed of light, are unchanged. The introduction indicated that dilatations seem to be either ignored (Sir Robert Ball [15]) or dismissed as a change in the units of length and time (Semple and Kneebone [16]). Semple and Kneebone give no account of why rotations and translations correspond to real motions, yet dilatations only correspond to a subjective change in the units of measurement in the mind of an observer. In contrast to Semple and Kneebone, we take the dilatations seriously as corresponding to motions as real as a rotation or a boost. Once this is accepted, it follows from the fact that we do not experience dilatations, that the group of congruences active in the real world is not that of Minkowski space-time and we should look for some other geometry, with a different group of congruences, which is in accord with our experience of the real world. As mentioned in the introduction, the space formed by the ultra-infinite points of hyperbolic geometry, which is called de Sitter space in General Relativity, has a natural light-cone structure, and a physically sensible group of congruences without any dilatations.

5.4 De Sitter space

The ultra-infinite points of hyperbolic geometry possess the natural light cone structure shown in figure 5.2. The treatment of 4-d hyperbolic geometry is an extension of section 5.3.1 on 3-d hyperbolic geometry. Let I be the absolute polarity of 4-d hyperbolic space. Take reference points $a_1, \ldots, a_5$ with $a_j I a_k = 0$ for $j \neq k$, $a_j I a_j = 1$ for $j = 1, 2, 3, 4$ and $a_5 I a_5 = -1$. The corresponding way in which the reference points for 3-d hyperbolic space were chosen in section 5.3.1 makes it clear that a_1 (say) can be chosen arbitrarily. This means that the space-time looks the same from any point. Figure 5.6 shows de Sitter space. If Alice is at a_1 then $a_1 - a_5$ and $a_1 + a_5$ are her infinite past and infinite future respectively. The light cone from a_1 touches the quadric at points $a_4 - a_5$ and $a_4 + a_5$. De Sitter space is standard hyperbolic geometry. All the formulae of chapter 3 just have to be worked out using the

[10] Section 7.4.2 shows how dilatations arise in the model of Euclidean geometry on the absolute quadric

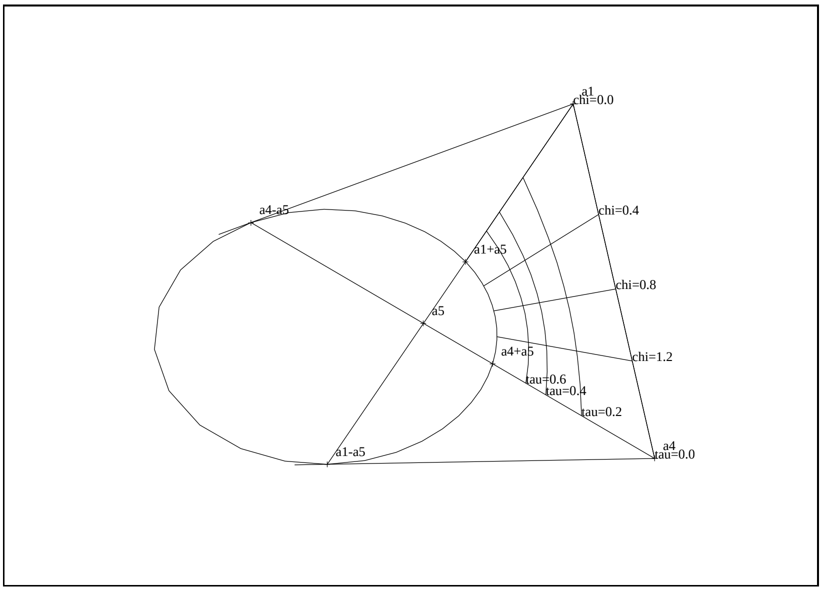

Figure 5.6: De Sitter space

above polarity.

5.4.1 The line element in natural coordinates

We get a natural system of coordinates for de Sitter space by exponentiating a single reference point as in section 3.6.2.

$$\begin{aligned} p(\chi,\theta,\phi,\tau) &= e^{\phi a_2 a_3 . I} e^{\theta a_4 a_2 . I} e^{\chi a_1 a_4 . I} e^{\tau a_1 a_5 . I} a_1 \\ &= a_1 \cosh\tau \cos\chi + a_2 \cosh\tau \sin\chi \sin\theta \cos\phi \\ +a_3 \cosh\tau \sin\chi \sin\theta \sin\phi &+ a_4 \cosh\tau \sin\chi \cos\theta + a_5 \sinh\tau \ . \end{aligned} \tag{5.13}$$

Table 5.1 gives the meaning of the invariant lines ILM for the de Sitter group of congruences used in the arguments of the exponentials.

Equation (3.35) gives the line element in Whitehead's algebra as,

$$(ds)^2 = \frac{-pdp.Ipdp}{(pIp)^2} \ .$$

Upon substituting $p(\chi,\theta,\phi,\tau)$ from equation (5.13) into this formula, we obtain, after some tedious calculation,

$$(ds)^2 = (d\tau)^2 - \cosh^2\tau((d\chi)^2 + \sin^2\chi((d\theta)^2 + \sin^2\theta(d\phi)^2)) \ . \tag{5.14}$$

ILM	LM	Congruence
a_1a_2	$a_3a_4a_5$	Translate along x
a_1a_3	$a_2a_4a_5$	Translate along y
a_1a_4	$a_2a_3a_5$	Translate along z
a_1a_5	$a_2a_3a_4$	Time translation
a_2a_3	$a_1a_4a_5$	Rotation about z
a_2a_4	$a_1a_3a_5$	Rotation about y
a_2a_5	$a_1a_3a_4$	Boost along x
a_3a_4	$a_1a_2a_5$	Rotation about x
a_3a_5	$a_1a_2a_4$	Boost along y
a_4a_5	$a_1a_2a_3$	Boost along z

Table 5.1: Congruences for observer at a_1

This is the same as the first line element in section 5.2 of the book by Hawking and Ellis [17]. It confirms that the space-time formed by the ultra-infinite points of 4-d hyperbolic geometry is in fact the same as de Sitter space studied in General Relativity. Figure 5.6 shows the hypersurfaces of constant χ and constant τ. This figure is equivalent to figure 16(i) in [17].

5.4.2 Radar coordinates for de Sitter space

The positivist's view of de Sitter space is obtained when a radar is used to measure coordinates. In figure 5.7 a radar transmits a pulse at a_1 along the positive z-axis. The pulse is reflected from a target p. The reflected pulse crosses the worldline of the radar set a time $\Delta\tau$ after transmission. The congruence of translation in time is $f_{\Delta\tau} = \exp(\Delta\tau a_1 a_5.I)$. We get the path of the reflected pulse by time translation of the light ray $a_1(a_4 - a_5)$. The reflected ray is therefore the line $f_{\Delta\tau}(a_1(a_4 - a_5))$. The transmitted ray is the line $a_1(a_4 + a_5)$. The target p is the intersection of these two rays.

$$\begin{aligned} p &= f_{\Delta\tau}(a_1(a_4 - a_5)).a_1(a_4 - a_5) \\ &= a_1 \cosh(\Delta\tau/2) + (a_4 + a_5)\sinh(\Delta\tau/2) \end{aligned} \tag{5.15}$$

This is the position of a target that is detected by a pulse transmitted at $\tau = 0$. A target that is detected by a pulse transmitted at $\tau = \tau_1$ is found by time translation of p in equation (5.15) to $f_{\tau_1}p$. Calculating, we find,

$$f_{\tau_1}p = a_1 \cosh(\tau_1 + (\Delta\tau/2)) + a_4 \sinh(\Delta\tau/2) + a_5 \sinh(\tau_1 + (\Delta\tau/2)) \ .$$

The round trip time for the radar pulse is $\Delta\tau$, so we *define* the time coordinate of the target as $t = \tau_1 + \Delta\tau/2$. Similarly, the range to the target is

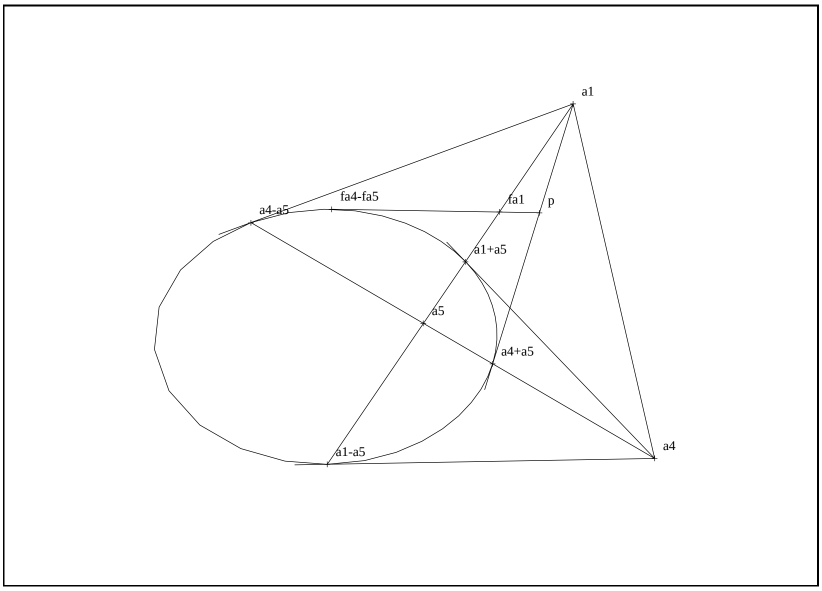

Figure 5.7: Operation of a radar in de Sitter space

defined as $r = \Delta\tau/2$. We drop the notation $f_{\tau_1}p$, and write the target as $p(r,t)$.

$$p(r,t) = a_1 \cosh t + a_4 \sinh r + a_5 \sinh t$$

The radar currently points along the positive z-axis. In order to detect targets all over the sky, we rotate the radar to point in the direction given by the polar angles (θ, ϕ). Fortunately, spatial rotations commute with time translations (see table 5.1) so this works. A target in a general position is now given the coordinates,

$$\begin{aligned} p(r,\theta,\phi,t) = e^{\phi a_2 a_3 . I} e^{\theta a_4 a_2 . I} p(r,t) &= a_1 \cosh t + a_2 \sinh r \sin\theta \cos\phi \\ &+ a_3 \sinh r \sin\theta \sin\phi + a_4 \sinh r \cos\theta + a_5 \sinh t \ . \end{aligned} \quad (5.16)$$

Repeating the calculations of section 5.4.1 for $p(r,\theta,\phi,t)$, we obtain the line element of de Sitter space in radar coordinates as,

$$(ds)^2 = \frac{(dt)^2 - (dr)^2 - \sinh^2 r((d\theta)^2 + \sin^2\theta(d\phi)^2))}{\cosh^2 r} \ . \quad (5.17)$$

Figure 5.8 shows the hypersurfaces of constant r and t. The figure clearly shows the horizon experienced by an observer whose world-line is $a_1(a_1 + a_5)$. The observer cannot measure anything that is beyond the line $a_4(a_1 + a_5)$ in the figure.

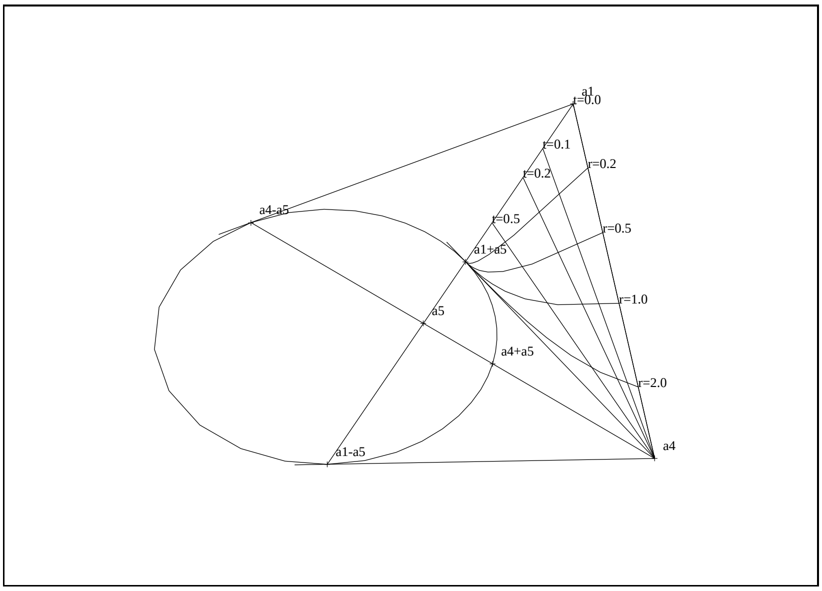

Figure 5.8: Radar coordinates in de Sitter space

It is interesting to calculate the volume of the spacelike shell of thickness dt. From section 4.1, the volume element of 4-d de Sitter space in radar coordinates is,

$$dV = \frac{p\frac{\partial p}{\partial r}\frac{\partial p}{\partial \theta}\frac{\partial p}{\partial \phi}\frac{\partial p}{\partial t}drd\theta d\phi dt}{(pIp)^{5/2}} = \left(\frac{\sinh^2 r \sin\theta drd\theta d\phi dt}{\cosh^4 r}\right) a_1a_2a_3a_4a_5$$

Upon integrating out the variables r, θ and ϕ, the volume of the spacelike shell of thickness dt is found to be

$$dV = \left(\frac{4\pi dt}{3}\right) a_1a_2a_3a_4a_5 \ .$$

So, in spite of the fact that the range r can be infinite, the volume of a space-like shell of constant t is finite and independent of the time.

5.4.3 Horospherical coordinates

We get another interesting natural system of coordinates for de Sitter space by exponentiating a single reference point with the following sequence of congruences.

$$\begin{aligned} p(\chi,\theta,\phi,\tau) &= e^{\phi a_2 a_3 .I} e^{\theta a_4 a_2 .I} e^{\chi(a_1-a_5)a_4 .I} e^{\tau a_1 a_5 .I} a_1 \\ &= a_1\cosh\tau + a_5\sinh\tau - (\chi^2/2)e^{\tau}(a_1-a_5) \\ &+ \chi e^{\tau}(a_2\sin\theta\cos\phi + a_3\sin\theta\sin\phi + a_4\cos\theta) \qquad (5.18) \end{aligned}$$

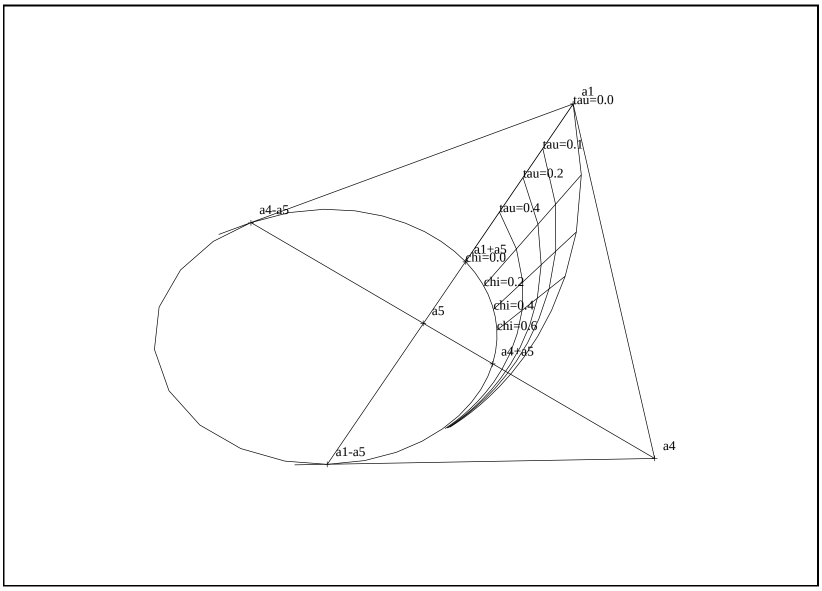

Figure 5.9: Horospherical coordinates in de Sitter space

This only difference between the coordinates of equation (5.13) and the above system is that the radial coordinate χ has been changed to the parameter of a parabolic congruence with invariant line $ILM = (a_1 - a_5)a_4$. The parabolic congruences were considered in section 3.10. Figure 5.9 shows the hypersurfaces of constant χ and constant τ. The constant τ hypersurfaces are made from curves generated by varying the χ parameter. These curves are a type of circle called *horocycles* in hyperbolic geometry. Hence the constant τ hypersurfaces are *horospherical.* The line element in horospherical coordinates is found to be,

$$(ds)^2 = (d\tau)^2 - e^{2\tau}((d\chi)^2 + \chi^2((d\theta)^2 + \sin^2\theta(d\phi)^2)) \ . \tag{5.19}$$

Let us change to Cartesian coordinates,

$$x = \chi \sin\theta \cos\phi \ , \ y = \chi \sin\theta \sin\phi \ , \ z = \chi \cos\theta$$

on the horospheres of constant τ, or equivalently, generate the Cartesian system by a sequence of parabolic congruences,

$$p(x, y, z, \tau) = e^{x(a_1-a_5)a_2.I} e^{y(a_1-a_5)a_3.I} e^{z(a_1-a_5)a_4.I} e^{\tau a_1 a_5.I} a_1 \ .$$

In both cases the point is,

$$p = a_1 \cosh\tau + a_5 \sinh\tau - (1/2)(x^2 + y^2 + z^2)e^{\tau}(a_1 - a_5) + e^{\tau}(xa_2 + ya_3 + za_4)$$

and the line element is,

$$(ds)^2 = (d\tau)^2 - e^{2\tau}((dx)^2 + (dy)^2 + (dz)^2) \ . \tag{5.20}$$

This shows that the geometry on the spacelike horospherical hypersurface is clearly Euclidean[11]. Equation (5.20) is the same as the second line element in section 5.2 of the book by Hawking and Ellis [17] and our figure 5.9 is equivalent to figure 16(ii) in [17]. This line element is also considered at greater length in section 6 of Schrödinger's book [18], where the coordinates are known as Lemaître's frame.

5.5 Summary and conclusion

Whitehead's algebra has been used to construct Minkowski space-time. It is an affine space. Its hyperplane at infinity is a hyperbolic space which supplies the light cone structure. The group of congruences of this affine space-time includes the dilatations in addition to the Poincaré group. Since we never experience a dilatation, then space-time cannot be Minkowskian. Another possible space-time is the ultra-infinite points of hyperbolic geometry. It has a natural light cone structure and it is the same as de Sitter space in General Relativity. There do not appear to be any other ways of getting a natural light cone structure in elliptic or hyperbolic geometry. The de Sitter group of congruences is physically sensible and it does not contain the problematic dilatations. We obtained some properties of de Sitter space that are normally obtained using the methods of General Relativity. In General Relativity, de Sitter space is a solution of the field equations. In contrast, de Sitter space emerges from Whitehead's algebra as the natural space-time prior to any consideration of the laws governing force and matter. Our proposal, and programme for the further development of Whitehead's algebra, is to study physics in a de Sitter space background instead of the usual Minkowski background.

[11] See also section 3.10.1.

Chapter 6

Müller's Theorems

6.1 Introduction

This chapter uses the fundamental formulae of section 1.14 to prove Müller's theorems which are found in article 106 of [10]. Müller's theorems investigate the conditions under which we can write relations of the form,

$$X(YZ) = \alpha(XY)Z + \beta(XZ)Y \tag{6.1}$$

between three linear subspaces X, Y, and Z where α and β are numerical weights. It is useful to have Müller's theorems to hand for proofs in Whitehead's algebra.

We will need some reference points. Let $a_1, \ldots, a_n$ be a set of reference points with $A_1, \ldots, A_n$ the dual hyperplanes so that $a_i A_j = \delta_{ij}$. Also, let $b_1, \ldots, b_n$ be another set of reference points with $B_1, \ldots, B_n$ the dual hyperplanes so that $b_i B_j = \delta_{ij}$.

6.2 Case 1: Y and Z intersecting.

Let us first consider the case in which Y and Z intersect. Let,

$$\begin{aligned} X_q &= b_1 \ldots b_q \\ Y_r &= A_1 \ldots A_r \\ Z_s &= A_{r+1} \ldots A_{r+s} \end{aligned}$$

so that $Y_r Z_s = A_1 \ldots A_{r+s}$ is the intersection of the subspaces. The hyperplane grade of $Y_r Z_s$ is $r + s$ whilst the hyperplane grade of X_q is $n - q$. In order to avoid the trivial case where $X_q Y_r Z_s$ is associative we require

$n-q+r+s > n$ which implies that $q < r+s$. Using formula (1.41) the lhs of (6.1) is,

$$X_q.Y_rZ_s = \sum_\lambda (X_q.[Y_rZ_s]_q^{(\lambda)})[Y_rZ_s]_{r+s-q}^{(\lambda)} \tag{6.2}$$

where we have factored Y_rZ_s into a product of $r+s-q$ hyperplanes and q hyperplanes denoted by $Y_rZ_s = [Y_rZ_s]_{r+s-q}^{(\lambda)}.[Y_rZ_s]_q^{(\lambda)}$ and λ is a superscript label which ranges over the ${}^{r+s}C_q$ ways of taking q hyperplanes out of the $r+s$ hyperplanes of $Y_rZ_s = A_1 \ldots A_{r+s}$. In other words, λ ranges over all the ways making the factorization. Notice that the term $(X_q.[Y_rZ_s]_q^{(\lambda)})$ in the summation is a number.

6.2.1 Case 1.1: X does not intersect Y or Z.

In order to make progress we have to make some assumptions about the factors X_qY_r and X_qZ_s. If we assume that $q \leq r$ and $q \leq s$ then this implies that X_q does not intersect Y_r and also does not intersect Z_s. The first term on the rhs of (6.1) is then,

$$(X_qY_r)Z_s = (b_1 \ldots b_q.A_1 \ldots A_r)A_{r+1} \ldots A_{r+s}$$

and it is a product of $r-q+s$ hyperplanes and in this product the hyperplanes $A_{r+1} \ldots A_{r+s}$ are present. Similarly, the second term on the rhs of (6.1) will always contain the hyperplane factors $A_1 \ldots A_r$ intact. This is a problem for the summation of terms on the rhs of (6.2) because the $r+s-q$ hyperplane product term $[Y_rZ_s]_{r+s-q}^{(\lambda)}$ will sometimes not contain intact the r hyperplanes $A_1 \ldots A_r$ and also not contain intact the $A_{r+1} \ldots A_{r+s}$. In this case the rhs of (6.1) cannot be equal to the lhs. The only way to avoid this situation is if $q = 1$ because then the term $[Y_rZ_s]_{r+s-q}^{(\lambda)}$ cannot simultaneously disrupt the hyperplane factors of Y_r and Z_s and so in each term in the summation on the rhs of (6.2) the factor Y_r or Z_s remains intact and so the summation can be replicated by the rhs of (6.1). In this case (6.1) becomes,

$$b_1.A_1 \ldots A_{r+s} = (b_1A_{r+s})A_1 \ldots A_{r+s-1} - (b_1A_{r+s-1})A_1 \ldots A_{r+s-2}A_{r+s} + \ldots$$

and the rhs follows from an application of (1.41). By keeping track of the signs it is straighforward to see that,

$$b_1(Y_rZ_s) = Y_r(b_1Z_s) + (-1)^s(b_1Y_r)Z_s \ . \tag{6.3}$$

This is the first one of Müller's theorems. The conditions are that Y_r and Z_s are subspaces of hyperplane grade r and s respectively and b_1 is a point.

6.2.2 Case 1.2: X does not intersect Y and X intersects Z.

If we assume that $q \leq r$ as before, but this time $s \leq q$ then this implies that X_q does not intersect Y_r but X_q and Z_s do intersect. Let's begin by examining the terms on the rhs of (6.1). The first term is,

$$X_q Y_r . Z_s = (b_1 \dots b_q . A_1 \dots A_r) A_{r+1} \dots A_{r+s}$$

so that $(b_1 \dots b_q . A_1 \dots A_r)$ will be a product of $r - q$ hyperplanes. Hence the overall result will be these $r - q$ hyperplanes times the s hyperplanes of $A_{r+1} \dots A_{r+s}$. The important thing to notice is that the overall product of $r - q + s$ hyperplanes will contain the s hyperplanes of $Z_s = A_{r+1} \dots A_{r+s}$ intact. The second term is,

$$X_q Z_s . Y_r = (b_1 \dots b_q . A_{r+1} \dots A_{r+s}) A_1 \dots A_r$$

so that $(b_1 \dots b_q . A_{r+1} \dots A_{r+s})$ will be a product of $q - s$ points. When this factor of point-grade $q - s$ is multiplied by $A_1 \dots A_r$ the overall result will be terms containing $r - (q - s)$ hyperplanes from the set $A_1 \dots A_r$. Now, the lhs of (6.1) is given by (6.2) which will be a sum of terms, each being a product of $r + s - q$ hyperplanes from the set $A_1 \dots A_r A_{r+1} \dots A_{r+s}$. However, if this is to agree with the rhs of (6.1), we must choose all of the hyperplanes $Z_s = A_{r+1} \dots A_{r+s}$ and $r - q$ from $Y_r = A_1 \dots A_r$ or else choose none of the hyperplanes from $Z_s = A_{r+1} \dots A_{r+s}$ and all $r + s - q$ from $Y_r = A_1 \dots A_r$. This is only possible[1] if $s = 1$. Now with the single hyperplane Z_1 in (6.2),

$$\begin{aligned}
X_q . Y_r Z_1 &= \sum_\lambda (X_q . [Y_r Z_1]_q^{(\lambda)}) [Y_r Z_1]_{r+1-q}^{(\lambda)} \\
&= (-1)^q \sum_\lambda (X_q . [Y_r]_q^{(\lambda)}) [Y_r]_{r-q}^{(\lambda)} Z_1 + \sum_\lambda (X_q . [Y_r]_{q-1}^{(\lambda)} Z_1) [Y_r]_{r-(q-1)}^{(\lambda)} \\
&= (-1)^q \sum_\lambda (X_q . [Y_r]_q^{(\lambda)}) [Y_r]_{r-q}^{(\lambda)} Z_1 + (-1)^{q-1} \sum_\lambda (X_q Z_1 . [Y_r]_{q-1}^{(\lambda)}) [Y_r]_{r-(q-1)}^{(\lambda)}
\end{aligned}$$

where the sign $(-1)^q$ is because, from (1.40),

$$Y_r Z_1 = (-1)^q [Y_r]_{r-q}^{(\lambda)} Z_1 [Y_r]_q^{(\lambda)}.$$

[1] An example is helpful here. Suppose $X = abc$, $Y = PQRS$, and $Z = TU$ where a, b, c are points and P, Q, R, S, T, U are hyperplanes. From (1.41),

$$abc.(PQRS.TU) = abc.(PQRSTU) = \dots - (abc.RST)PQU + \dots$$

and a term such as PQU is not allowed on the rhs of (6.1) because it does not contain all of the hyperplanes of $Z = TU$.

and we have to jump over q hyperplane factors in order to move Z_1 into the correct position. Similarly, the sign $(-1)^{q-1}$ results from jumping over $q-1$ hyperplane factors in the second term. Notice that we have also used the fact that the numerical factor in the second term is associative by result (1.44). We therefore have the second of Müller's theorems,

$$(-1)^q X_q(Y_r Z_1) = (X_q Y_r)Z_1 - (X_q Z_1)Y_r \tag{6.4}$$

and the conditions are that X_q is a subspace of point-grade q, Y_r is a subspace of hyperplane grade r with $q \le r$ and Z_1 is a hyperplane.

6.2.3 Case 1.3: X intersects Y and Z.

This time we take $r \le q$ and $s \le q$. As usual, we examine the terms on the rhs of (6.1). The first term is,

$$X_q Y_r.Z_s = (b_1 \ldots b_q.A_1 \ldots A_r)A_{r+1} \ldots A_{r+s}$$

so that $(b_1 \ldots b_q.A_1 \ldots A_r)$ will be a product of $q-r$ points. Hence the overall result will be these $q-r$ points times the s hyperplanes of $A_{r+1} \ldots A_{r+s}$. Now since the assumption in all of these cases so far has been $q < r+s$ we find $q-r < s$ so that the overall result of the product $X_q Y_r.Z_s$ is a sum of terms each made of $s-(q-r) = r+s-q$ hyperplane factors taken from the set $A_{r+1} \ldots A_{r+s}$. Similarly, the overall result of the product $X_q Z_s.Y_r$ is a sum of terms each made of $r-(q-s) = r+s-q$ hyperplane factors taken from the set $A_1 \ldots A_r$. However, as in all the other cases, the lhs of (6.1) will be a sum of terms each made from $r+s-q$ hyperplane factors taken from the set $A_1 \ldots A_r A_{r+1} \ldots A_{r+s}$. The only way in which both sides of (6.1) are compatible is if $r+s-q = 1$ because each term in the sum on the lhs has got to be made of $r+s-q$ hyperplanes from $A_{r+1} \ldots A_{r+s}$ or $A_1 \ldots A_r$ but *not* taken from both sets. We now evaluate the lhs of (6.1) using expansion (6.2). Since $r+s-q=1$ we always factor $Y_r Z_s$ into a product of $q = r+s-1$ hyperplanes times a single hyperplane. There are two different types of factorization depending on whether the single hyperplane factor is taken from Y_r or Z_s. These two types of factorization are,

$$Y_r Z_s = [Y_r]_1^{(\lambda)}.[Y_r]_{r-1}^{(\lambda)} Z_s$$
$$Y_r Z_s = Y_r[Z_s]_1^{(\lambda)}.[Z_s]_{s-1}^{(\lambda)} = (-1)^{rs}[Z_s]_1^{(\lambda)}.[Z_s]_{s-1}^{(\lambda)} Y_r$$

and so the expansion is,

$$X_q.Y_r Z_s = \sum_\lambda (X_q.[Y_r]_{r-1}^{(\lambda)} Z_s)[Y_r]_1^{(\lambda)} + (-1)^{rs} \sum_\lambda (X_q.[Z_s]_{s-1}^{(\lambda)} Y_r)[Z_s]_1^{(\lambda)} \ .$$

Using the fact (1.44) that the numerical factor is associative, we move some factors together by jumping over the intervening hyperplanes and keeping track of the signs.

$$X_q.Y_rZ_s = (-1)^{(r-1)s}\sum_\lambda (X_qZ_s.[Y_r]^{(\lambda)}_{r-1})[Y_r]^{(\lambda)}_1$$
$$+(-1)^{rs}(-1)^{r(s-1)}\sum_\lambda (X_qY_r.[Z_s]^{(\lambda)}_{s-1})[Z_s]^{(\lambda)}_1 \ .$$

This gives the third of Müller's theorems,

$$X_q.Y_rZ_s = (-1)^{(r-1)s}X_qZ_s.Y_r + (-1)^r X_qY_r.Z_s \tag{6.5}$$

where X_q is a subspace of point-grade q, and Y_r and Z_s are subspaces of hyperplane grade r and s respectively. The conditions are $q < r + s$ and $r \leq q$ and $s \leq q$ and $r + s - q = 1$. Although we have proved these theorems using simple elements, by linearity we can replace our simple elements by compound elements in all three theorems.

6.3 Case 2: Y and Z not intersecting.

Article 106 of [10] proves another three theorems due to Müller. However, these extra theorems are just the duals of the first three and can be obtained by applying a polarity to each side of the original three theorems and using the outermorphism property to change each product of hyperplanes to a product of points and vice-versa.

6.4 Example: The general projection formula

In (6.3), set $b_1 = p$, $Y_r = IH_1 \dots IH_s$ and $Z_s = H_1 \dots H_s$ where p is a point and the H_j are s hyperplanes. Hence $r = n - s$ and (6.3) becomes,

$$p = \frac{(IH_1 \dots IH_s).(p.H_1 \dots H_s)}{(IH_1 \dots IH_s.H_1 \dots H_s)} + \frac{(-1)^s(p.IH_1 \dots IH_s).(H_1 \dots H_s)}{(IH_1 \dots IH_s.H_1 \dots H_s)} \tag{6.6}$$

where we have divided throughout by $(IH_1 \dots IH_s.H_1 \dots H_s)$ because it is a number. Now suppose that p lies in the subspace $H_1 \dots H_s$ so that $p.H_1 \dots H_s = 0$. In this case the first term on the RHS of (6.6) is zero, so the second term is the formula for the projection of p into the subspace $H_1 \dots H_s$. So, we can write (6.6) as $p = p_\perp + p_\parallel$ where,

$$p_\perp \;=\; \frac{(IH_1 \dots IH_s).(p.H_1 \dots H_s)}{(IH_1 \dots IH_s.H_1 \dots H_s)} \tag{6.7}$$

$$p_{\|} \quad = \quad \frac{(-1)^s(p.IH_1 \dots IH_s).(H_1 \dots H_s)}{(IH_1 \dots IH_s.H_1 \dots H_s)} \; . \tag{6.8}$$

Here, $p_{\|}$ is the projection of p into the subspace $H_1 \dots H_s$ and $p_\perp$ is the projection of p into the dual subspace $IH_1 \dots IH_s$. The projections have the property $p_\perp Ip_{\|} = 0$. To see this, note that if $p = p_\perp$ then $p_{\|} = 0$ which means $p_\perp.IH_1 \dots IH_s = 0$. In other words, $p_\perp$ is a linear combination of the s points IH_j. Now we have already noted that $p_{\|}$ lies in the intersection $H_1 \dots H_s$ so that $p_{\|}H_j = 0$ for all s hyperplanes. Now,

$$p_\perp Ip_{\|} = \sum_{j=1}^{s} \xi_j IH_j.Ip_{\|} = (-1)^{n-1} \sum_{j=1}^{s} \xi_j (p_{\|}H_j) = 0 \tag{6.9}$$

which proves the relation.

Equation (3.25) that was used in the derivation of the canonical form of a simple congruence is now seen to be a special case of the general projection formula (6.6). In (6.6), set $s = 2$ and $H_1H_2 = LM$ and we recover (3.25) exactly. Equation (3.25) is therefore seen to be a projection of p onto a line ILM.

A useful formula for projecting a hyperplane $Y = Ip$ is obtained by applying the polarity I to equation (6.6). After some simplification using (1.13),

$$\begin{aligned} Y = Y_{\|} + Y_\perp \quad &= \quad \frac{(IH_1 \dots IH_s.Y).(H_1 \dots H_s)}{(IH_1 \dots IH_s.H_1 \dots H_s)} \\ &\quad + \quad \frac{(IH_1 \dots IH_s).(YH_1 \dots H_s)}{(IH_1 \dots IH_s.H_1 \dots H_s)} \end{aligned} \tag{6.10}$$

and, of course, $IY_\perp.Y_{\|} = 0$.

Chapter 7

Computational Euclidean Geometry Modelled on the Absolute Quadric of Elliptic Space

7.1 Introduction

There are two ways to model Euclidean geometry in the Kleinian framework. Firstly, we can proceed as in chapter 5 where we start with projective space, make it an affine space by picking an absolute hyperplane at infinity, then make it into a Euclidean geometry by making the hyperplane at infinity into an elliptic hyperplane. The second posssibility, which is the one we explore in this chapter, is to model Euclidean geometry[1] on an absolute quadric hyperface using the result found in section 3.10.1.

The theoretical material in this chapter has been used to make some function packages for serious computational geometry that run under the scientific interpreter Yorick. These packages are more sophisticated versions of the toy package that was described in chapter 2. So, in the current chapter, each item of theoretical material is illustrated by an example using the Yorick functions. By seeing the constructs of Whitehead's algebra juxtaposed to the computations using the Yorick functions, it is hoped that the reader will gradually gain enough familiarity with the techniques to be able to use the functions in engineering applications.

[1]Equation (3.52) has shown that any flat-space geometry can be modelled on the absolute quadric. Consequently, with very little modification, the current chapter could be applied to a Minkowski space-time.

7.2 The choice of the absolute quadric

The example in section 3.10.1 constructed 2-d Euclidean geometry on the 2-d quadric surface of 3-d hyperbolic space. The reason for choosing a hyperbolic space in the example was because the quadric surface $pIp = 0$ is real. To see this, write a general point in 3-d space as,

$$p = \xi_1 a_1 + \xi_2 a_2 + \xi_3 a_3 + \xi_4 a_4 \ . \tag{7.1}$$

This becomes a point in hyperbolic space if we choose a hyperbolic polarity $Ia_1 = A_1, Ia_2 = A_2, Ia_3 = A_3, Ia_4 = -A_4$. The equation of the absolute quadric is now,

$$0 = pIp = \xi_1^2 + \xi_2^2 + \xi_1^3 - \xi_4^2 \ . \tag{7.2}$$

This is a real surface in the sense that if real values are assigned to the weights ξ_1, ξ_2, ξ_3, then (7.2) fixes real values for ξ_4. In other words, the points on the quadric surface have real weights and are real points.

Equation (3.52) showed that the intrinsic geometry of *any* absolute quadric hypersurface $pIp = 0$ is always a flat-space geometry. Furthermore, section 3.4 showed that the elliptic polarity defined by equation (3.13) is, in a sense, the most natural polarity. Consequently, a geometry modelled on the absolute elliptic quadric will be a flat-space geometry. The only problem is that the absolute quadric of 3-d elliptic space has equation,

$$0 = pIp = \xi_1^2 + \xi_2^2 + \xi_1^3 + \xi_4^2 \ . \tag{7.3}$$

which does not give real weights. Consequently, the points on the elliptic quadric are not real points. Nevertheless, if we modify (7.1) to read,

$$p = \xi_1 a_1 + \xi_2 a_2 + \xi_3 a_3 + \xi_4 i a_4$$

where $i = \sqrt{-1}$ then the equation of the elliptic quadric becomes,

$$0 = pIp = \xi_1^2 + \xi_2^2 + \xi_1^3 - \xi_4^2 \ .$$

This is exactly the same as equation (7.2) and has been achieved by simply thinking of the point ia_4 as a reference point in elliptic space. In other words, we can use the natural elliptic polarity and get the same effect as a hyperbolic polarity by judiciously using an imaginary reference point.

At this point, the reader may wonder, "Ok, the elliptic polarity is the most natural one, but why not use whatever polarity is appropriate - elliptic or hyperbolic - instead of introducing an imaginary reference point?"

The answer is to do with the way pseudonumbers are handled in the computational framework. Suppose we are in 4-d hyperbolic space[2] in which everything is real. The reference points are $a_1, \ldots, a_5$ and the dual hyperplanes are $A_1, \ldots, A_5$ with $a_j A_k = \delta_{jk}$. The hyperbolic polarity is $Ia_j = A_j$ for $j = 1, 2, 3, 4$ and $Ia_5 = -A_5$. Now, we would like to be able to assume that the pseudonumber $a_1 a_2 a_3 a_4 a_5 = 1$. In section 1.14, it was shown that this assumption means we do not have to distinguish between numbers and pseudonumbers. This is helpful in the computer implementation because scientific interpreters allow multiplication of arrays by numbers. So, if the elements points, lines, ..., hyperplanes are modelled by arrays, then multiplication by numerical weights is already taken care of by the interpreter syntax. However, if there are two types of weights - numbers and pseudonumbers - then the task of implementing the system in a ready-made scientific interpreter such as Yorick [12] becomes more difficult. Now, if we apply the hyperbolic polarity to the pseudonumber we get,

$$I(a_1 a_2 a_3 a_4 a_5) = -A_1 A_2 A_3 A_4 A_5$$

which leads to a contradiction if $a_1 a_2 a_3 a_4 a_5 = 1$ because equation (1.24) says that $A_1 A_2 A_3 A_4 A_5 = 1$ and so we get $I(1) = 1 = -1$. This contradiction can be removed by always working with an elliptic polarity which avoids the problematical minus sign on the RHS of the above equation[3].

So, here is our model of Euclidean geometry. We are in a (n-1)-d projective space with reference points $a_1, \ldots, a_n$ and dual hyperplanes $A_1, \ldots, A_n$ with $a_j A_k = \delta_{jk}$. We turn this into an (n-1)-d elliptic space by picking the

[2]The example of a 4-d space is used because that is where we shall end up - in order to model 3-d Euclidean geometry on a 3-d quadric hypersurface in the 4-d space.

[3]Another solution would be to allow any sort of polarity, and not make the assumption $a_1 \ldots a_n = 1$. In fact, numbers and pseudonumbers would be treated as different kinds of things with pseudonumbers being regarded as elements of the algebra alongside points, lines and so forth. In this scheme, whenever one encountered a product of a pseudonumber and another element, say, for simplicity $a_1 \ldots a_n . A_1 \ldots A_r$, it could be worked out by the fundamental formulae of section 1.14. In this particular case, it can be immediately evaluated by the rule of the middle factor (1.33) as,

$$a_1 \ldots a_n . A_1 \ldots A_r = a_{r+1} \ldots a_n \ .$$

This approach would be closer to that of [2, 3, 5, 6] where the Grassmann product of n vectors is called a *pseudoscalar* and it is not regarded as a number. However, this scheme is avoided in the current text because the writer is more influenced by how things transform between observers as a guide to their nature. Pseudonumbers transform according to (1.9). Now $\det(f) = 1$ for transformations between equivalent (inertial) observers, so pseudonumbers transform like ordinary numbers and so the mathematics seems to be saying that pseudonumbers and ordinary numbers are the same kind of things.

absolute polarity as the natural elliptic polarity $IA_j = a_j$ for $j = 1, \dots, n$. The point at infinity IL and the origin IN are taken as points on the absolute quadric hypersurface $pIp = 0$ so that have $IL.L = IN.N = 0$. If we need to specify these points they are usually taken as $IL = a_{n-1} - ia_n$ and $IN = a_{n-1} + ia_n$ where $i = \sqrt{-1}$. Then, $L = A_{n-1} - iA_n$ and $N = A_{n-1} + iA_n$. In section 3.10.1 we generated an arbitrary point p on the quadric hypersurface by the action of a parabolic congruence with invariant line ILM where the point IM is in the subspace LN. If we take the generator as,

$$IM = \sum_{j=1}^{n-2} x_j a_j \tag{7.4}$$

$$M = (-1)^{n-1} I^2 M = \sum_{j=1}^{n-2} x_j A_j \tag{7.5}$$

where the x_j are weights, then IM is in LN since $LN = 2iA_{n-1}A_n$. Now, equation (3.50) gives the point on the quadric as,

$$p = e^{ILM.I} IN = IN + (-1)^{n-1}(IN.L)\left(IM - \frac{(-1)^{n-1}}{2}(IM.M)IL\right) \tag{7.6}$$

and,

$$IM.M = \sum_{j=1}^{n-2} x_j^2 \tag{7.7}$$

is the Euclidean squared distance of the point p from the origin IN as given by (3.51). It is clear that the generator IM given by (7.4) is the position vector of a point in the classical treatment of vector analysis [1] and the reference points a_j for $j = 1, \dots, n-2$ in the subspace LN are the unit vectors. This is similar to the situation in section 5.2.3 where a vector was found to be a point in the hyperplane at infinity of affine space. Here the unit direction vectors are points in the subspace LN. Finally, if set $n = 5$ then we have a model for 3-d Euclidean space on the quadric hypersurface of 4-d elliptic space.

7.3 Computational 4-d elliptic geometry

The theoretical material in this chapter will be illustrated by computations using the Yorick packages `Whitehead4d.i`, `Euclid.i` and `utilities.i` which can be down-loaded from the url [21] . The package `Whitehead4d.i` handles 4-d elliptic geometry with complex weights. The package `Euclid.i` handles 3-d computational Euclidean geometry by modelling it on the absolute quadric of 4-d elliptic space as explained in section 7.2. The package `utilities.i` contains useful functions that are regarded as applications of the functions in the other two packages. Consequently, `Whitehead4d.i` is the fundamental package. It implements the theory of chapter 1 and chapter 3 in 4-d. The current section explains the internal working of `Whitehead4d.i`. It is a straightforward generalisation of the toy 2-d package `Whitehead2d.i` which has already been described in chapter 2.

7.3.1 Points, lines, planes and hyperplanes

In 4-d we have 5 reference points $a_1, \ldots, a_5$ and a point x is,

$$x = x_1 a_1 + x_2 a_2 + x_3 a_3 + x_4 a_4 + x_5 a_5$$

where the x_j are weights. A point is represented by the 1-d array `x(j)`. The following example from a Yorick session illustrates how a point is defined and how it is stored as a 1-d array. Notice that complex weights are allowed.

```
> #include "Whitehead4d.i"
> x=5.0*a1+6.0*a2+7.0*a3-4.7*(a4-1.0i*a5);
> print_element, x
+5.000a1+6.000a2+7.000a3-4.700a4+4.700ia5
> x
[5+0i,6+0i,7+0i,-4.7+0i,0+4.7i]
>
```

A line l is,

$$\begin{aligned} l = l_{12} a_1 a_2 + l_{13} a_1 a_3 + l_{14} a_1 a_4 + l_{15} a_1 a_5 + l_{23} a_2 a_3 \\ + l_{24} a_2 a_4 + l_{25} a_2 a_5 + l_{34} a_3 a_4 + l_{35} a_3 a_5 + l_{45} a_4 a_5 \end{aligned}$$

where the l_{jk} are weights. A line is represented by an upper triangle 2-d array `l(j,k)`. We anticipate the `wap` function that implements the antisymmetric (Whitehead algebra) product in order to be able to illustrate the definition of a line.

```
> l=4.0*wap(a1,a2)+7.0*wap(a3,1.0i*a5);
> print_element, l;
+4.000a1a2+7.000ia3a5
```

A plane P is,

$$\begin{aligned} P = P_{123} a_1 a_2 a_3 + P_{124} a_1 a_2 a_4 + P_{125} a_1 a_2 a_5 + P_{134} a_1 a_3 a_4 + P_{135} a_1 a_3 a_5 \\ + P_{145} a_1 a_4 a_5 + P_{234} a_2 a_3 a_4 + P_{235} a_2 a_3 a_5 + P_{245} a_2 a_4 a_5 + P_{345} a_3 a_4 a_5 \end{aligned}$$

where the P_{jkl} are weights. A plane is represented by a 3-d array `P(j,k,l)` with non-zero components addressed by increasing indices.

```
> P=3.0*wap(a1,wap(a2,a3))-4.0*wap(wap(a5,a4),a3);
> print_element, P;
+3.000a1a2a3+4.000a3a4a5
```

A hyperplane H is,

$$H = H_{1234}a_1a_2a_3a_4 + H_{1235}a_1a_2a_3a_5 + H_{1245}a_1a_2a_4a_5 \\ +H_{1345}a_1a_3a_4a_5 + H_{2345}a_2a_3a_4a_5$$

where the H_{jklm} are weights. A hyperplane is represented by a 4-d array `H(j,k,l,m)` with only 5 non-zero components which are addressed by increasing indices.

7.3.2 Duality and the elliptic polarity

Section 3.4 explained that each element of the algebra can also be represented dually. For example, a plane is normally represented as a sum of products of three points as a 3-d array. However, the same plane can also be represented dually as the intersection of pairs of hyperplanes as a 2-d array. The function `I` is used to convert an element's array to its dual array representation. In the following Yorick session the array elements of a plane are examined in its point and dual hyperplane representations.

```
> print_element, P;
+3.000a1a2a3+4.000a3a4a5
> P(1,2,3);
3+0i
> I(P)(4,5);
3+0i
```

Furthermore, as explained in section 3.4, the function `I` can also be regarded as the furnishing the natural (elliptic) polarity. So, in our example, if P is a plane, IP is a line.

```
> print_element, I(P);
+4.000a1a2+3.000a4a5
```

Let's check the involutary nature of the polarity as given by equation (3.7) or (3.10).

```
> print_element, I(I(P));
+3.000a1a2a3+4.000a3a4a5
```

Let's check the outermorphism property given by equation (3.4).

```
> print_element, I(wap(a1,wap(a2,a3)));
+1.000a4a5
> print_element, wap(I(a1),wap(I(a2),I(a3)));
+1.000a4a5
```

Finally, on a general note, the packages are self-documenting in the sense that Yorick's `help` function always provides on-line documentation. So, to get help on the polarity function just enter `help, I`.

7.3.3 The antisymmetric product

Section 2.4 has already shown how to implement the antisymmetric product (`wap` function) in the toy example of 2-d projective space. In this section we give the formulae to implement the `wap` function for 4-d projective space. The method is the same as in section 2.4 but now the Einstein summation convention means we sum from 1 to 5 over any pairs of repeated indices.

In order to calculate XY where X and Y are points we write X in point-form as $X = X_i a_i$ and Y in hyperplane form as $Y = Y_{jklm} A_j A_k A_l A_m$. Using (1.41),

$$
\begin{aligned}
XY &= X_i Y_{jklm} a_i . A_j A_k A_l A_m \\
&= X_i Y_{jklm} (\delta_{im} A_j A_k A_l - \delta_{il} A_j A_k A_m + \delta_{ik} A_j A_l A_m - \delta_{ij} A_k A_l A_m) \\
&= X_i Y_{jkli} A_j A_k A_l - X_i Y_{jkim} A_j A_k A_m + X_i Y_{jilm} A_j A_l A_m - X_i Y_{iklm} A_k A_l A_m
\end{aligned}
$$

The formula for multiplying *two points* is obtained by changing the dummy indices on the hyperplane factors so that they can be taken out as a common factor.

$$XY = (X_i Y_{jkli} - X_i Y_{jkil} + X_i Y_{jikl} - X_i Y_{ijkl}) A_j A_k A_l \tag{7.8}$$

The 3-d array of coefficients gives the result in hyperplane-form which can be converted to point-form by taking the dual representation.

The calculations of all the other cases follow the same pattern, so only the results are just stated without the intervening manipulation of indices. In all cases we write the first factor X in point-form and the second factor Y in hyperplane form.

The formula for multiplying a *point and a line* is,

$$XY = (X_i Y_{jki} - X_i Y_{jik} + X_i Y_{ijk}) A_j A_k \tag{7.9}$$

and the 2-d array of coefficients gives the result in hyperplane-form.

The formula for multiplying a *point and a plane* is,

$$XY = (X_i Y_{ji} - X_i Y_{ij}) A_j \tag{7.10}$$

and the 1-d array of coefficients gives the result in hyperplane-form.

The formula for multiplying a *point and a hyperplane* is,

$$XY = X_i Y_i \tag{7.11}$$

and the result is a number.

The formula[4] for multiplying a *line and a line* is,

$$XY = (X_{ij} Y_{kij} - X_{ij} Y_{ikj} + X_{ij} Y_{ijk}) A_k \tag{7.12}$$

and the 1-d array of coefficients gives the result in hyperplane-form.

The formula for multiplying a *line and a plane* is,

$$XY = X_{ij} Y_{ij} \tag{7.13}$$

and the result is a number.

There is no need to go any further because all the other cases can be found by duality. For example, the next case would logically be a *line and a hyperplane*. Let l be the line and H the hyperplane. The product lH is the point at the intersection of the line and the hyperplane. Hence, by (3.7) and (1.13) we can arrange things so that,

$$lH = (-1)^{5-1} I^2 (lH) = I(Il.IH) = (-1)^3 I(IH.Il) = -I(IH.Il)$$

and since IH is a point and Il is a plane, the product on the right can be evaluated using (7.10).

As an example of the use of the `wap` function, we verify Müller's third theorem (6.5) for the situation in which X, Y, and Z are three planes. Since X is a plane it is of point-grade $q = 3$. Y and Z are also planes so they are both of hyperplane grade $r = s = 5 - 3 = 2$. These grades satisfy the conditions attached to Müller's third theorem and so it becomes the following theorem about three planes in 4-d projective space.

$$X.YZ = XZ.Y + XY.Z \tag{7.14}$$

[4] In deriving this formula, terms like $X_{ij} Y_{kji}$ turn up, but can be discarded because the coefficients X_{ij} and Y_{kji} are only non-zero for $i < j$ and $k < j < i$. In this case this is impossible so the term is zero. Three terms can be discarded in this way in order to end up with the result (7.12)

In order to check this theorem numerically we need to create the three planes X, Y and Z. We use the fact that the package `Whitehead4d.i` declares an array of reference points `a` which can be used to easily get reference points a_j in loops.

```
> print_element, a(1,)+a(2,)+a(3,)+a(4,)+a(5,);
+1.000a1+1.000a2+1.000a3+1.000a4+1.000a5
```

Now we make an array `v` of 9 random points by combining the reference points with a 9 by 5 matrix of random coefficients. The second of the random points is printed out to show what has been done.

```
> v=random([2,9,5])(,+)*a(+,);
> print_element, v(2,);
+0.258a1+0.708a2+0.519a3+0.290a4+0.116a5
```

Next, we make our three random planes out of the random points and form both sides of theorem (7.14). Each side evaluates to the same hyperplane thus verifying the theorem for these random planes.

```
> X=10.0*wap(v(1,),wap(v(2,),v(3,)));
> Y=10.0*wap(v(4,),wap(v(5,),v(6,)));
> Z=10.0*wap(v(7,),wap(v(8,),v(9,)));
> print_element, wap(X,wap(Y,Z));
-3.457a1a2a3a4+7.512a1a2a3a5-10.496a1a2a4a5-6.686a1a3a4a5+12.764a2a3a4a5
> print_element, wap(wap(X,Z),Y)+wap(wap(X,Y),Z);
-3.457a1a2a3a4+7.512a1a2a3a5-10.496a1a2a4a5-6.686a1a3a4a5+12.764a2a3a4a5
```

7.3.4 Elliptic congruence

The package `Whitehead4d` has a function `con` which implements the canonical formula (3.26) for a simple congruence in the special case of 4-d elliptic space. As an example, let's use this function to numerically calculate some events in de Sitter space.

Section 5.4 studied de Sitter space as the ultra-infinite points of 4-d hyperbolic space. Although the package `Whitehead4d.i` is for 4-d elliptic space, we can easily do calculations in hyperbolic space, as explained in section 7.2, by working with an imaginary reference point. All of the calculations in section 5.4 make sense if we work in 4-d elliptic space with reference point ia_5. Also, all of the diagrams in 5.4 are the same if we re-label the reference point a_5 inside the quadric hypersurface to ia_5.

Using table 5.1 we see that the invariant line for time translation is $a_1.ia_5$. So, points $p(t)$ on the world-line of an observer can be obtained by time-translating the origin event a_1 as $p(t) = f_t a_1$ where f_t is the congruence. In the following Yorick session, we calculate the observer's position when his proper time reads $t = 1$.

```
> a1ia5=wap(a1,1.0i*a5);
> ft_a1=con(1.0,a1ia5,a1);
> print_element, ft_a1;
+1.543a1+1.175ia5
```

From table 5.1 we also see that the invariant line for a spatial translation along the observer's x-axis is a_1a_2. Suppose we consider two observer's Alice and Bob. Suppose Alice moves away along Bob's x-axis by a distance $\pi/4$ in cosmological units[5]. Alice's world-line a_1ia_5 now appears to Bob as the line $f_x(a_1ia_5)$ where f_x is the congruence.

```
> fx_a1ia5=con(1.0i*pi/4.0,wap(a1,a2),a1ia5);
> print_element, fx_a1ia5;
+0.707ia1a5+0.707ia2a5
```

In other words, Alice's world-line has just been rotated through $\pi/4$ about the point ia_5 in the direction of point a_2 in figure 5.7. Figure 5.7 also shows that the tangent hyperplane $I(a_1 + ia_5)$ is Bob's future event horizon. Once an object goes across Bob's horizon it is lost forever from Bob's universe. Let's determine the point on Alice's world-line at which she leaves Bob's universe.

```
> p=wap(fx_a1ia5,I(a1+ia5));
> print_element, p;
+0.707a1+0.707a2+0.707ia5
```

This point p is on Alice's world-line as seen by Bob. It is also given by $p = f_x f_t a_1$ where t is Alice's proper time. Of course, we could solve this equation analytically[6] to get Alice's proper time t at which she disappears over Bob's horizon. However, the purpose of this section is to show how the package `Whitehead4d.i` can be used. By re-arranging this equation, $f_{-t} f_{-x} p = a1$ and trying a few values of t, we get the solution $t = 0.881$.

```
> print_element, con(0.881,-a1ia5,con(1.0i*pi/4.0,-wap(a1,a2),p));
+0.707a1+0.000ia5
```

[5]The length of a space-like line in de Sitter space is π units.

[6]The analytic solution is $t = -\ln(\tan(\theta/2))$ where θ is the distance of the translation along the x-axis.

7.3.5 Parabolic congruence

The package `Whitehead4d` has the function `pcon` which implements the canonical formula (3.47) for a parabolic congruence in the special case of 4-d elliptic space with the parameter $\tau = 1$. In other words, it evaluates the action of $\exp(ILM.I)$ on any element in the parabolic case $ILM.LM = 0$.

This function can be used to generate points on the absolute quadric according to equation (7.6). In the following example, we define the point at infinity IL, the origin IN and a generator IM in the subspace LN. We produce a point p on the quadric by acting on the origin with the parabolic congruence. Finally, we check that p is on the quadric so that $pIp = 0$.

```
> IL=a4-1.0i*a5;
> IN=a4+1.0i*a5;
> IM=a1+2.0*a2+3.0*a3;
> p=pcon(IL,IM,IN);
> print_element, wap(p,I(p));
+0.000a1a2a3a4a5
```

In the next example we generate another point q on the quadric and then check the formula (3.49) for the distance between two points on the quadric.

```
> q=pcon(IL,2.0*a1+4.0*a2+6.0*a3,IN);
> L=I(IL);
> print_element, -2.0*wap(p,I(q))/(wap(p,L)*wap(q,L));
+14.000a1a2a3a4a5
```

The result for the squared distance is 14 and this is indeed the squared distance that would be obtained by treating the generators (the IM points that live in subspace LN) as position vectors in classical vector analysis.

7.3.6 Other functions

The package `Whitehead4d.i` also contains the vectorized functions `I2`, `wap2` and `con2`. These are versions of the standard fuctions `I`, `wap` and `con` which apply to arrays of elements of the algebra. They work faster than calling the corresponding standard function in a loop and are implemented for applications to computer vision where speed of execution is important. They are self-documenting, for example, typing `help, wap2` produces comprehensive documentation on the function `wap2`.

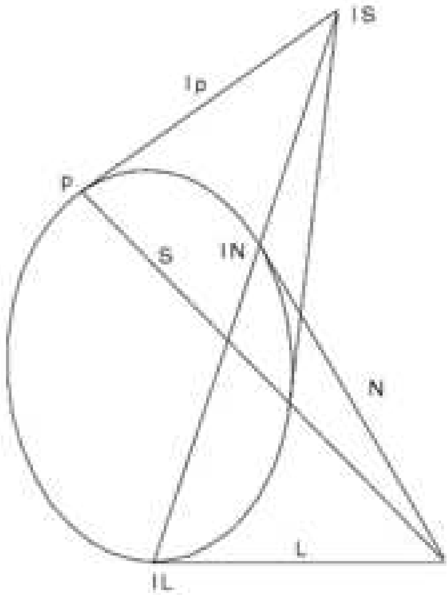

Figure 7.1: Hyperplane intersecting the quadric

7.4 Hyperspheres in Euclidean space

We now resume the study of the model of Euclidean geometry on the elliptic quadric by studying the properties of hyperspheres on their own. This will provide the motivation to regard a hyperplane in Euclidean space as a hypersphere of infinite radius. We then interrupt the study of hyperspheres to explore, in section 7.5, the aforementioned concept of a hyperplane in Euclidean space. Then, having become familiar with handling hyperplanes, we will be able to take up the study of hyperspheres again in section 7.6 and obtain closed formulae for the centre and radius of the subspace formed by the intersection of r hyperspheres.

7.4.1 The single hypersphere

Figure 7.1 shows the quadric as an ellipse. Of course this is only schematic because the elliptic quadric does not consist of real points, and anyway, the figure is only a 2-d slice through n-d elliptic space. The point at infinity IL is shown on the quadric. IL is on L so $IL.L = 0$. S is an arbitrary hyperplane. Let us determine the set of points p on the quadric which also lie in S. In other words, we want to solve $pS = 0$. Since the points on the quadric are the points of (n-2)-d Euclidean space in our model, the intersection of hyperplane

S with the quadric will be a set of points which are a (n-3)-d subspace of Euclidean space. When n=5, we have a 2-d subspace of 3-d Euclidean space. In order to find out what this subspace represents, it is convenient to choose the origin point IN as the intersection of the line $IL.IS$ with the quadric as shown[7] in figure 7.1. The point IN is, in fact, the reflection of point IL in the hyperplane S. Using the formula (3.39) for a reflection, we find,

$$IN = f(IL) = \frac{2(IL.S)IS}{IS.S} - IL \tag{7.15}$$

where f is the operator of reflection in S. The hyperplane N is,

$$N = \frac{2(IL.S)S}{IS.S} - L \tag{7.16}$$

and it is easy to check that $IN.N = 0$ so that IN does indeed lie on the quadric. Now let's find the distance between the origin IN and a point p which satisfies $pS = 0$. The distance between two points on the quadric is given by equation (3.49). Thus, the squared distance between IN and p is,

$$\tau^2 = \frac{-2(pIIN)}{(pL)(IN.L)} = \frac{-(-1)^{n-1}2(pN)}{(pL)(IN.L)} \ .$$

Upon substituting for N from (7.16) and using $pS = 0$, we find,

$$\tau^2 = \frac{(-1)^{n-1}(IS.S)}{(IS.L)^2} = \frac{(SIS)}{(IS.L)^2} \ . \tag{7.17}$$

The distance is independent of the point p. So, all the points p on the quadric which also lie on hyperplane S are at the same distance from the origin point IN. In other words, the points all belong to a hypersphere embedded in the Euclidean space on the quadric. In the case n=5, the points are on the surface of a sphere in 3-d Euclidean space. The origin point IN is the centre of the hypersphere. The radius τ of the hypersphere is given by (7.17).

Suppose that we write $S = p_1 \dots p_{n-1}$ where each of the points lies on the quadric. Since, by definition, each point p_j also lies on S, then they must also be on the hypersphere represented by S. In other words, the hypersphere that goes through these $n-1$ points is just $S = p_1 \dots p_{n-1}$. Let's check these findings using Yorick.

We start by loading the package `utilities.i`. By loading this package, the packages `Euclid.i` and `Whitehead4d.i` are automatically loaded so everything is ready for 3-d Euclidean computational geometry on the quadric.

[7] It is easy to explain why the point IS in figure 7.1 lies on the tangent hyperplane Ip. Since $pS = 0$, then taking the polarity, $Ip.IS = 0$ so IS lies on Ip.

Furthermore, the package Euclid.i defines the point at infinity IL and the origin IN according to the default elements suggested in section 7.2. The elements L, N, LN, $ILIN$, $IN.L$ and $i = \sqrt{-1}$ are also defined in Euclid.i.

```
> #include "utilities.i"
> print_element, IL;
+1.000a4-1.000ia5
> print_element, IN;
+1.000a4+1.000ia5
```

Now we define four points that live on the surface of a sphere of radius 5 units with centre a distance of 5 units along the positive z axis. The Cartesian coordinates of these four points are (5,0,5),(0,5,5),(0,0,10) and (0,0,0). For us, as already explained at the end of section 7.3.5, the (x, y, z) coordinates of a point, or a position vector of classical vector analysis, really mean a generating point $IM = xa_1 + ya_2 + za_3$ in the subspace LN. We think of our points starting out as generators in the subspace LN, but we must then generate the actual points on the quadric. We can then compute the sphere as $S = p_1p_2p_3p_4$.

```
> p1=pcon(IL,5*a1+5*a3,IN);
> p2=pcon(IL,5*a2+5*a3,IN);
> p3=pcon(IL,10*a3,IN);
> p4=IN;
> S=wap(p1,wap(p2,wap(p3,p4)));
> print_element, S;
+2000.000a1a2a3a4+2000.000ia1a2a3a5-20000.000ia1a2a4a5
```

We can check that the element S does represent the correct sphere because the point p with coordinates $5(3^{-1/2}, 3^{-1/2}, 3^{-1/2}+1)$ should lie on the sphere. We verify that $pS = 0$.

```
> p=pcon(IL,5*(a1+a2+a3)/sqrt(3)+5*a3,IN);
> print_element, wap(p,S);
+0.000a1a2a3a4a5
```

We can also check that formula (7.17) gives the correct radius for the sphere.

```
> sqrt(wap(I(S),S)/(wap(I(S),L)^2));
5-0i
```

Using equation (7.16), it is easy to see that, up to an unimportant weight, a hypersphere can be written as

$$S = L + N \tag{7.18}$$

where IN is the point at the centre of the hypersphere. This equation disguises the fact that the radius of the hypersphere depends on the weight absorbed into N. If we calculate the squared radius of S from (7.18) using the formula (7.17), the result is $\tau^2 = 2(-1)^{n-1}/(IN.L)$. In other words, if we increase the intrinsic weight of N, the radius decreases because the sphere $S = L + N$ is pulled towards the element N. Ultimately, a sphere given by N is just a point at the origin. This can also be appreciated by noting that S is a linear combination of L and N as drawn in figure 7.1.

Using the default values of L and N declared in package `Euclid.i`, we find that the sphere of unit radius centred on the origin is $L + N$.

```
> s=L+N;
> sqrt(wap(I(s),s)/(wap(I(s),L)^2));
1+0i
```

7.4.2 Dilatations and a neat way to define spheres

Instead of defining a sphere by a product of four points that lie on it, we can make a sphere of unit radius centred on the origin and then apply a dilatation to change the radius, and a parabolic congruence to move it into position. A dilatation is an expansion or contraction of space[8]. A dilatation in the model of Euclidean geometry on the quadric is just a congruence with invariant line $IL.IN$ in figure 7.1. We could just evaluate the canonical formula (3.26) using this invariant line with IL and IN both points on the quadric so that $IL.L = IN.N = 0$. However, the properties of such a congruence can be easily obtained using formula (3.49) for the distance between two points on the quadric.

Let f_τ be the congruence with invariant line ILN and congruence parameter τ. Let p_1 and p_2 be two points on the quadric so that $p_1 I p_1 = p_2 I p_2 = 0$. From equation (3.44) we know that the new points $f_\tau p_1$ and $f_\tau p_2$ will still be on the quadric. From (3.49) the distance between the transformed points is,

$$(-1)^{n-1}\sqrt{\frac{-2(f_\tau p_1.I f_\tau p_2)}{(f_\tau p_1 L)(f_\tau p_2 L)}} = (-1)^{n-1}\sqrt{\frac{-2(p_1 I p_2)}{(f_\tau p_1.L)(f_\tau p_2.L)}}$$

where $f_\tau p_1 I f_\tau p_2 = p_1 I p_2$ because I commutes with f_τ. In the denominator,

$$f_\tau p_1.L = p_1.f_{-\tau}L = p_1.\exp\left((-1)^{n-1}\tau\right)L = \exp\left((-1)^{n-1}\tau\right)p_1 L$$

[8]The absence of dilatations in our world was used in chapter 5 to argue that space-time cannot be Minkowskian and that we should regard de Sitter space as the natural space-time arena for special relativity.

where we have used equation (3.2), the eigenvalue equation (3.45) and the fact that I commutes with the congruence. The distance between the transformed points is therefore,

$$(-1)^{n-1}\sqrt{\frac{-2(f_\tau p_1.If_\tau p_2)}{(f_\tau p_1L)(f_\tau p_2L)}} = e^{(-1)^n\tau}(-1)^{n-1}\sqrt{\frac{-2(p_1Ip_2)}{(p_1L)(p_2L)}}$$

which shows that the effect of the congruence is to change the distance between any two points by a factor of $\exp((-1)^n\tau)$. In other words, the congruence is a dilatation[9].

In the following example, we use a dilatation to expand the unit sphere $L+N$ to a radius of 5 units. We check that the radius of the expanded sphere S_1 is 5 units. This technique can be applied to any element or set of elements (using `con2`).

```
> S1=con(-log(5.0),ILIN,L+N);
> sqrt(wap(I(S1),S1)/(wap(I(S1),L)^2));
5-0i
```

In the next example, we create a sphere S_2 by expanding the unit sphere to a radius of 5 units and translating it 5 units along the z-axis using a parabolic congruence. In this way, we can easily create arbitrary spheres at any point. We check the radius of the new sphere and also check that it has been moved along the z-axis because the origin point IN now lies on S_2.

```
> S2=pcon(IL,5.0*a3,con(-log(5.0),ILIN,L+N));
> sqrt(wap(I(S2),S2)/(wap(I(S2),L)^2));
5-0i
> print_element, wap(IN,S2);
+0.000a1a2a3a4a5
```

7.4.3 Rotation of a hypersphere into itself

It is interesting to find the form of the congruence which rotates a hypersphere into itself. The congruence must leave the centre IN of the hypersphere invariant. It must also leave the point at infinity IL invariant. Let the congruence be f_τ. Since the congruence commutes with the polarity, then $f_\tau IL = IL$ implies $f_\tau L = L$ and similarly $f_\tau N = N$. From equation (7.18), this implies that $f_\tau S = S$ so that the congruence maps the hypersphere into

[9]The invariant line of the dilatation sticks out of the quadric hypersurface in figure 7.1. Coupled with the fact that we do not experience dilatations, this suggests that there are no extra dimensions in our world.

itself. In other words, such a congruence is a rotation of the hypersphere about its centre. The line ILN is a part of the subspace in which the congruence acts like the identity[10], so the invariant line of the congruence must lie in the subspace LN. We have already met the subspace LN. It is the subspace containing the generators [11] of points on the quadric. Therefore, a rotation is a simple congruence with invariant line in the generating subspace LN.

In the following example, we make an invariant line for a rotation and check that it lies in LN. We then make a unit sphere centred on the origin and rotate it by $\pi/2$ and check that it has been rotated into itself.

```
> line=wap(a1,a2);
> print_element, wap(line,LN);
+0.000a1a2a3a4a5
> print_element, L+N;
-2.000a1a2a3a5
> print_element, con(1.0i*pi/2.0,line,L+N);
-2.000a1a2a3a5
```

Notice that in the previous example, the congruence parameter for the rotation through $\pi/2$ was $\tau = i\pi/2$. The reason for this is that the congruence function `con` implements equation (3.26),

$$f_\tau = \exp\left(\frac{\tau ILM.I}{\sqrt{-ILM.LM}}\right) .$$

For a congruence which is a rotation we always have $ILM.LM > 0$. In the above example this factor is,

```
> print_element, wap(line,I(line));
+1.000a1a2a3a4a5
```

and so the square root in the denominator evaluates to $i = \sqrt{-1}$. Therefore, the congruence parameter τ needs to be imaginary in order to keep the argument of the exponential real. In order to see why this is important, we see that $f_{i\pi/2}a_1 = a_2$,

```
> print_element, con(i*pi/2.0,line,a1);
+0.000a1+1.000a2
```

whilst real τ results in complex weights on points in the generating subspace LN.

[10] See section 3.5.
[11] See equation (7.4).

```
> print_element, con(pi/2.0,line,a1);
+2.509a1-2.301ia2
```

This must be avoided in applications to Euclidean geometry[12] because the Euclidean distance (7.7) relies on the assumption that the weights of points in the generating subspace LN are real.

7.5 Linear subspaces of Euclidean space

We have seen that a hypersphere in (n-2)-d Euclidean space can be written as a product $S = p_1 \ldots p_{n-1}$ of points which lie on the hypersphere. If one of the points is taken as the point at infinity IL, the hypersphere S satisfies $IL.S = 0$ and the radius of the hypersphere is infinite according to equation (7.17). In other words, the hypersphere has become a hyperplane. So, a (n-3)-d hyperplane H in the (n-2)-d Euclidean space on the quadric is represented by an element[13] of the form $H = IL.p_1 \ldots p_{n-2}$. In general, linear subspaces (points, lines and planes) in Euclidean space are represented by elements which contain the point at infinity as a factor. This section explores the properties of linear subspaces defined in the aforementioned way.

7.5.1 Planes, lines and points

The following example defines C as the x-y plane in 3-d Euclidean space. It does this by defining two vertices v_1 and v_2 which are on the plane. v_1 is the point on the quadric generated by the Cartesian coordinates (1,0,0) and v_2 is the point corresponding to coordinates (0,1,0). The plane is formed by the product $C = IL.IN.v_1.v_2$. The plane goes through the origin because IN is in the product.

```
> v1=pcon(IL,a1,IN);
> v2=pcon(IL,a2,IN);
```

[12] The use of points with imaginary weights in the generating subspace LN would be the way to implement 2+1-d Minkowski space without any modification of the existing packages.

[13] To clarify things, H is a hyperplane in the (n-1)-d elliptic space. However, the set of points which lie in the intersection of H with the quadric are the points which belong to a hyperplane in the model of (n-2)-Euclidean space on the quadric. This may be a little confusing at first because H is a hyperplane in the elliptic space, and it also represents a hyperplane of the Euclidean space by virtue of its intersection with quadric. The situation for a hypersphere S is similar, but is not so confusing because S is a *hyperplane* element of the elliptic space, but the set of points which lie on the intersection of S with the quadric constitute the points of a *hypersphere* in the model of Euclidean space on the quadric.

```
> C=wap(IL,wap(IN,wap(v1,v2)));
> print_element, C;
+8.000ia1a2a4a5
```

We can check that the C really does represent the x-y plane through the origin by checking that it contains the correct points. For example, the point v_3 with coordinates (0,0,1) does not lie on the plane. This is evident because the number v_3C is non-zero.

```
> v3=pcon(IL,a3,IN);
> print_element, wap(v3,C);
+16.000ia1a2a3a4a5
```

The point with coordinates (10,20,0) does lie on the plane.

```
> print_element, wap(pcon(IL,10.0*a1+20.0*a2,IN),C);
+0.000a1a2a3a4a5
```

The next example makes the x-z plane through the origin. We use B for this plane. Instead of defining this plane as a product of points, we simply rotate the x-y plane through $\pi/2$ about the x-axis. We already have the vertex v_1 on the axis. Consequently, the points IL, IN and v_1 will be invariant points of the rotation. The subspace in which the rotation acts like the identity is therefore $IL.IN.v_1$ and so the invariant line of the rotation is $I(IL.IN.v_1) = LNIv_1$. Having obtained the x-z plane, we check that it is sensible by showing that the point (3,0,4) is on the new plane.

```
> B=con(i*pi/2.0,I(wap(IL,wap(IN,v1))),C);
> print_element, wap(pcon(IL,3.0*a1+4.0*a3,IN),B);
+0.000a1a2a3a4a5
```

Incidentally, the invariant line of the rotation was a_2a_3 modulo an unimportant weight.

```
> print_element, I(wap(IL,wap(IN,v1)));
+4.000ia2a3
```

This is in agreement with section 7.4.3 where general considerations showed that the invariant line for a rotation about the origin had to lie in the generating subspace LN.

It is clear that the intersection of the planes B and C is the line along the x-axis. We can compute this intersecting line as BC. The reason is that the set of points in Euclidean space represented by product BC is the set of points which lie on the intersection of element BC with the quadric. In the following example, we compute BC and check that it represents the x-axis because v_1 is on the x-axis, whilst v_2 is on the y-axis.

```
> BC=wap(B,C);
> print_element, wap(v1,BC);
0
> print_element, wap(v2,BC);
+128.000a1a2a4a5
```

The line $BC = IL.IN.v_1$ modulo an unimportant weight.

```
> print_element, BC;
-64.000a1a4a5
> print_element, wap(wap(IL,IN),v1);
+4.000ia1a4a5
```

This shows that a straight line can be written in the form ILp_1p_2 where p_1 and p_2 are points on the line in the quadric. In general, we can represent an (r-1)-d linear subspace of the (n-2)-d Euclidean space on the quadric by an element of the form $X = IL.p_1 \ldots p_r$ and we can compute intersections of linear subspaces by just multiplying such elements with the antisymmetric product. The result of multiplying two such elements will always be a linear subspace because the product will always contain the point at infinity IL as a factor. This is obvious physically and is easily shown formally by an application of Müller's theorem (6.3).

Continuing with the example, let's set up the y-z plane and translate it so that it sits at the point x=1. In the example, we do this by rotating the x-y plane through $\pi/2$ about the y-axis and then moving the resulting plane into position at x=1. The final result is the plane denoted A. We check that A makes sense by verifying that it contains the point (1,2,-4). Notice that $ILIN$ is pre-defined in package `Euclid.i`.

```
> A=pcon(IL,a1,con(i*pi/2.0,I(wap(ILIN,v2)),C));
> print_element, wap(pcon(IL,a1+2.0*a2-4.0*a3,IN),A);
+0.000a1a2a3a4a5
```

Now let's obtain the point at the intersection of the x-axis with the plane A. The x-axis has already been obtained as line BC. The intersection of the axis with the plane is $A.BC$. We know that it must represent the point v_1 which was defined to be the point (1,0,0). The example shows that printing out the result $A.BC$ is pretty unintelligible. This is because $A.BC = IL.v_1$ modulo an unimportant weight as the example shows.

```
> print_element, wap(A,BC);
-512.000a1a4+512.000ia1a5+512.000ia4a5
> print_element, wap(IL,v1);
-2.000a1a4+2.000ia1a5+2.000ia4a5
```

The last example shows that we need a way of taking a point in the form $IL.p$ and extracting the Cartesian coordinates. We can calculate with our elements representing Euclidean subspaces, but at some point we will need to send the coordinates of a point to some other device. Fortunately, the package `utilities.i` contains several functions for changing back and forth between various representations of a point. These are described in section 7.5.2.

7.5.2 Some utility functions

In this section we describe the functions in the package `utilities.i` for changing back and forth between different representations of a point. In each case we give an example of the use of the function, followed by the theoretical background.

In the first example we generate the point p on the quadric that has coordinates (4,-3,9). We check that p is on the quadric because it satisfies the quadric equation $pIp = 0$. However, this is why p looks messy when printed out. The function `q_from_p` takes a point p on the quadric and returns the generating point in the subspace LN. The generating point is the equivalent of a position vector in classical vector analysis.

```
> #include "utilities.i"
> p=pcon(IL,4.0*a1-3.0*a2+9.0*a3,IN);
> print_element, wap(p,I(p));
+0.000a1a2a3a4a5
> print_element, p;
+8.000a1-6.000a2+18.000a3-105.000a4+107.000ia5
> print_element, q_from_p(p);
+4.000a1-3.000a2+9.000a3
```

Equation (7.6) for generating a point p on the quadric uses IM as the generator. However, the function `q_from_p` denotes the generator by $q = IM$. With this replacement, equation (7.6) becomes,

$$p = e^{ILq.I}IN = IN + (-1)^{n-1}(IN.L)\left(q - \frac{(qIq)}{2}IL\right) . \tag{7.19}$$

In practice, the point may contain an arbitrary weight, so we have to solve,

$$p = \lambda\left(IN + (-1)^{n-1}(IN.L)\left(q - \frac{(qIq)}{2}IL\right)\right)$$

for the generator q. In order to extract q we multiply by $ILIN$.

$$IL.IN.p = \lambda(-1)^{n-1}(IN.L)IL.IN.q$$

Multiplication by LN, and using $IL.L = IN.N = 0$ and $qL = qN = 0$,

$$(IL.IN.p).LN = \lambda(-1)^{n-1}(IN.L)(IL.IN.q).LN = \lambda(-1)^{n}(IN.L)^3 q \ .$$

The weight λ can be found by calculating $pL = \lambda IN.L$ so that the result for the generator is,

$$q = \frac{(-1)^n (IL.IN.p).LN}{(pL)(IN.L)^2} \ . \tag{7.20}$$

Suppose that we have a linear subspace that goes through the origin IN. It must be of the form $X = IL.IN.p_1 \dots p_r$ where the p_i are points on the quadric. Each point p_i can be generated from a point q_i in the subspace LN according to equation (7.19). Substituting for the p_i in this way, and using antisymmetry,

$$X = IL.IN.p_1 \dots p_r = \lambda IL.IN.q_1 \dots q_r \tag{7.21}$$

where λ is an unimportant weight. Thus, in order to define a linear subspace through the origin, it can be made out of points in the generating subspace LN without bothering to send the points onto the quadric. In the following example, we make the x-y plane $IL.IN.a_1.a_2$ and set it in position at z=5, naming it as plane C.

```
> C=pcon(IL,5.0*a3,wap(ILIN,wap(a1,a2)));
```

We now set up a line l through the origin using the angles $\phi = \pi/4$ and $\theta = \pi/6$ of polar coordinates. This is done by making a line which is the z-axis $IL.IN.a_3$ and rotating it into position. The first rotation is about the y-axis by the polar angle θ using invariant line a_3a_1 and then about the z-axis by the azimuth angle ϕ using invariant line a_1a_2. Then we compute the intersection lC of the line with the plane.

```
> l=con(i*pi/4.0,wap(a1,a2),con(i*pi/6.0,wap(a3,a1),wap(ILIN,a3)));
> lC=wap(l,C);
> print_element, lC;
-7.071ia1a4-7.071ia2a4-17.321ia3a4-7.071a1a5-7.071a2a5-17.321a3a5-3.464a4a5
```

The element lC represents the point at the intersection of the line and the plane, but it is returned in the general form of a linear subspace, that is, $lC = \lambda IL.p$ where p is the actual point of intersection on the quadric and λ is some unimportant weight. In order to read out the coordinates of the point lC we use the function `q_from_Z`.

```
> print_element, q_from_Z(lC);
+2.041a1+2.041a2+5.000a3
```

Historically, in the development of this package, the representation of a point p in the form $Z = IL.p$ was called the *Z-form* of the point. Hence, the function `q_from_Z` takes a Z-form of a point and returns the generator of the point as in the last example. From (7.19),

$$Z = \lambda IL.p = \lambda IL(IN + (-1)^{n-1}(IN.L)q) \tag{7.22}$$

and we solve for the generator q. Multiplying by n and L using the fact that q is in LN gives,

$$\begin{aligned} ZN &= \lambda((IN.L)IN + (-1)^{n-1}(IN.L)^2 q) \\ ZNL &= \lambda(IN.L)^2 \end{aligned}$$

so that the solution for the generator is,

$$q = (-1)^{n-1}\left(\frac{ZN}{ZNL} - \frac{IN}{IN.L}\right) . \tag{7.23}$$

There is also the function `p_from_Z` which returns a point p on the quadric given a point specified in $Z = \lambda ILp$ form. This function returns

$$p = \frac{(-1)^{n-1}2(ZNL)ZN}{ZN.IZN} - IL \tag{7.24}$$

and this formula can be found by substituting (7.23) into (7.19). Alternatively, the formula can be obtained by reflecting IL in the hyperplane IZN.

Suppose we have a point specified in $Z = \lambda ILp$ form. Sometimes it is useful to work with the point r at which Z intersects the hyperplane N. This is shown in figure 7.2. The point r is given, up to a weight, by ZN. Therefore, using (7.23), the point r is,

$$r = \frac{(-1)^{n-1}ZN}{ZNL} = \frac{(-1)^{n-1}IN}{IN.L} + q \tag{7.25}$$

This representation is called the *r-form* or the *position vector form* of a point[14]. The function `r_from_Z` takes a point in Z-form and returns the r-form of the point. The following example shows what happens if we take the r-form of the point lC from the previous example.

[14] This is because it mimics the way a position vector is handled in classical vector analysis. In vector analysis, the vectors are really points in a subspace at infinity. In figure 7.2 this subspace is LN. The generators q live in LN so they are analogous to vectors. In vector analysis a position vector is given with respect to an origin point which is the null vector. This practice just covers up the fact that the origin is a point. In equation (7.25) the origin point is IN up to a weight.

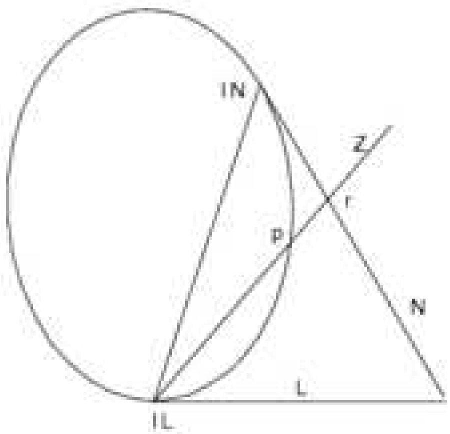

Figure 7.2: Representing a point on hyperplane N

```
> print_element, r_from_Z(lC);
+2.041a1+2.041a2+5.000a3+0.500a4+0.500ia5
> re_part(r_from_Z(lC))(1:3)
[2.04124,2.04124,5]
```

Notice that the Cartesian coordinates are available in the first three slots of the result, so `r_from_Z` is just as useful as `q_from_Z` for sending coordinates to other devices. However, `q_from_Z` is not very useful if its output needs to be processed by further operations of Whitehead's algebra. The problem is that,

```
> print_element, q_from_Z(ILIN);
0
```

cannot be processed any further by operations of Whitehead's algebra, whereas,

```
> print_element, r_from_Z(ILIN);
+0.500a4+0.500ia5
```

gives the sensible result $IN/IN.L$. So, the function `r_from_Z` tends to be used by package functions to output points in situations where we are not sure exactly how the output will be used by an application program. We might send the coordinates to another device or function that does not understand

Whitehead's algebra. Alternatively, we might process the points further using the operations of Whitehead's algebra. Both situations can be handled if the points are output from `r_from_Z`.

7.5.3 Integration on the absolute quadric

We run into a difficultly when we try to integrate on the absolute quadric using the infinitesimal volume element defined in (4.3) because $pIp = 0$ and so all the factors blow up. Clearly, some other expression needs to be used for the volume element on the quadric. Suppose that we want to integrate over some subspace on the quadric. A point on the subspace in the quadric can be specified as $p = p(\theta_1, \ldots, \theta_r)$ using the coordinates θ_i. we can define a linear subspace in the Euclidean space that goes through the points,

$$p, p + d\theta_1 \frac{\partial p}{\partial \theta_1}, \ldots, p + d\theta_r \frac{\partial p}{\partial \theta_r} \ .$$

Such a linear subspace is represented by,

$$d\theta_1 \ldots d\theta_{n-2} IL.p.\frac{\partial p}{\partial \theta_1} \ldots \frac{\partial p}{\partial \theta_{n-2}} \ .$$

It is a linear subspace because it includes the point at infinity IL. However, the above expression cannot be the infinitesimal volume element because it depends on the weight of the point. We have to normalise the point, but we cannot use pIp because it is always zero. The only way to a get a number out of a point is to use pL. So, our volume element must be something like[15],

$$dV = d\theta_1 \ldots d\theta_r IL.\frac{p}{pL}.\frac{\partial}{\partial \theta_1}\left(\frac{p}{pL}\right) \ldots \frac{\partial}{\partial \theta_r}\left(\frac{p}{pL}\right) \ . \tag{7.26}$$

This is an element of point-grade $r + 2$ to integrate over a subspace of dimension r. If we want the volume element to be a number we take it as $\sqrt{-dV.IdV}$. The minus sign is necessary to make the volume element a positive number. From (7.26),

$$\sqrt{-dV.IdV} = d\theta_1 \ldots d\theta_r \sqrt{\frac{\partial}{\partial \theta_1}\left(\frac{p}{pL}\right) \ldots \frac{\partial}{\partial \theta_r}\left(\frac{p}{pL}\right) I \frac{\partial}{\partial \theta_1}\left(\frac{p}{pL}\right) \ldots \frac{\partial}{\partial \theta_r}\left(\frac{p}{pL}\right)} \ . \tag{7.27}$$

[15]See (9.59) for the relation between the volume of a cell of a hypersurface embedded in the absolute quadric and the normal to the hypersurface.

As an example, let us evaluate the volume element for Euclidean space. Take $r = n-2$ and use (7.19) to generate a point on the quadric as,

$$p = IN + (-1)^{n-1}(IN.L)\left(q - \frac{(qIq)}{2}IL\right) \ .$$

The Cartesian coordinates x_i of the point are the components of the generator,

$$q = \sum_{i=1}^{n-2} x_i a_i \ .$$

The normalising factors are $pL = IN.L$. The derivatives are[16],

$$\frac{\partial}{\partial x_i}\left(\frac{p}{pL}\right) = (-1)^{n-1}a_i - x_i IL \ .$$

In this example, it is easier to get $\sqrt{-dV.IdV}$ by substituting in (7.26) than to use the general formula (7.27). So, from (7.26) by employing antisymmetry,

$$dV = \frac{dx_1 \dots dx_{n-2} IL.IN.a_1 \dots a_{n-2}}{IN.L} \ .$$

Calculating,

$$\begin{aligned} (IL.IN.a_1 \dots a_{n-2}).I(IL.IN.a_1 \dots a_{n-2}) &= \\ (-1)^{(n-1)(n-2)}(IL.IN.a_1 \dots a_{n-2}).(L.N.A_1 \dots A_{n-2}) &= -(IN.L)^2 \end{aligned}$$

we find the number-valued volume element is,

$$\sqrt{-dV.IdV} = dx_1 \dots dx_{n-2} \ .$$

7.6 Intersections of hyperspheres

This section takes up the study of hyperspheres from the point at which it was left at the end of section 7.4.

Consider r hyperspheres $S_1, \dots, S_r$. The hypersphere S_1 (say) in the (n-2)-d Euclidean space is the set of points on the quadric at which the element S_1, considered as a hyperplane in the (n-1)-d elliptic space, cuts the quadric.

[16] Remember that $IA_i = a_i$ from (3.13) so,

$$\frac{\partial qIq}{\partial x_i} = 2qI\frac{\partial q}{\partial x_i} = 2qIa_i = (-1)^{n-1}2qA_i = (-1)^{n-1}2x_i \ .$$

However, we have seen that we can think about the hypersphere itself as the element S_1, and not worry about the points it represents by the way it cuts into the quadric. Therefore, the intersection of the r hyperspheres can be studied as the element $S_1 \ldots S_r$ of hyperplane grade r. The set of points represented by $S_1 \ldots S_r$ is the set of points at which $S_1 \ldots S_r$ cuts the quadric. In 3-d Euclidean space, the intersection of two spheres will be a circle. In general, the intersection of r hyperspheres will be a lower dimensional hypersphere. Let us determine the centre and radius of the hyperspherical subspace $S_1 \ldots S_r$ formed by the intersection of r hyperspheres.

The linear subspace which contains the intersection of the hyperspheres is $IL.S_1 \ldots S_r$. In 3-d Euclidean space the intersection of two spheres S_1 and S_2 is the circle S_1S_2. The linear subspace $IL.S_1S_2$ is the plane containing the circle. For example, following section 7.4.2, we create a sphere S_1 of radius 5 units centred on the origin and a sphere S_2 of radius 5 units centred on the point (0,0,5). Then we create the circle S_1S_2 formed by the intersection of the two spheres,

```
> S1=con(-log(5.0),ILIN,L+N);
> S2=pcon(IL,5.0*a3,S1);
> S1S2=wap(S1,S2);
```

and check that it is the correct one in the plane z=5/2 with radius $r = \sqrt{75}/2$. The examples show, by calculating cases of $p.S_1S_2$, that points (r,0,2.5) and (0,r,2.5) are on the circle, and (0,1,2.5) is not on the circle.

```
> r=sqrt(75.0)/2.0;
> print_element, wap(S1S2, pcon(IL,2.5*a3+r*a1,IN));
-0.000a1a2a3a4+0.000a1a2a3a5+0.000a1a2a4a5
> print_element, wap(S1S2, pcon(IL,2.5*a3+r*a2,IN));
-0.000a1a2a3a4+0.000a1a2a3a5+0.000a1a2a4a5
> print_element, wap(S1S2, pcon(IL,2.5*a3+a2,IN));
+35.500ia1a2a3a4+35.500a1a2a3a5+14.200a1a2a4a5
```

In the next example we calculate the linear subspace $IL.S_1S_2$ and show that, modulo a weight, it is the same as the plane z=2.5 which we make by translating the x-y plane.

```
> print_element, wap(IL,S1S2);
-2.000ia1a2a3a4-2.000a1a2a3a5-0.800a1a2a4a5
> print_element, 0.4*i*pcon(IL,2.5*a3,wap(ILIN,wap(a1,a2)));
-2.000ia1a2a3a4-2.000a1a2a3a5-0.800a1a2a4a5
```

So, this example has illustrated that $IL.S_1 \dots S_r$ is the linear subspace which contains the intersection of the hyperspheres.

Now consider the action of a simple rotation which takes $S_1 \dots S_r$ into itself. Let the centres of the individual hyperspheres be $IN_1, \dots, IN_r$. These centres are each invariant under the action of the rotation[17]. In addition, IL is invariant. Therefore, the rotation acts like the identity in the subspace $ILIN_1 \dots IN_r$. Substituting for the N_i using equation (7.16), we find that this subspace can be written as $ILIS_1 \dots IS_r$ up to a weight. The intersection of this subspace with the linear subspace containing $S_1 \dots S_r$ is $ILIS_1 \dots IS_r.(IL.S_1 \dots S_r)$ and this element is of point grade 2. In other words it is a line[18]. Furthermore, this line goes through IL. To see this formally, we can use Müller's first theorem (6.3) to show that,

$$\begin{aligned} &IL(ILIS_1 \dots IS_r.S_1 \dots S_r) \\ &= (-1)^r(ILILIS_1 \dots IS_r)(S_1 \dots S_r) + (ILIS_1 \dots IS_r)(IL.S_1 \dots S_r) \end{aligned}$$

and the first term on the RHS is zero by antisymmetry.

$$IL(ILIS_1 \dots IS_r.S_1 \dots S_r) = (ILIS_1 \dots IS_r)(IL.S_1 \dots S_r) \tag{7.28}$$

The LHS of (7.28) shows that our line passes through IL. We therefore have a Z-form which represents a point on the quadric where the line through IL intersects the quadric. Tentatively, it seems that this point must be the centre of the hyperspherical subspace $S_1 \dots S_r$. It certainly is the centre in our example of a pair of intersecting spheres.

```
> print_element, q_from_Z(wap(wap(IL,I(S1S2)),wap(IL,S1S2)));
+2.500a3
```

Now things turn out rather neatly if we choose our notation carefully at this juncture. Consider the hyperplane L and use the general projection formula (6.10) to write,

$$\begin{aligned} L = L_\parallel + L_\perp \quad &= \quad \frac{(IS_1 \dots IS_r.L).(S_1 \dots S_r)}{(IS_1 \dots IS_r.S_1 \dots S_r)} \\ &+ \quad \frac{(IS_1 \dots IS_r).(LS_1 \dots S_r)}{(IS_1 \dots IS_r.S_1 \dots S_r)} \end{aligned} \tag{7.29}$$

[17] For example, in 3-d Euclidean space, the centres of two spheres that intersect to form a circle are each invariant under the rotation which takes the circle into itself.

[18] $ILIS_1 \dots IS_r$ is of point grade $r+1$ so it is of hyperplane grade $n-r-1$. The other element $IL.S_1 \dots S_r$ is of hyperplane grade $r-1$. The hyperplane grade of the product $ILIS_1 \dots IS_r.(IL.S_1 \dots S_r)$ is therefore $n-r-1+r-1 = n-2$. The point grade of the product is therefore 2.

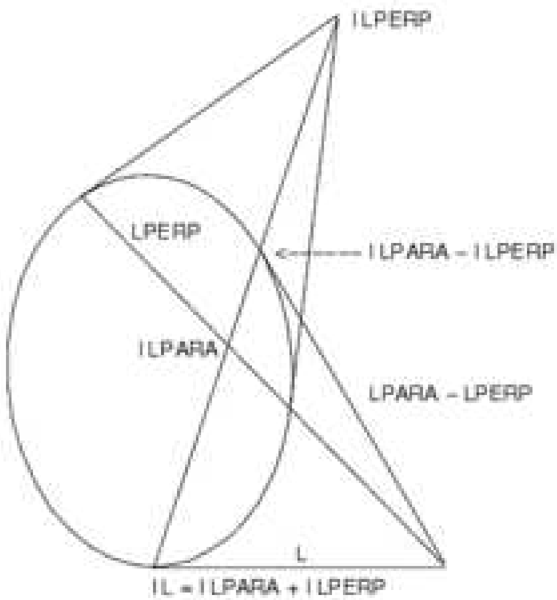

Figure 7.3: Centre of $S_1 \dots S_r$ is $IL_\| - IL_\perp$

and, $IL_\perp.L_\| = 0$. The tentative centre given by equation (7.28) is, up to an unimportant weight, $ILIL_\perp$. We can now draw figure 7.3. Point $IL_\|$ is on $L_\perp$ since $IL_\|.L_\perp = 0$ from the projection formula. It is then clear that the point on the quadric which is the tentative centre of $S_1 \dots S_r$ is $IL_\| - IL_\perp$. We can easily check that $IL_\| - IL_\perp$ lies on the quadric as follows.

$$0 = IL.L = (IL_\| + IL_\perp).(L_\| + L_\perp) = IL_\|.L_\| + IL_\perp.L_\perp$$

Hence,

$$(IL_\| - IL_\perp).(L_\| - L_\perp) = IL_\|.L_\| + IL_\perp.L_\perp = 0 \ .$$

Of course, we could have reflected IL in the hyperplane $L_\perp$ to get $IL_\| - IL_\perp$ on th quadric, but the harmonic conjugate[19] relations are neater.

We can show that $IL_\| - IL_\perp$ really is the centre of $S_1 \dots S_r$ by using the distance formula (3.49) to calculate the distance between $IL_\| - IL_\perp$ and any point p in the subspace $S_1 \dots S_r$ on the quadric. The squared distance is,

$$\begin{aligned} \tau^2 &= -\frac{2(IL_\| - IL_\perp).Ip}{((IL_\| - IL_\perp).L)(pL)} = -\frac{2(IL - 2IL_\perp).Ip}{((IL - 2IL_\perp).L)(pL)} \\ &= \frac{(IL - 2IL_\perp).Ip}{(IL_\perp.L)(pL)} = \frac{(-1)^{n-1}(pL - 2pL_\perp)}{(IL_\perp.L)(pL)} \ . \end{aligned} \quad (7.30)$$

[19]See section 1.15.

Now let's evaluate $pL_\perp$. Note that $p.S_1 \ldots S_r = 0$ so that $pS_j = 0$ for $j = 1, \ldots, r$. From (7.29),

$$
\begin{aligned}
pL_\perp &= \frac{p[(IS_1 \ldots IS_r).(LS_1 \ldots S_r)]}{(IS_1 \ldots IS_r.S_1 \ldots S_r)} \\
&= \frac{p[L(IS_1 \ldots IS_r.S_1 \ldots S_r) + \text{ terms in } S_j]}{(IS_1 \ldots IS_r.S_1 \ldots S_r)} = pL\ . \qquad (7.31)
\end{aligned}
$$

Hence (7.30) becomes,

$$
\tau^2 = \frac{(-1)^n}{(IL_\perp.L)} = \frac{(-1)^{n-1}}{(IL_\parallel.L)} \qquad (7.32)
$$

and we see that this distance is *independent* of the point p on $S_1 \ldots S_r$. Therefore, all the points of $S_1 \ldots S_r$ on the quadric are equidistant from the point $IL_\parallel - IL_\perp$. The tentative conclusion was correct and $IL_\parallel - IL_\perp$ is in fact the centre of $S_1 \ldots S_r$. Furthermore, (7.32) is the squared radius of $S_1 \ldots S_r$. Our final result for the squared radius of the subspace $S_1 \ldots S_r$ formed by the intersection of r hyperspheres is obtained by substituting for $L_\parallel$ from (7.29) into (7.32).

$$
\tau^2 = \frac{(-1)^{n-1}(IS_1 \ldots IS_r.S_1 \ldots S_r)}{IL.[(IS_1 \ldots IS_r.L).(S_1 \ldots S_r)]} \qquad (7.33)
$$

For the example of the intersection of two spheres, the radius of the circle S_1S_2 should be $\sqrt{75}/2$. Evaluating (7.33) gives the correct result.

```
> sqrt(wap(I(S1S2),S1S2)/wap(IL,wap(wap(I(S1S2),L),S1S2)));
4.33013+0i
> sqrt(75)/2.0
4.33013
```

As a further check on formula (7.33), we recover the formula (7.17) for the radius of a single hypersphere by using a single hypersphere in (7.33).

7.6.1 Example: Soddy's generalised formula

When a pair of hyperspheres S_1 and S_2 touch instead of intersecting, the radius of their intersecting subspace is zero. Using equation (7.33), the condition for this to happen is $IS_1IS_2.S_1S_2 = 0$. Now it turns out that in (n-2)-d Euclidean space, n hyperspheres can be arranged so that each hypersphere is touching the other (n-1) hyperspheres. Figure 7.4 shows this for the case

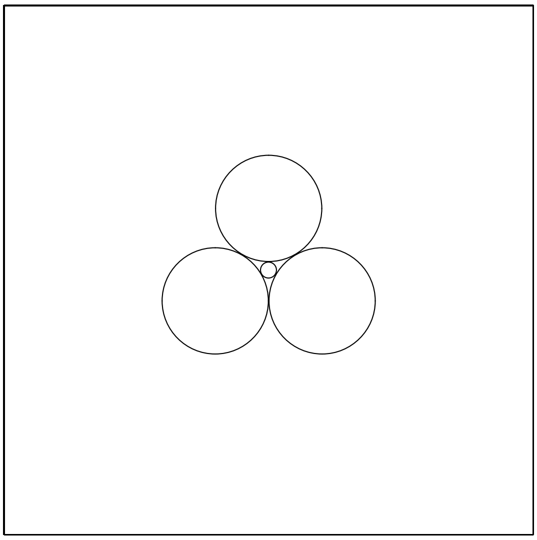

Figure 7.4: Four mutually touching circles

of 4 circles in the Euclidean plane. There is the following interesting relation between the radii of the touching hyperspheres.

$$(\eta_1 + \ldots + \eta_n)^2 = (n-2)(\eta_1^2 + \ldots + \eta_n^2) \tag{7.34}$$

Here, η_i is the signed curvature of the hypersphere S_i. The magnitude of the curvature is the inverse of the radius and when the signs are all the same it indicates that the hyperspheres all touch on their outsides as in figure 7.4. When the sign changes it means that the particular hypersphere touches the others by enclosing them[20]. Coxeter gives an interesting account of the history of this formula in chapter 1 of [14]. The case of $n = 4$ circles in the Euclidean plane (see figure 7.4) was first obtained by Descartes in 1642. The result was re-discovered by Sir Frederick Soddy in 1936. Soddy extended the formula to the case of $n = 5$ spheres in 3-d Euclidean space. The general formula (7.34) was obtained in the 1960s. References to the general case are at the url [24]. We now derive the generalised Soddy formula (7.34) using Whitehead's algebra. It turns out that (7.34) is just a disguised form of the condition $IL.L = 0$ that the point at infinity IL is on the quadric.

As already mentioned, the condition that a pair of hyperspheres S_i, S_j touch is,

$$0 = IS_iIS_j.S_iS_j = (IS_i.S_i)(IS_j.S_j) - (IS_i.S_j)^2 \ . \tag{7.35}$$

From equation (7.17), the radius of hypersphere S_i is real-valued if S_iIS_i is real and positive. There is no loss in generality in taking $S_iIS_i = 1$ so

[20] Section A.9.5 gives a numerical example.

that $IS_i.S_i = (-1)^{n-1}$. From the condition (7.35), the off-diagonal matrix elements are then $IS_i.S_j = \pm 1$. The matrix is of full rank[21] if we take the off-diagonal elements as $IS_i.S_j = -(-1)^{n-1}$ given that the diagonal elements are $IS_i.S_i = (-1)^{n-1}$.

The points IS_i can now be regarded as a set of n reference points[22] for the elliptic space. As in section 1.8, we can then set up hyperplanes H_i dual[23] to these reference points so that $IS_i.H_j = \delta_{ij}$. The hyperspheres S_i can then be expanded in terms of the reference hyperplanes H_k as,

$$S_i = \sum_{k=1}^{n} (IS_k.S_i)H_k$$

so that multiplying by IH_j gives,

$$\delta_{ij} = IS_i.H_j = IH_j.S_i = \sum_{k=1}^{n} (IS_k.S_i)(IH_j.H_k) = \sum_{k=1}^{n} (IS_i.S_k)(IH_k.H_j)$$

and so the matrix with entries $IH_i.H_j$ is the inverse of the matrix with entries $IS_i.S_j$. Since we have just set up the matrix with entries $IS_i.S_j$, it is straightforward to show that the inverse matrix has diagonal entries,

$$IH_i.H_i = \frac{(-1)^{n-1}(n-3)}{2(n-2)}$$

[21]Later in the argument, it will be obvious that the maximum number of mutually touching hyperspheres is n because otherwise the matrix with entries $IS_i.S_j$ would not be invertible.

[22]This set of reference points is interesting in its own right, regardless of its interpretation as a set of touching hyperspheres, because the reference figure with vertices IS_i has edges IS_iIS_j which are all parabolic lines.

[23]The dual hyperplanes H_i also represent hyperspheres in the Euclidean space in the same way that the hyperplanes S_i represent hyperspheres. From equation (1.18),

$$H_j = \frac{(-1)^{j-1} IS_1 \ldots \check{IS_j} \ldots IS_n}{IS_1 \ldots IS_n}$$

and so, for example,

$$H_1H_2 = \frac{IS_2 \ldots IS_n.H_2}{IS_1 \ldots IS_n} = \frac{(IS_2.H_2)IS_3 \ldots IS_n + 0 + \ldots + 0}{IS_1 \ldots IS_n} = \frac{IS_3 \ldots IS_n}{IS_1 \ldots IS_n}$$

and then,

$$IH_1IH_2.H_1H_2 = \frac{IS_3 \ldots IS_n.S_3 \ldots S_n}{IS_1 \ldots IS_n.S_1 \ldots S_n} \ .$$

When $n = 4$ we get $IH_1IH_2.H_1H_2 = 0$ because $IS_3IS_4.S_3S_4 = 0$. So, in the Euclidean plane, the dual hyperplanes represent another set of 4 touching circles. These are drawn in figure 1.5d of Coxeter's book [14]. However, this property of the dual hyperspheres does not hold for $n > 4$. In the higher dimensions, the dual hyperspheres intersect one another.

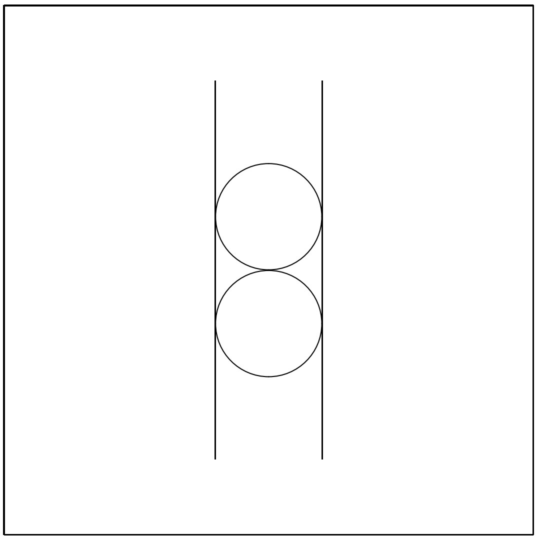

Figure 7.5: Another example of Soddy's formula in the plane.

and off-diagonal entries,

$$IH_i.H_j = \frac{-(-1)^{n-1}}{2(n-2)} \ .$$

Now consider the point at infinity IL and expand L in terms of the reference hyperplanes H_i.

$$L = \sum_{j=1}^{n} \eta_j H_j \tag{7.36}$$

The weights are $\eta_i = IS_i.L$ and, since we have taken $S_i IS_i = 1$, the squared radius of the hypersphere S_i is η_i^{-2} from equation (7.17). Therefore, the η_i are the signed curvatures which appear in the generalised Soddy formula (7.34). The condition,

$$0 = IL.L = \sum_i \sum_j \eta_i \eta_j IH_i.H_j = \sum_{i=1}^{n} \eta_i^2 IH_i.H_i + 2 \sum_{1 \leq i < j \leq n} \eta_i \eta_j IH_i.H_j$$

and substitution of the matrix entries $IH_i.H_j$, gives, after a little re-arrangement, the generalised Soddy formula (7.34).

In the above proof, the touching hyperspheres make a set of reference points IS_i for the elliptic space. However, the radii of the hyperspheres are left unspecified until the relation of the point at infinity IL to the touching hyperspheres is established by equation (7.36). For example, take the case of $n = 4$ touching circles. Most choices for the point at infinity in relation to the touching circles will result in a diagram like figure 7.4. However, if we take the point at infinity as $IL = IS_1IS_2.S_1$, then $IL.S_2 = IS_1IS_2.S_1S_2 = 0$ so that IL lies on S_1 and on S_2. Since S_1 and S_2 now pass through the point at infinity, they are a pair of parallel lines and instead of figure 7.4 we have figure 7.5.

Section A.9 contains additional material on touching hyperspheres as part of an extended example on the use of a Scilab function package for Whitehead's algebra that is useful for calculations in higher dimensional geometry.

7.7 Screws

7.7.1 Eigenpoints of a general congruence

A simple congruence f given by equation (3.26) has two eigenpoints[24] on its invariant line ILM and it acts like the identity in the subspace LM. In other words it has n eigenpoints, but $n-2$ of these are in the subspace LM with each eigenvalue being unity. A more general congruence[25] can be formed as gf where g is a congruence with an invariant line $IL'M'$ which lies in the subspace LM. The congruence gf has two eigenpoints on line ILM, another two on $IL'M'$ and $n-4$ eigenpoints with unity eigenvalues in the subspace $L'M'LM$. Continuing in this way, the most general congruence will be a product of m simple congruences where m is an integer and $2m <= n$. This general congruence will have $2m$ non-unity eigenvalues. If n is odd there will be, additionally, one eigenvalue of unity[26].

7.7.2 Decomposition of a screw

We are interested in the congruence group of 3-d Euclidean space. The elements of this group are the screws. Our 3-d Euclidean space is on the quadric of 4-d elliptic space so $n = 5$. Therefore, the screw will be a product of two simple congruences $h = gf$. In addition to commuting with the absolute polarity I, a screw h must leave the point at infinity IL invariant. Hence IL must be an eigenpoint. Let's take IL as one eigenpoint of the simple congruence g in $h = gf$. We know from equations (3.45,3.46) that

[24] See equations (3.45) and (3.46).

[25] In other words, a collineation which commutes with the absolute polarity.

[26] In the case of n odd, the final congruence could be a reflection given by an equation of the form (3.39), but we ignore this case, as our interest is in congruences which are continuously connected to the identity.

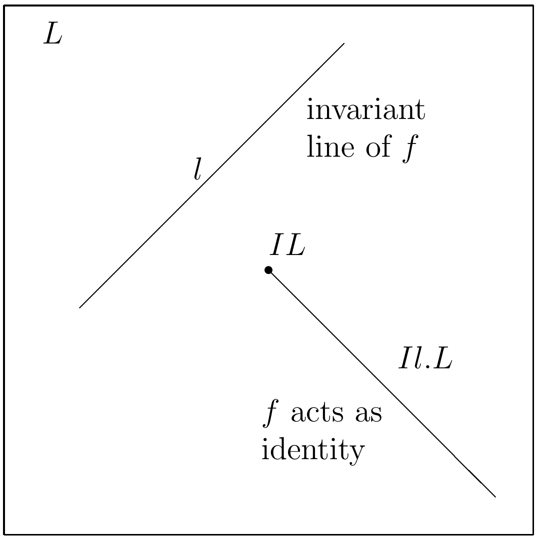

Figure 7.6: Lines in the decomposition of a screw

the pair of eigenpoints of a simple congruence lie on the quadric. Let's take the second eigenpoint of g as the origin IN. There is no loss of generality in doing this because IN is just any point on the quadric. The invariant line of simple congruence g is now $ILIN$. We have already studied such a congruence in section 7.4.2 where it was identified as a dilatation. Dilatations are not experienced in our world[27] so $ILIN$ cannot be the invariant line of g. We must preserve IL as an eigenpoint of g, so the only solution is to move the invariant line so that it passes through IL and lies in the hyperplane L. In this case we have the degenerate case of the parabolic congruence that was studied in section 3.10. In the degenerate case, the two eigenpoints have coalesced to a single eigenpoint at IL. Furthermore, the parabolic congruence has been used in section 7.3.5 to produce translations in the context of Euclidean geometry on the quadric. So, general considerations have led to the screw being decomposed as $h = gf$, where g is a parabolic congruence which has the meaning of a translation. The other simple congruence f will turn out to be a rotation.

We have seen that IL is an eigenpoint of the screw with an eigenvalue of unity (otherwise we would have a dilatation). So, $hIL = IL$ and since h commutes with the polarity, $hL = L$. The screw takes L into itself, but the individual points of L are not invariant. This means that the invariant line l of the simple congruence f must lie in L so $l.L = 0$. The subspace Il of grade 3 in which f acts like the identity contains IL since $IL.Il = I(L.l) = 0$.

[27]See chapter 5.

Therefore, according to section 7.5, Il is the type of element that defines a line in the 3-d Euclidean space on the quadric. The simple congruence f acts like the identity on this line, so the line must be the axis of a rotation in the 3-d Euclidean space. This shows that any screw can be decomposed as $h = gf$ where g is a parabolic congruence (translation) and f is a simple congruence (rotation) with invariant line l in L. Figure 7.6 shows the lines l and $Il.L$ in the hyperplane L. Notice that the lines do not intersect each other. They are actually a pair of skew lines because the hyperplane L is a 3-d subspace.

Let the invariant line of the translation g be $IL.q_2$ where q_2 is a point in L and let p be an arbitrary point in the elliptic space. The action of the translation is given by equation (3.47) with $q_2 = IM$ and $\tau = 1$.

$$gp = \exp(ILq_2.I)p = p - (pIq_2)IL + (pL)q_2 - \frac{(pL)(q_2Iq_2)IL}{2} \tag{7.37}$$

This equation is similar to (7.19). The difference is that (7.37) acts on a general point p whilst (7.19) acted on the origin IN. Also (7.37) is for the case $n = 5$. Now replace p in (7.37) by fp, and note that the term $fp.L = p.f^{-1}L = pL$ by (3.2)[28]. This gives the following formula for a screw[29].

$$hp = gfp = \exp(ILq_2.I)fp = fp - (fp.Iq_2)IL + (pL)q_2 - \frac{(pL)(q_2Iq_2)IL}{2} \tag{7.39}$$

There will be many ways in which a particular screw h can be decomposed into the product $h = gf$. However, there is one decomposition that stands out

[28] Clarification, $f(IL) = IL$ because IL is in the subspace in which f acts like the identity since $IL.Il = 0$. f commutes with the polarity so $fL = L$ and therefore $f^{-1}L = L$.

[29] It is interesting to establish the connection between formula (7.39) for a screw and the representation in classical vector analysis. We have already seen that classical vector analysis takes place in the subspace LN of generators of points on the quadric. So, we use (7.39) to calculate gfp with p given by (7.19) for $n = 5$. Start with fp and note that $f(IN) = IN$ since the rotation axis is through IN.

$$fp = f(IN)+(IN.L)f(q)-(1/2)(IN.L)(qIq)f(IL) = IN+(IN.L)fq-(1/2)(IN.L)(qIq)IL$$

Now calculate gfp using (7.39) and the fact that $q_2Iq_2 = fq_2.Ifq_2$ to get,

$$\begin{aligned} gfp &= IN + (IN.L)fq - (1/2)(IN.L)(qIq)IL \\ &\quad -(IN.L)(fq.Iq_2)IL + (IN.L)q_2 - (1/2)(IN.L)(q_2Iq_2)IL \\ &= IN + (IN.L)(fq + q_2) - (1/2)(IN.L)[(fq + q_2)I(fq + q_2)]IL \ . \end{aligned} \tag{7.38}$$

This equation is in the form of (7.19) with the generator q replaced by $(fq+q_2)$. Therefore the action of a screw can be found by generating the quadric point gfp using the generator $fq + q_2$ where q is the generator of p. This recovers the recipe of classical vector analysis.

as canonical and can be used to study the general properties of an arbitrary screw. Suppose that the invariant line $IL.q_2$ of the translation coincides with the line $Il.L$ in figure 7.6 in which the rotation acts as the identity. In this case $fq_2 = q_2$ and the term $fp.Iq_2$ in (7.39) becomes $fp.Iq_2 = fp.Ifq_2 = pIq_2$. The last three terms on the right of equation (7.39) are now the same as the corresponding terms in (7.37)so,

$$hp = gfp = fp + gp - p \ . \tag{7.40}$$

In the canonical decomposition f and g commute so $h = gf = fg$ because g only has an effect on points on its invariant line $Il.L$ which is where f acts like the identity. Similarly, f only has an effect on points on its invariant line l which lies in the subspace $I(Il.L) = l.IL$ where g acts like the identity. Let p be a point on the quadric. From (7.40) we find,

$$\begin{aligned} gfp.Ip &= (fp + gp - p)Ip = fp.Ip + gp.Ip = fp.Ip + f(gp.Ip) \\ &= fp.Ip + fgp.Ifp = fp.Ip + gfp.Ifp \ . \end{aligned} \tag{7.41}$$

This equation is just Pythagoras' theorem for the right-angled triangle with vertices p, fp, gfp using the formula (3.49) for the squared distance between two points on the quadric. The squared distance between p and gfp is proportional to $gfp.Ip$. This is the hypotenuse of the triangle. The squared lengths of the sides adjacent to the right angle are proportional to $fp.Ip$ and $gfp.Ifp$. This relation shows that the rotational motion $p \to fp$ is perpendicular to the translational motion $fp \to gfp$ and so the motion $p \to gfp$ is a screw motion. The axis of the screw is along the translation. An arbitrary screw will always have a canonical decomposition into a rotation and a translation along the axis of the rotation. This means that an arbitrary screw will always be characterised by a pair of skew invariant lines l and $Il.L$ in L as shown in figure 7.6. The screw leaves each line invariant so $h(l) = l$ and $h(Il.L) = Il.l$. These results follow because in the canonical decomposition $h = gf$, l is the invariant line of f and $Il.L$ is the invariant line of g, and each line lies in the subspace in which the other simple congrence acts like the identity.

In applications it is more useful to work with decomposition (7.39) in which there is no special relation between the invariant line $IL.q_2$ of g and the line $Il.L$ in which f acts like the identity (see figure 7.6). Nevertheless, in such cases, it is easy to find one of the invariant lines of the canonical decomposition. Suppose we have a screw $h = gf$ where f is a rotation with axis through the points p_1 and p_2 on the quadric. The subspace in which the rotation acts like the identity is then $Il = IL.p_1.p_2$. The line in which this subspace intersects L is,

$$Il.L = (IL.p_1.p_2).L = -(p_1L)IL.p_2 + (p_2L)IL.p_1 \ . \tag{7.42}$$

This line is shown in figure 7.6. It is not the same line as the invariant line $IL.q_2$ of the translation g because we have said that the decomposition $h = gf$ is not the canonical one for this screw h. Nevertheless, let us calculate the action of h on $Il.L$. From (7.39) and (7.42),

$$\begin{aligned} h((IL.p_1.p_2).L) &= -(p_1L)IL.h(p_2) + (p_2L)IL.h(p_1) \\ &= -(p_1L)IL.(fp_2 + (p_2L)q_2) + (p_2L)IL.(fp_1 + (p_1L)q_2) \\ &= -(p_1L)IL.(p_2 + (p_2L)q_2) + (p_2L)IL.(p_1 + (p_1L)q_2) \\ &= -(p_1L)IL.p_2 + (p_2L)IL.p_1 = (IL.p_1.p_2).L \ . \end{aligned} \tag{7.43}$$

So, the line $(IL.p_1.p_2).L$ is invariant under the action of the screw. However, we know that a screw will always have two invariant lines, so this line $(IL.p_1.p_2).L$ must be the invariant line through IL. In other words it is the invariant line of the translation in the canonical decomposition of the screw. The action of a translation is to move everthing in a particular direction. So, we see that if we make a screw from a rotation about some axis followed by a translation in some non-canonical direction - not along the axis of the rotation - the screw action of the combination will have its axis parallel to the axis of the original rotation[30].

7.7.3 The screw as a pair of generating points

Consider a non-canonical decomposition of a screw as $h = gf$. Let the axis of the rotation f be through the origin IN in direction q_1. The rotation axis is therefore $IL.IN.q_1$ where q_1 lies in the generating subspace LN. The invariant line of the rotation is $I(IL.IN.q_1) = LNIq_1$. The term $\sqrt{-ILM.LM}$

[30]Some clarification may be needed. Suppose that the axis of the rotation f is $IL.IN.p_2$. This is the same as the axis in the text except that point p_1 has been replaced by the origin IN. There is no loss of generality because IN can be any point on the quadric. The rotation axis $IL.IN.p_2$ is a line through the points IN and p_2. However, we know from (7.21) that $IL.IN.p_2$ can be written, modulo a weight, as $IL.IN.q$ where q is the generator of p_2 and q lies in the generating subspace LN. The intersection of $IL.IN.q$ on L is $(IL.IN.q).L = -(IN.L)IL.q$ and so from (7.43), $IL.q$ is the invariant line of the translation in the canonical decomposition of the screw. The axis of the screw motion is therefore in the direction represented by the generator q. But, the axis of the original rotation was $IL.IN.q$ which is a line through the origin in the direction represented by q. Therefore, if we have a screw which is made by a rotation about an axis through some point IN pointing in direction q, followed by a translation in some other direction, the resulting screw motion has the axis of the screw pointing in the direction q.

that appears in (3.26) is,

$$LNIq_1.(IL.IN.q_1) = (IL.IN.q_1).LNIq_1 = -(IN.L)^2(q_1Iq_1)$$

and so from (3.26) the rotation is,

$$f = \exp\left(\frac{\tau LNIq_1.I}{IN.L\sqrt{q_1Iq_1}}\right) .$$

Now,

```
> print_element, LN/IN_L;
+1.000ia1a2a3
```

so that the congruence parameter needs to be $\tau = -i\theta$ where θ is the rotation angle. If we take the magnitude of the rotation angle as $\theta = \sqrt{q_1Iq_1}$ with the direction encoded by the sign of q_1 in the invariant line, then,

$$f = \exp\left(\frac{-iLNIq_1.I}{IN.L}\right) \tag{7.44}$$

and a rotation about an axis through IN depends on a generating point q_1. Similarly, the translation $g = \exp(ILq_2.I)$ and so the screw $h = gf$ can be represented by a pair of generators (q_1, q_2).

The following example illustrates the use of the function `screw` from package `Euclid.i`. The screw is represented by an array of two points. A rotation of $\pi/2$ about the z-axis is set up in q_1 and a translation of 10 units along the z-axis is set up in q_2. A point with coordinates (1,1,0) is represented on the quadric as p. The screw is allowed to act on this point and the result is printed out as coordinates as described in section 7.5.2.

```
> q=array(complex,2,5);
> q(1,)=0.5*pi*a3;
> q(2,)=10.0*a3;
> p=pcon(IL,a1+a2,IN);
> print_element, q_from_p(screw(q,p));
-1.000a1+1.000a2+10.000a3
```

The next example shows that `screw` is not doing anything that cannot be done by a combination of `con` and `pcon`.

```
> print_element, q_from_p(pcon(IL,10.0*a3,con(-i*pi/2.0,I(wap(ILIN,a3)),p)));
-1.000a1+1.000a2+10.000a3
```

The advantage of using `screw` becomes evident when we consider how to combine chains of screws.

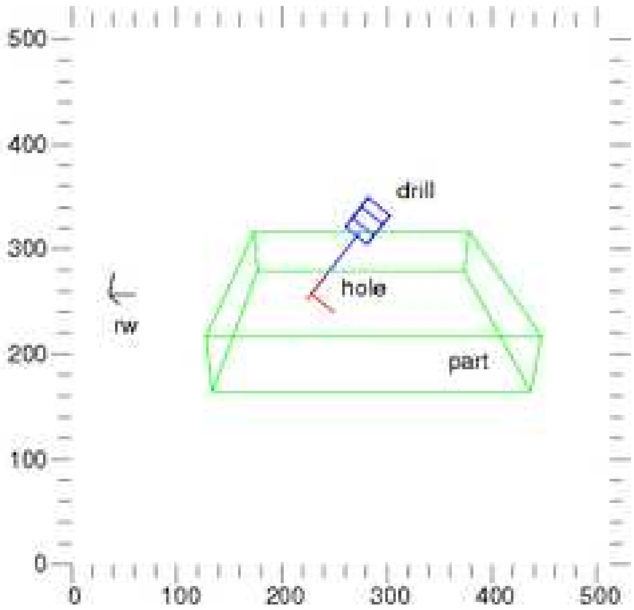

Figure 7.7: Reference frames in a toy drilling program

7.7.4 A notational convention for reference frames

A reference frame is a set of reference points $a_1, \ldots, a_n$. An observer is considered to be co-located with some point in the reference frame such as the origin IN. This text always uses the active point of view. Consider two reference frames/observers[31] $\mathcal{A}$ and $\mathcal{B}$. Initially they are the same. Then the points of $\mathcal{B}$ are moved away from $\mathcal{A}$ by the action of a screw $h_{\mathcal{BA}}$. We label the screw by the subscript $\mathcal{BA}$ to indicate that the motion produced by the screw is $\mathcal{B} \to \mathcal{A}$. Consider the origin IN of frame $\mathcal{A}$. From the point of view of frame $\mathcal{B}$, the point at the origin of frame $\mathcal{A}$ appears to be at point $h_{\mathcal{BA}}IN$. Similarly, an arbitrary element X in frame $\mathcal{A}$ appears as element $h_{\mathcal{BA}}X$ in frame $\mathcal{B}$.

As an example consider figure 7.7 which shows a robotic drilling machine making a hole in a part. We have reference frames in each of the objects. The `rw` frame stands for *robot world* which is a frame in the working volume of the robot. The frame `part` is fixed in the part. In the example, the part is a slab of material of size 10x10x2 units. The part frame is in the centre of the slab. The part is positioned in a vice. It has been rolled by 45 degrees about the x-axis of the robot world and shifted by 10 units along the x-axis of the robot world. The position of the part is specified by the screw variable `rw_part`

[31] Transformations between the reference frames used by two observers have been considered at various points in the current text. See sections 1.2, 3.2 and 7.3.4.

which represents the screw which produces the movement `rw -> part`. The screw variable `rw_part` represents the screw by a pair of generating points as explained in section 7.7.3.

```
> rw_part=array(complex,2,5);
> rw_part(1,)=0.25*pi*a1;
> rw_part(2,)=10.0*a1;
```

In the next example, we find out how the origin of the part frame appears in the rw frame. The function `q_from_p` is used because the result of the screw function is a point on the quadric[32].

```
> print_element, q_from_p(screw(rw_part,IN));
+10.000a1
```

Incidentally, the screw could have been defined in a single line. The transpose is necessary because Yorick fills the columns of arrays first.

```
> rw_part=transpose([0.25*pi*a1,10.0*a1]);
```

Continuing with the example, suppose that the robot is to drill a hole into the part at the coordinates (-2,-2,1) in the part frame. The hole is not normal to the surface of the part, instead the axis of the hole is pass through the point (-1,-1,2) in the part frame. We start by setting up the point at the break-out of the hole and the axis of the hole, both elements being evaluated in the part frame.

```
> base=pcon(IL,-2.0*a1-2.0*a2+a3,IN);
> pt=pcon(IL,-1.0*a1-1.0*a2+2.0*a3,IN);
> axis=wap(IL,wap(base,pt));
```

The task is to position the drill to make this hole. To solve this problem, we need a reference frame `hole` with origin IN on the point at which the hole breaks out of the surface of the part, and with its z-axis along the axis of the hole, with the positive direction being out of the hole. This task often arises in applications, so we consider it in general terms before applying the method to the example of figure 7.7.

The problem is to find a screw gf which accomplishes the motion $\mathcal{B} \rightarrow \mathcal{A}$ such that the origin IN of frame $\mathcal{A}$ appears at point $g(IN)$ in frame $\mathcal{B}$ and the z-axis $IL.IN.a_3$ of frame $\mathcal{A}$ appears as the known line l in frame $\mathcal{B}$. The z-axis of frame $\mathcal{A}$ appears in frame $\mathcal{B}$ as line, $gf(IL.IN.a_3) = g(IL.IN.fa_3)$

[32]See section 7.5.2.

because f is a rotation about an axis through the origin so IL and IN are invariant under f. Similarly IL is invariant under g, and from (7.37),

$$gf(IL.IN.a_3) = IL.(IN + (IN.L)q_2).fa_3$$

where q_2 is the generator of the translation g. On multiplying by LN we extract fa_3. Hence, up to a weight, fa_3 is given by $l.LN$. Since f takes a_3 to the known point fa_3, this immediately gives f as the rotation with invariant line $a_3.fa_3$ and so the screw gf is determined. Let's see how this works on the example.

The rotation has to go $a_3 \to fa_3$ and fa_3 is the point calculated as e_2 in the example. In order not to get the rotation wrong by an angle of π, which could happen if the signs are not correct, we extract e_1 as the point on LN from the z-axis $IL.IN.a_3$ in the same way as the extraction of e_2. The rotation angle θ is, modulo a factor of i, the congruence parameter needed to take a_3 to fa_3. This is given by the `distance` function in package `Whitehead4d.i` which implements equation (3.32)[33].

```
> e1=wap(wap(ILIN,a3),LN);
> print_element, e1;
-4.000a3
> e2=wap(axis,LN);
> print_element, e2;
-8.000a1-8.000a2-8.000a3
> theta=-i*distance(e1,e2);
```

[33]In general, the distance between two points p_1 and p_2 is the value of the congruence parameter τ in the congruence f_τ which takes the first point into the second point as $p_2 = f_\tau p_1$. In elliptic or hyperbolic space there is not really any distinction between distance and angle as shown by the formulae in sections 3.8 and 3.8.2. This is why we can have a single function that can be used to get distance or angle. There are two reasons behind this property.

1. The distance/angle between two elements is the value of the congruence parameter that carries the first element into the second. The elements can be points, lines, planes or hyperplanes.
2. There is one formula (3.26) for a simple congruence.

However, in Euclidean space the situation is messier because distance between two points is given by the parameter (3.49) of a parabolic congruence. When considering angles, the parameter is always that of an elliptic congruence given by (3.26). So, in Euclidean space, the distinction between distance and angle is because distance uses a parabolic congruence to map between points on the quadric, whilst angle needs an elliptic congruence to map between elements that do not live on the quadric. The distinction between distance and angle would still turn up if we treated Euclidean geometry using the approach of chapter 5.

```
> theta*180.0/pi;
54.7356+0i
```

The invariant line of the rotation is e_1e_2 but we want to represent the screw gf by a pair of generating points. For this we need the rotation axis Ie_1e_2 expressed as a generating point in LN. Furthermore, the magitude of the rotation generator must be the rotation angle. Hence, we first normalize the generator and then multiply by the rotation angle. Finally, we define the frame on the hole by setting up the screw for the movement `part -> hole`.

```
> e3=wap(I(wap(e1,e2)),-i*LN);
> e3=e3/sqrt(wap(e3,I(e3)));
> print_element, e3;
-0.707a1+0.707a2
> part_hole=transpose([theta*e3,q_from_p(base)]);
```

We can check that the screw `part_hole` is sensible by checking that a point on the z-axis axis of the hole frame appears on the known axis of the hole in the part frame.

```
> print_element, wap(axis,screw(part_hole,pcon(IL,5.0*a3,IN)));
+0.000a1a3a4a5-0.000a2a3a4a5
```

The drill frame has its origin IN at the tip of the drill and the z-axis of the frame extends upwards along the shaft of the drill into the drill body. Hence, the drill can now be put into position by the following assignment.

```
> part_drill=part_hole;
```

The drill is positioned in the working volume of the robot by commands which are screws for the movements of the form `rw -> drill`. Thus in order to command the robot drill into position we have to calculate the screw `rw_drill`. In order to do this we need to combine the screws `rw -> part` and `part -> drill` because such a combination of motions will produce `rw -> drill`. The package `Euclid.i` contains the function `Donkin` which uses Donkin's theorem[34] to combine pairs of screws. The internal working of `Donkin` will be described in section 7.7.5. However, for now, we just accept that it works. Notice that it is easy to combine screws using `Donkin` and the notational convention we have employed for the screws. The names of the screws correspond to the motions they produce and so chains of screws can be readily produced and understood. This notation is a great assistance in programming robotic applications where there are many reference frames.

[34]See section 3.9.3.

```
> rw_drill=Donkin(rw_part,part_drill);
```

As a check, the position of the drill in the rw frame should correspond to the position of the hole in the rw frame.

```
> print_element, rw_drill(2,);
+8.000a1-2.121a2-0.707a3
> print_element, q_from_p(screw(rw_part,base));
+8.000a1-2.121a2-0.707a3
```

There are two other functions which are useful for robotic applications. The package `Euclid.i` contains the function `reverse` which is used to output (say) `drill_rw` from `rw_drill`. Also, a robotic drilling machine would be unlikely to understand our screw description, and it would probably need to be commanded by a shift vector for the translation and some Euler angles for the orientation of the drill. The package `utilities.i` provides the function `Euler_angles` for this purpose. In this, the end of the drilling example, we output the translation and rotation angles for the drill.

```
> print_element, rw_drill(2,);
+8.000a1-2.121a2-0.707a3
> (180.0/pi)*Euler_angles(rw_drill(1,));
[10.3711,33.8953,18.1683]
```

7.7.5 Combining screws using Donkin's theorem

Section 7.7.3 has shown that a screw can be represented by a pair of points (q_1, q_2) in the generating plane LN. The generator q_1 specifies the rotation and q_2 specifies the translation. Section 7.7.4 introduced the function `Donkin` for combining two screws which are each represented as generator pairs. This function returns the generator pair (q_1'', q_2'') that specifies the composite screw $g'f'gf$ where gf is represented by (q_1, q_2) and $g'f'$ is represented by (q_1', q_2'). The current section explains the working of `Donkin`.

Consider two screws given by equation (7.39).

$$\begin{aligned} hp &= gfp = fp - (fp.Iq_2)IL + (pL)q_2 - (1/2)(pL)(q_2Iq_2)IL \\ h'p &= g'f'p = f'p - (f'p.Iq_2')IL + (pL)q_2' - (1/2)(pL)(q_2'Iq_2')IL \end{aligned}$$

We start calculating the combination $h'hp$ and note that $gfp.L = gfp.gfL = pL$.

$$g'f'gfp = f'(gfp) - (f'(gfp).Iq_2')IL + (gfp.L)q_2' - (1/2)(gfp.L)(q_2'Iq_2')IL$$

$$
\begin{aligned}
&= f'[fp - (fp.Iq_2)IL + (pL)q_2 - (1/2)(pL)(q_2Iq_2)IL] \\
&\quad -(f'[fp - (fp.Iq_2)IL + (pL)q_2 - (1/2)(pL)(q_2Iq_2)IL].Iq_2')IL \\
&\quad +(pL)q_2' - (1/2)(pL)(q_2'Iq_2')IL \\
&= f'fp - (fp.Iq_2)IL + (pL)f'q_2 - (1/2)(pL)(q_2Iq_2)IL \\
&\quad -([f'fp - (fp.Iq_2)IL + (pL)f'q_2 - (1/2)(pL)(q_2Iq_2)IL].Iq_2')IL \\
&\quad +(pL)q_2' - (1/2)(pL)(q_2'Iq_2')IL
\end{aligned}
$$

The factor $IL.Iq_2' = 0$ so this removes some terms.

$$
\begin{aligned}
g'f'gfp &= f'fp - (fp.Iq_2)IL + (pL)f'q_2 - (1/2)(pL)(q_2Iq_2)IL \\
&\quad -([f'fp + (pL)f'q_2].Iq_2')IL + (pL)q_2' - (1/2)(pL)(q_2'Iq_2')IL
\end{aligned}
$$

Now replace every occurence of q_2 by $f'q_2$. So, $q_2Iq_2 = f'q_2.If'q_2$ and $fp.Iq_2 = f'fp.If'q_2$.

$$
\begin{aligned}
g'f'gfp &= f'fp - (f'fp.If'q_2)IL + (pL)f'q_2 - (1/2)(pL)(f'q_2.If'q_2)IL \\
&\quad -([f'fp + (pL)f'q_2].Iq_2')IL + (pL)q_2' - (1/2)(pL)(q_2'Iq_2')IL
\end{aligned}
$$

Finally, by grouping the terms together the result can be put into the form (7.39) of a single screw with rotation $f'f$ and a translation with invariant line $IL.(f'q_2 + q_2')$.

$$
\begin{aligned}
g'f'gfp &= f'fp - [f'fp.I(f'q_2 + q_2')]IL + (pL)(f'q_2 + q_2') \\
&\quad -\frac{(pL)[(f'q_2 + q_2).I(f'q_2 + q_2)]IL}{2}
\end{aligned}
\tag{7.45}
$$

So, if gf is represented by the generator pair (q_1, q_2) and $g'f'$ is represented by (q_1', q_2'), then $g'f'gf$ is represented by (q_1'', q_2'') and equation (7.45) shows that the second composite generator is $q_2'' = (f'q_2 + q_2')$. However, we still need a way of getting the composite generator q_1'' that stands for $f'f$. Donkin's theorem in section 3.9.3 solves this problem. The axes of the rotations are both through the origin IN so that the invariant lines of f and f' both lie in plane LN. Therefore, we can construct Donkin's triangle of figure 3.2 in order to find the invariant line and angle for the composite rotation $f'f$ and so fill in the composite generator q_1''.

The end of section 7.7.4 also mentioned the function `reverse` which finds the inverse of a screw gf given by (q_1, q_2). From (7.45) it can be seen that $(gf)^{-1}$ is represented by the generator pair $(-q_1, -f^{-1}q_2)$.

Chapter 8

Computer Vision Applications

8.1 Introduction

A camera maps straight lines in its object space to straight lines in the image in the graphics window. This means that the action of a camera is naturally modelled as a collineation[1] within the framework of projective geometry. However, in computer vision, the object space of the camera is a 3-d Euclidean space. So, in computer vision we need to be able to mix notions of projective geometry with those of Euclidean geometry. This makes Whitehead's algebra a natural mathematical arena for computer vision.

8.2 Collineations revisited

A collineation is a linear map of points to points. Some properties of collineations were considered in section 3.2. In this section we show that exactly $r+1$ point correspondences are needed to define a collineation between two linear subspaces of point grade r.

Let $a_1, \ldots, a_n$ be arbitrary reference points for the entire projective space. The dual hyperplanes are $A_1, \ldots, A_n$ with $a_i A_j = \delta_{ij}$. The domain of the collineation f is assumed to be the linear subspace $a_1 \ldots a_r$. Pick an extra reference point a_0 in this subspace. Similarly, let $b_1, \ldots, b_n$ be another set of arbitrary reference points for the entire projective space. The dual hyperplanes are $B_1, \ldots, B_n$ and $b_i B_j = \delta_{ij}$. The range of the collineation f is assumed to be the linear subspace $b_1 \ldots b_r$. Also, there is an extra reference point b_0 in this subspace. We want the collineation f to map the $r+1$ points a_i into the corresponding points b_i. However, since a point is only defined up

[1]See section 3.2.

to multiplication by an arbitrary weight, we can have $r+1$ free weights λ_i and only demand that $fa_i = \lambda_i b_i$ for $i = 0, 1, \ldots, r$. The weights are found by considering the action $fa_0 = \lambda_0 b_0$. The extra points a_0 and b_0 will have the following expansions in terms of the other reference points.

$$a_0 = \sum_{i=1}^{r} (a_0 A_i) a_i \tag{8.1}$$

$$b_0 = \sum_{i=1}^{r} (b_0 B_i) b_i \tag{8.2}$$

Now,

$$fa_0 = \sum_{i=1}^{r} (a_0 A_i) fa_i = \sum_{i=1}^{r} (a_0 A_i) \lambda_i b_i$$

but also,

$$fa_0 = \lambda_0 b_0 = \lambda_0 \sum_{i=1}^{r} (b_0 B_i) b_i$$

so that,

$$\lambda_i = \frac{\lambda_0 (b_0 B_i)}{(a_0 A_i)} \ .$$

The collineation is therefore completely defined by its action on the r reference points $a_1, \ldots, a_r$.

$$fa_i = \lambda_0 \frac{(b_0 B_i)}{(a_0 A_i)} b_i \tag{8.3}$$

Notice that equation (8.3) reproduces $fa_0 = \lambda_0 b_0$ by substituting the expansions (8.1) and (8.2). Summarising, we have shown that $r+1$ point correspondences are needed to fix a collineation that maps between two linear subspaces of point grade r. Furthermore, equation (8.3) defines the collineation in terms of the point correspondences. The weight λ_0 is a constant overall weight so there is no loss of generality in setting $\lambda_0 = 1$.

Everything becomes a lot simpler if we work with re-weighted reference points $a'_i = (a_0 A_i) a_i$ and $b'_i = (b_0 B_i) b_i$ for $i = 1, \ldots, r$ because then,

$$a_0 = \sum_{i=1}^{r} a'_i \tag{8.4}$$

$$b_0 = \sum_{i=1}^{r} b'_i \tag{8.5}$$

$$fa'_i = b'_i \text{ for } i = 1, \ldots, r \tag{8.6}$$

$$fa_0 = b_0 \ . \tag{8.7}$$

The hyperplanes dual to the re-weighted reference points are given by,

$$A'_i = \frac{A_i}{a_0 A_i} \tag{8.8}$$

$$B'_i = \frac{B_i}{b_0 B_i} \tag{8.9}$$

because then $a'_i A'_j = a_i A_j = \delta_{ij}$ and $b'_i B'_j = \delta_{ij}$ for $i, j = 1, \ldots, r$. The problem of working out the action of the collineation on an arbitrary point p in the subspace $a_1 \ldots a_r$ is easily done using the re-weighted reference points.

$$fp = f\left(\sum_{i=1}^{r}(pA'_i)a'_i\right) = \sum_{i=1}^{r}(pA'_i)b'_i \tag{8.10}$$

8.3 Example: The collineation for a camera

The theory of section 8.2 is now used to model a camera as a collineation. The Euclidean geometry in the object space of the camera is handled using the methods of chapter 7.

The camera produces a sharp image of an object plane in a graphics window of the computer. The object plane of the camera is $IL.v_1v_2v_3$ where IL is the point at infinity and v_1, v_2, v_3 are points on the quadric[2]. The graphics window of the computer is $IL.w_1w_2w_3$ where w_1, w_2, w_3 are points on the quadric. The camera is a collineation f which maps the object plane subspace $IL.v_1v_2v_3$ into the graphics window subspace $IL.w_1w_2w_3$. These subspaces are of point grade 4, so, according to section 8.2, we need 5 point correspondences between the two subspaces to define the collineation. The first of these point correspondences must be $IL \mapsto IL$ because otherwise the graphics window would represent a sphere[3] instead of a plane. The next three point correspondences are $v_j \mapsto w_j$ for $j = 1, 2, 3$. These are point correspondences between object plane reference points and their corresponding image points in the graphics window. Finally, we need a fifth point correspondence $v_0 \mapsto w_0$. Here, v_0 is a point on the quadric which represents a point on the object reference plane. Similarly, w_0 is a quadric point representing the image of v_0 in the graphics window. Let's set up these correspondences using Yorick and calculate a collineation for a camera.

Let's take the object reference plane as the plane z=5. The four object plane reference points $v_0, \ldots v_3$ are taken as (-10,0,5), (-10,-10,5), (10,-10,5),

[2]See section 7.5.
[3]See section 7.4.

(10,10,5) respectively[4]. We set up the four object plane reference points $v_0, \ldots, v_3$ on the quadric.

```
> v0=pcon(IL,-10.0*a1+5.0*a3,IN);
> v1=pcon(IL,10.0*(-a1-a2)+5.0*a3,IN);
> v2=pcon(IL,10.0*(a1-a2)+5.0*a3,IN);
> v3=pcon(IL,10.0*(a1+a2)+5.0*a3,IN);
```

Let's assume that the graphics window is a 512x512 image with the (0,0) at the lower left. It is assumed that the 20x20 object plane fills the 512x512 image. The images $w_0, \ldots, w_3$ that correspond to the object points $v_0, \ldots, v_3$ are therefore (0,256), (0,0), (512,0) and (512,512).

```
> w0=pcon(IL,256.0*a2,IN);
> w1=IN;
> w2=pcon(IL,512.0*a1,IN);
> w3=pcon(IL,512.0*a1+512.0*a2,IN);
```

The theory in section 8.2 shows that we need to compute some dual hyperplanes, so it helps to have a full set of reference points $v_1, \ldots, v_5$. To this end, we take $v_4 = IL$ so that our grade $r = 4$ linear subspace is $v_1v_2v_3IL = v_1v_2v_3v_4$. The remaining point v_5 that completes the set of reference points for the entire projective space can be anything, provided it does not lie in the subspace $IL.v_1v_2v_3$. A suitable choice is a_3. The example checks that this is the case. The same goes for the graphics window subspace, so we create $w_4 = IL$ and $w_5 = a_3$.

```
> v4=w4=IL;
> v5=w5=a3;
> print_element, wap(a3,wap(IL,wap(v1,wap(v2,v3))));
-800.000ia1a2a3a4a5
```

We can now compute the dual hyperplanes V_i and W_i using the formula (1.18). Notice that V_5 and W_5 are the object plane and graphics window subspaces, but they are not used in the formulae of section 8.2, so we do not calculate them.

[4]Three of these points are the vertices of a 20x20 square. The reason that v_0 is set at (-10,0,5) in the middle of the left hand edge of the square instead of on the upper left vertex at (-10,10,5) is because the latter position would make v_0 linearly dependent on the other three vertices and so we would be defining a collineation from the circle $v_1v_2v_3$ instead of the plane $ILv_1v_2v_3$.

```
> v1v2v3v4v5=wap(v1,wap(v2,wap(v3,wap(v4,v5))));
> V1=wap(v2,wap(v3,wap(v4,v5)))/v1v2v3v4v5;
> V2=-wap(v1,wap(v3,wap(v4,v5)))/v1v2v3v4v5;
> V3=wap(v1,wap(v2,wap(v4,v5)))/v1v2v3v4v5;
> V4=-wap(v1,wap(v2,wap(v3,v5)))/v1v2v3v4v5;
> w1w2w3w4w5=wap(w1,wap(w2,wap(w3,wap(w4,w5))));
> W1=wap(w2,wap(w3,wap(w4,w5)))/w1w2w3w4w5;
> W2=-wap(w1,wap(w3,wap(w4,w5)))/w1w2w3w4w5;
> W3=wap(w1,wap(w2,wap(w4,w5)))/w1w2w3w4w5;
> W4=-wap(w1,wap(w2,wap(w3,w5)))/w1w2w3w4w5;
```

These points and hyperplanes obey $v_i V_j = w_i W_j = \delta_{ij}$.

```
> print_element, wap(v1,V1);
+1.000a1a2a3a4a5
> print_element, wap(v1,V2);
-0.000a1a2a3a4a5
```

We just need the re-weighted versions of the points and dual hyperplanes sufficient to calculate the terms in equation (8.10).

```
V1=V1/wap(v0,V1);V2=V2/wap(v0,V2);
V3=V3/wap(v0,V3);V4=V4/wap(v0,V4);
w1=w1*wap(w0,W1);w2=w2*wap(w0,W2);
w3=w3*wap(w0,W3);w4=w4*wap(w0,W4);
```

The collineation for the camera is now available. The point (0,0,5) on the object plane has not been used to define the collineation. It is clear that this object point should form an image point in the graphics window at (256,256). To check this, we define the point as p and compute fp using equation (8.10).

```
> p=pcon(IL,5.0*a3,IN);
> fp=wap(p,V1)*w1+wap(p,V2)*w2+wap(p,V3)*w3+wap(p,V4)*w4;
> print_element, q_from_p(fp);
+256.000a1+256.000a2
```

In obtaining these results we have ignored the fact that in the object reference plane $IL.v_1v_2v_3$, a point is $IL.p$ as shown in sections 7.5 and 7.5.2. The collineation f is not a congruence of Euclidean space, so it does not commute with the polarity. Consequently, if p is on the quadric so that $pIp = 0$, then $fp.Ifp \neq 0$ in general. In this example, the corresponding points $v_i \mapsto w_i$ for $i = 0, 1, 2, 3$ are all on the quadric and points p on the quadric which are close to the object plane reference points will map to points fp which are on the quadric to a good approximation. However, if we take p as a point on the quadric representing a point on the object reference plane, but far away from the reference points, we find that fp is no longer on the quadric.

```
> p=pcon(IL,-1574.0*a1+6175.0*a2+5.0*a3,IN);
> fp=wap(p,V1)*w1+wap(p,V2)*w2+wap(p,V3)*w3+wap(p,V4)*w4;
> print_element, wap(fp,I(fp))
-53.760a1a2a3a4a5
```

This does not matter because we are really representing a point in the Euclidean space as $IL.p$ and the action of the camera collineation is $f(IL.p) = IL.fp$ which is the representation of a point in the graphics window. So, instead of evaluating fp using (8.10), we should evaluate $IL.fp$ and then use the functions described in section 7.5.2 to extract the most appropriate representation of the point. Furthermore, since $IL.p$ represents the point on the object reference plane, it follows that p does not have to be on the quadric. Instead of using p, we can use any point $IL + \xi p$ on $IL.p$ because $f(IL.(IL + \xi p)) = IL.fp$ up to a weight. In the next example we take the point p as (0,0,5) on the object reference plane, and transform the point $(ILp).N$ in order to show that the point itself does not have to lie on the quadric[5]. Notice that there is no need to keep the term in $w_4 = IL$ in the formula for fp when evaluating $IL.fp$ because it is removed by antisymmetry.

```
> p=pcon(IL,5.0*a3,IN);
> p=wap(wap(IL,p),N);
> print_element, wap(p,I(p));
+400.000a1a2a3a4a5
> ILfp=wap(IL,wap(p,V1)*w1+wap(p,V2)*w2+wap(p,V3)*w3);
> print_element, q_from_Z(ILfp);
+256.000a1+256.000a2
```

It is clear that the theory of section 8.2 is completely symmetrical and we can easily take a point in the graphics world and obtain the corresponding point in the object space reference plane. In the next example, we set up the image point (256,256) and calculate the corresponding point on the object reference plane using equation (8.10) with the linear subspaces swopped[6].

```
> fp=pcon(IL,256.0*a1+256.0*a2,IN);
> ILp=wap(IL,wap(fp,W1)*v1+wap(fp,W2)*v2+wap(fp,W3)*v3);
```

[5]Notice that the point $(ILp).N$ is the point r shown in figure 7.2.

[6]This example will not work as shown because, to save space, the earlier examples in the text did not calculate the re-weighted hyperplanes W_i' and the re-weighted points v_i'. Once these elements are obtained, the example will work as shown.

```
> print_element, q_from_Z(ILp);
+5.000a3
```

This example has effectively shown how to calibrate the camera of a computer vision system. One sets up four reference points $v_0, \ldots, v_3$ at known positions in an object reference plane in front of the camera. One then measures the coordinates of the four corresponding image points in the graphics window. Using the theory of section 8.2, together with the example given in the current section, we can determine the collineation that maps both ways between the object reference plane and the graphics window. Then, given an arbitrary image point in the graphics window, one can determine the corresponding point in the object reference plane. In this way, measurements made on images in the graphics window can be mapped to measurements of physical objects in the object reference plane.

The camera can be calibrated for 3-d vision by using a pair of object reference planes. Assume that we have calibrated the camera on each object reference plane. Pick an arbitrary point w in the graphics window and use the collineations to determine the corresponding points p_1 and p_2 in the two object reference frames. The line p_1p_2 is the ray of light that passes through the object space of the camera and is imaged at the point w in the graphics window. As we vary the image point w we obtain a family of rays p_1p_2 in the object space of the camera. For an ideal camera, these rays will all intersect at a common point c, called the *perspective centre*, in the object space of the camera. In practice, they will intersect in a small region. Using least squares processing, one can determine a single best-fit point c. Now, given an image point w in the graphics window, we can use one of the collineations to determine the corresponding point p on one of the object reference planes. The line cp is then the path of the light ray, in the object space of the camera, that is imaged at point w in the graphics window. In this way, image points are mapped to known rays in the object space of the camera. This information can be used to make 3-d measurements using two or more cameras, or by moving the single camera.

8.4 Projective invariants

The method of camera calibration that has been described at the end of section 8.3 is the same as the one described in reference [23]. However, the description in [23] is in terms of *projective invariants*. A projective invariant is a number-valued function of points which is invariant under a collineation and is also invariant under arbitrary re-weighting of the points in the function[7]. This section explains how projective invariants arise in Whitehead's algebra.

From equation (8.10),

$$p = \sum_{i=1}^{r} (pA'_i) a'_i = (pA'_1) \left(a'_1 + \sum_{i=2}^{r} \frac{(pA'_i)}{(pA'_1)} a'_i \right) \ .$$

The weights under the summation sign are the projective invariants ρ_i. Using (8.8),

$$\rho_i = \frac{(pA'_i)}{(pA'_1)} = \frac{(pA_i)(a_0 A_1)}{(pA_1)(a_0 A_i)} \ .$$

Now substitute for the hyperplanes using (1.18),

$$\rho_i = \frac{(pa_1 \ldots \breve{a}_i \ldots a_n)(a_0 a_2 \ldots a_n)}{(pa_2 \ldots a_n)(a_0 a_1 \ldots \breve{a}_i \ldots a_n)} \ .$$

The points p and a_0 have no particular significance in this expression. We can just have a list of $n+2$ arbitrary points $a_1, \ldots, a_{n+2}$ and we can form projective invariants $\rho(i,j;k,l)$ labelled by the points which are omitted in each pseudonumber. Thus,

$$\rho_{i,j;k,l} = \frac{(a_1 \ldots \breve{a}_i \ldots \breve{a}_j \ldots a_{n+2})(a_1 \ldots \breve{a}_k \ldots \breve{a}_l \ldots a_{n+2})}{(a_1 \ldots \breve{a}_i \ldots \breve{a}_l \ldots a_{n+2})(a_1 \ldots \breve{a}_k \ldots \breve{a}_j \ldots a_{n+2})} \ . \tag{8.11}$$

As an example, let's create the five vertices $v_1, \ldots, v_5$ of a regular pentagon, project them into a graphics window as image points $w_1, \ldots, w_5$ and show that the projective invariant derived from the vertices of the pentagon is the same number when calculated using the object plane points and when calculated using the graphics world points. The first thing to do is to create the vertices of the regular pentagon and check that some of them make sense.

```
> v=w=array(complex,5,5);
> v(1,)=pcon(IL,a1,IN);
> for (j=2;j<=5;++j) v(j,)=con(1.0i*2.0*pi*(j-1.0)/5.0,wap(a1,a2),v(1,));
> print_element, q_from_p(v(1,));
+1.000a1
> print_element, q_from_p(v(2,));
+0.309a1+0.951a2
```

[7] In Whitehead's algebra, numbers are always invariant under collineations (section 1.2), so a projective invariant is just a number-valued homogeneous function (section 4.2.1) of points of degree zero.

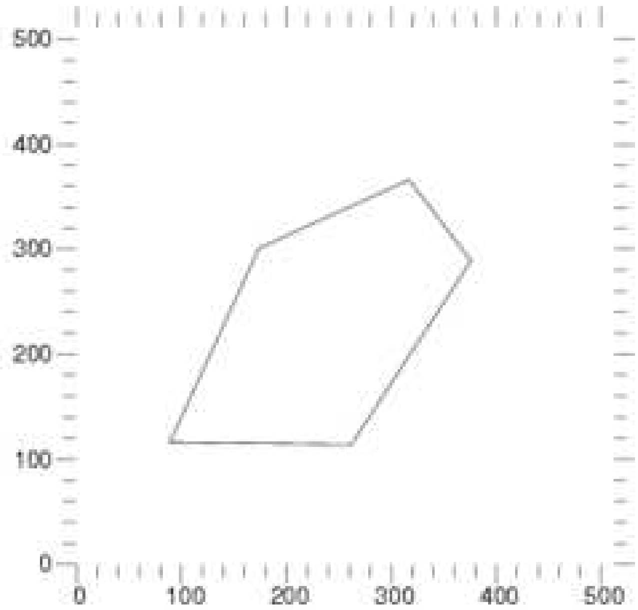

Figure 8.1: A regular pentagon distorted by a collineation

The next task is to define the collineation of the camera[8] by giving the four point correspondences between the object reference plane and the graphics window. The four points in the object reference plane are in the array `pRW` and the corresponding four points in the graphics window are in the array `pGW`. Here the letters `RW` and `GW` are meant to remind us that the points are defined with respect to the *robot world* or *graphics window* reference frames. In this case the point correspondences have been set up to give a distorted image of the regular pentagon in the graphics window as shown in figure figure 8.1.

```
> pRW=pGW=array(complex,4,5);
> pRW(1,)=pcon(IL,1.0*(-a1-a2),IN);pRW(2,)=pcon(IL,2.0*(a1-a2),IN);
> pRW(3,)=pcon(IL,3.0*(a1+a2),IN);pRW(4,)=pcon(IL,4.0*(-a1+a2),IN);
> pGW(1,)=IN;pGW(2,)=pcon(IL,512.0*a1,IN);
> pGW(3,)=pcon(IL,512.0*a1+512.0*a2,IN);pGW(4,)=pcon(IL,512.0*a2,IN);
```

The package `utilities.i` contains the function `project` that takes the point correspondences and sends a point in the object reference plane into the graphics window[9]. As usual, on-line documentation is available by typing `help, project`, and this states that the resulting point is not guaranteed to

[8]See section 8.3.

[9]There are a number of functions in the packages for projecting into the graphics window. The different versions project a single point, arrays of points, project and display

be on the quadric. Consequently, we can use the utility functions described in section 7.5.2 to get the points in whatever representation is required. In this case, `r_from_Z` is used to get position vectors for the points on the graphics window.

```
> for (j=1;j<=5;++j) w(j,)=r_from_Z(wap(IL,project(pGW,pRW,v(j,))));
```

We evaluate a projective invariant using formula (8.11) with the points from the regular pentagon on the object plane.

```
> ILa3v3v4v5=wap(IL,wap(a3,wap(v(3,),wap(v(4,),v(5,)))));
> ILa3v1v2v5=wap(IL,wap(a3,wap(v(1,),wap(v(2,),v(5,)))));
> ILa3v2v3v5=wap(IL,wap(a3,wap(v(2,),wap(v(3,),v(5,)))));
> ILa3v1v4v5=wap(IL,wap(a3,wap(v(1,),wap(v(4,),v(5,)))));
> print_element, ILa3v3v4v5*ILa3v1v2v5/(ILa3v2v3v5*ILa3v1v4v5);
> +0.618a1a2a3a4a5
```

The same number is obtained when the same projective invariant is evaluated using the points points of the pentagon in the graphics window. This demonstrates the use of projective invariants.

```
> ILa3w3w4w5=wap(IL,wap(a3,wap(w(3,),wap(w(4,),w(5,)))));
> ILa3w1w2w5=wap(IL,wap(a3,wap(w(1,),wap(w(2,),w(5,)))));
> ILa3w2w3w5=wap(IL,wap(a3,wap(w(2,),wap(w(3,),w(5,)))));
> ILa3w1w4w5=wap(IL,wap(a3,wap(w(1,),wap(w(4,),w(5,)))));
> print_element, ILa3w3w4w5*ILa3w1w2w5/(ILa3w2w3w5*ILa3w1w4w5);
> +0.618a1a2a3a4a5
```

The reason for getting the points in the graphics window in position vector form is because this example is also used to show how to display vertices in the graphics window. The function package `utilities.i` contains the function `plifs` which stands for *plot indexed face set*. This function plots an array of (x,y) coordinates in the graphics window. As explained at the end of section 7.5.2, the position vector output from `r_from_Z` can be directly

points, slice holes into a surface and project the result to the graphics window. Sometimes there are vectorized versions that take advantage of Yorick's array syntax to avoid for loops and work quickly for computer vision applications. In principle, they all work using the theory of section 8.2. However, when they were written, the writer followed reference [23]. This reference obscures the theory in section 8.2 by doing everthing in terms of projective invariants. Consequently, the projection functions could all be written more neatly by following the example in section 8.3. However, since the projection functions work, and the vectorized ones are reasonably fast, there is a certain amount of inertia to be overcome on the part of the writer before they are re-written. Of course, contributions from readers would be greatly appreciated.

used because the first two slots will be the (x,y) coordinates. The function plifs joins up the points in facets. In this case we have a single facet called shape containing all five vertices. The result is shown in figure 8.1.

```
shape=array(int,1,5);shape(1,)=[1,2,3,4,5];
fma;plifs, w,shape,"black";
```

Chapter 9

Theory of a hypersurface embedded in a Euclidean space

9.1 Introduction

This chapter develops the theory of a smooth hypersurface embedded in Euclidean space[1]. The $(n-2)$-d Euclidean space is modelled on the quadric of $(n-1)$-d elliptic space as in chapter 7. Consequently, the hypersurface has dimension n-3. The case of a smooth 2-d surface embedded in 3-d Euclidean space is obtained by setting $n = 5$. In this case, the theory will be illustrated by numerical examples using the Yorick functions introduced in chapters 7 and 8. Our main result will be a formula for the curvature operator.

9.2 The equation of a hypersurface

Section 4.2.1 has shown that a (n-2)-d hypersurface in a (n-1)-d projective space is defined by the equation $\phi(p) = 0$ where ϕ is a number-valued homogeneous function of a point p in the space. In the model of Euclidean geometry on the quadric, the (n-1)-d space is elliptic and the (n-2)-d hypersurface $\phi(p) = 0$ will intersect the (n-2)-d quadric $pIp = 0$ in a (n-3)-d subspace which will be a hypersurface embedded in the Euclidean space. Thus, in the model of Euclidean geometry on the quadric, the equation $\phi(p) = 0$ represents a hypersurface of the Euclidean space.

If the homogeneous function ϕ is of degree m then section 4.2.1 has shown that Euler's theorem on homogeneous functions is $p\partial\phi/\partial p = m\phi(p)$. The

[1] Equation (3.52) has shown that any flat-space geometry can be modelled on the absolute quadric. Consequently, with very little modification, the current chapter could be applied to a smooth hypersurface embedded in (say) a Minkowski space-time.

derivative $\partial\phi/\partial p$ is a hyperplane so we define,

$$F(p) = \frac{\partial\phi(p)}{\partial p} \ . \tag{9.1}$$

Notice that $F(p)$ is a map from points to hyperplanes so it is like a correlation that was studied in section 3.3. However, in general, $F(p)$ in (9.1) is not linear in p so it is not a correlation. The equation of the hypersurface is $0 = m\phi(p) = p.F(p)$ so that $F(p)$ is the tangent hyperplane to the hypersurface at p.

9.2.1 Example: A saddle-shaped surface

We can use a homogeneous polynomial of degree m,

$$\phi(p) = \sum_{\pi} c_\pi (pA_{\pi(1)}) \dots (pA_{\pi(m)}) \tag{9.2}$$

to define a hypersurface. Here, the A_j are reference hyperplanes and the c_π are numerical coefficients in which the subscript π denotes a selection of m indices from $1, 2, \dots, n$ in which the order of the indices is unimportant and copies are allowed[2]. For example, for $m = 5$ and $n = 7$, a particular selection could be $\pi = (1, 2, 2, 2, 4)$. It turns out that the number of selections is ${}^{n+m-1}C_m$.

A Yorick package `curvature.i` can be down-loaded from the url [21]. This package contains some functions for handling smooth surfaces in 3-d Euclidean space which are defined by quadratic polynomials. Since 3-d Euclidean space is modelled on the absolute quadric of $n = 5$ elliptic space, the number of coefficients for a quadratic $m = 2$ polynomial is ${}^{5+2-1}C_2 = 15$. These coefficients can be stored on the diagonal and upper triangle of a 5x5 matrix c_{ij}. So, the package `curvature.i` defines surfaces in terms of the quadratic polynomial,

$$\phi(p) = \sum_{1\le i\le j\le 5} c_{ij}(pE_i)(pE_j) \tag{9.3}$$

where the reference hyperplanes are $E_i = Ia_i$ for $i = 1, 2, 3$ and $E_4 = L$ and $E_5 = N$. The dual reference points are then $e_i = a_i$ for $i = 1, 2, 3$ with $e_4 = IN/(IN.L)$ and $e_5 = IL/(IN.L)$. It is easily verified that $e_iE_j = \delta_{ij}$ for $i, j = 1, \dots, 5$. The reason for using such a set of reference points and dual

[2]The homogeneous functions are not exhausted by the polynomials of the form of equation (9.2) because $\phi(x, y, z) = z^2 \ln(x/y)$ is homogeneous of degree 2 but it is not a polynomial

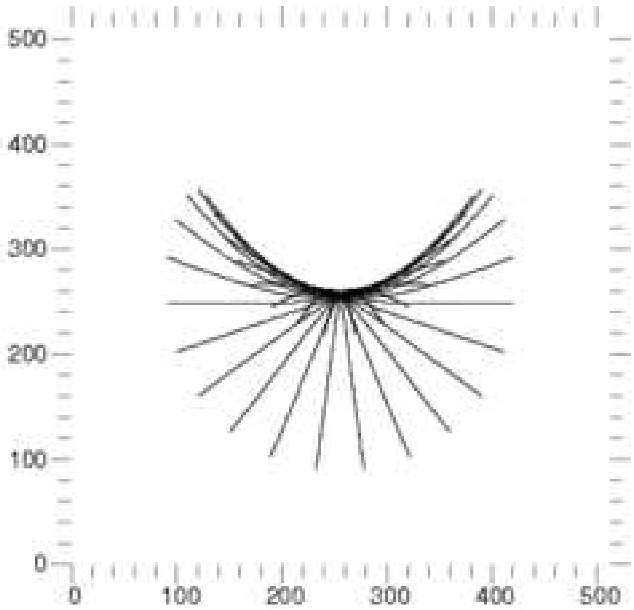

Figure 9.1: Saddle-shaped surface

hyperplanes instead of the standard ones a_i and A_i is so that when we fit the coefficients to a surface specified as an array of discrete points, we get real coefficients[3].

The saddle-shaped surface shown in figure 9.1 is formed[4] by setting all the coefficients c_{jk} to zero except for $c_{12} = -1/5$ and $c_{34} = 1$. Equation (9.3) is now $\phi = -(pA_1)(pA_2)/5 + (pA_3)(pL)$. Using equation (7.19) the point on the quadric can be written,

$$p = IN + (IN.L)\left(xa_1 + ya_2 + za_3 - \frac{(x^2+y^2+z^2)}{2}IL\right)$$

so that the polynomial is $\phi = (IN.L)^2(-xy/5 + z)$ and so the surface $\phi = 0$ is the saddle-shaped surface $z = xy/5$. In the following Yorick session[5], the saddle-shaped surface is created by setting up the coefficients c_{jk}. The function `homopoly1` evaluates equation (9.3) to return $\phi(p)$. We check that the point (1,2,2/5) is in the surface.

[3]For a point p on the quadric, pE_i is always a real number, but this would not be the case if we used the standard A_i as reference hyperplanes because then pA_5 would be imaginary. This would not be a problem except that matrix inversion routines in Yorick only work with real matrices.

[4]Figure 9.1 has been generated by starting from the origin IN which is on the surface, and drawing a set of geodesic curves in all directions in the surface.

[5]All the Yorick sessions used to illustrate the theory in this chapter will develop this saddle-shaped surface example in stages.

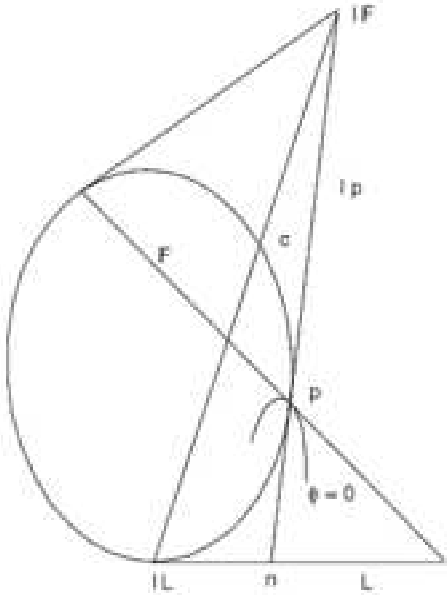

Figure 9.2: Normal vector to a hypersurface

```
> #include "curvature.i"
> c(1,2)=-0.2;
> c(3,4)=1.0;
> x=1.0;y=2.0;z=0.2*x*y;
> p=pcon(IL,x*a1+y*a2+z*a3,IN);
> homopoly1(p,c,E);
0+0i
```

9.3 The normal vector to a hypersurface

Figure 9.2 is a diagram of the hypersurface $\phi = 0$. The figure shows a point p at the intersection of the quadric and the hypersurface. The figure can only show one such point, but in fact these points form an (n-3)-d sub-manifold. The figure also shows the tangent hyperplane F at p and the point IF. The point n is the intersection of the line $p.IF$ with L. This point is the normal vector to the hypersurface at p. In order to see this, note that, according to section 7.4, F represents a hypersphere in the Euclidean space. This hypersphere is tangential to the hypersurface $\phi = 0$ at p. The centre of this tangent hypersphere is the point c shown in figure 9.2. The straight line in the Euclidean space through c and p is represented by $IL.c.p$

or $IL.p.n$ modulo a weight. Hence n is in the direction of the line from c to p. Now, if we think of the point p as an origin, then n lies in the subspace LIp that we have associated with vectors[6]. Therefore, the point n is the normal vector to the hypersurface at p. We shall work with the normalised point $n_1 = n/\sqrt{nIn}$ so that $n_1 I n_1 = 1$. The use of n_1 also avoids the confusion that might result because the normal vector n uses the same symbol as the number n of reference points in the projective space. So, the unit normal vector, or direction, is,

$$n_1 = \frac{(p.IF).L}{(pL)\sqrt{FIF}} = \frac{IF}{\sqrt{FIF}} - \frac{(IF.L)}{\sqrt{FIF}}\left(\frac{p}{pL}\right) \tag{9.4}$$

so that $n_1 I n_1 = 1$ by using $pF = 0$ and $pIp = 0$. Also, notice that $n_1 L = 0$ and $n_1 Ip = 0$.

9.3.1 Newton's method to solve for a point on a hypersurface

It is often necessary to solve $\phi(p) = 0$ to find the points p on the hypersurface. We start with a point p on the quadric which is a guess for a point on the hypersurface. Since p is only a guess, we have $\phi(p) \neq 0$ and we use Newton's method to iteratively improve the guess. In the case where p is on the hypersurface, then from figure 9.2, the line $(IL.IF.p).L$ is the invariant line of the parabolic congruence which moves p along the normal to the hypersurface. Consequently, if p is not on the hypersurface, but is close to it, the parabolic congruence can be expected to move a point approximately normal to the hypersurface. If the action of the congruence is denoted by $f_\tau p$, then,

$$f_\tau p = \exp(\tau ILIFp.L.I)p = p + \tau(IL.IF.p).(L.Ip)$$

and this new point will be on the hypersurface to first order if,

$$\begin{aligned} 0 &= \phi(p + \tau(IL.IF.p).(L.Ip)) = \phi(p) + \tau((IL.IF.p).(L.Ip)).\frac{\partial\phi(p)}{\partial p} \\ &= \phi(p) + \tau((IL.IF.p).(L.Ip.F)) \end{aligned}$$

so that,

$$p' = f_\tau p = p - \frac{\phi(p)(IL.IF.p).(L.Ip)}{(IL.IF.p).(L.Ip.F)} \tag{9.5}$$

[6]See, for example, section 7.3.5.

is a better guess for the point on the hypersurface. The function `homopoly3` in the package `curvature.i` does a few iterations of equation (9.5) to solve for points on a hypersurface.

In next stage of the saddle-shaped surface example, we use the function `homopoly3` to project the point (1,2,0) onto the surface at (0.866,1.933,0.335) and then check that this new point is really on the surface.

```
> p=pcon(IL,a1+2.0*a2,IN);
> pnew=homopoly3(p,c,E);
> print_element, q_from_p(pnew);
+0.866a1+1.933a2+0.335a3
> homopoly1(pnew,c,E);
-4.02239e-16+0i
```

9.4 Parallel-transport on a hypersurface

Parallel-transport is interesting because it provides the natural way to define the curvature of a manifold. In section 4.3.1 we studied parallel-transport in a homogeneous space with absolute polarity I. As we found in section 4.3.1, parallel-transport in a homogeneous space is neatly described by the action of the congruence group of the space. However, in a hypersurface $\phi = 0$, there is no analogous congruence group which can used to define parallel-transport.

In order to make progress, let us consider the restricted problem of how to parallel-transport a vector in the hypersurface. On numerous occasions[7] we have observed that a classical vector is always a point at infinity in Whitehead's algebra. For example, when we pick an origin point IN on the quadric, the generators q that live in the subspace LN are analogous to classical vectors. So, at a point p in the hypersurface $\phi(p) = pF = 0$ which is also on the quadric $pIp = 0$, a vector at p is a point q in the subspace $L.Ip.F$. The reason for specifying that q lies in F in addition to L and Ip is that the vector q should lie in the tangent space to the hypersurface at p. However, it turns out that the hyperplane In_1 is more important that the hyperplane F. From equation (9.4) it is clear that In_1 is a linear combination of F and Ip and so if q is in $L.Ip.F$ it is also in $L.Ip.In_1$. This is an important subspace, so we write,

$$l = L.Ip.In_1 \tag{9.6}$$

and l is the subspace containing the points q which are the tangent vectors to the hypersurface at p. Using the general projection formula (6.6), we can

[7]For example, see equation (7.4) and the associated text.

write q in terms of a part in the subspace l and another part in the dual subspace Il.

$$\begin{aligned} q &= \frac{(IL.p.n_1).(q.(L.Ip.In_1))}{(IL.p.n_1).(L.Ip.In_1)} - \frac{(q.IL.p.n_1).(L.Ip.In_1)}{(IL.p.n_1).(L.Ip.In_1)} \\ &= \frac{Il.(q.l)}{Il.l} - \frac{(q.Il).l}{Il.l}\ . \end{aligned} \tag{9.7}$$

However, q is in $l = L.Ip.In_1$ so this equation simplifies to,

$$q = -\frac{(q.Il).l}{Il.l}\ . \tag{9.8}$$

The rhs is the projection of q into l, but since q is already in l, the projection does not change q. The subspace $l = L.Ip.In_1$ depends on p because F and n_1 depend on p. Let's write $l = l(p)$ to remind ourselves of this fact. Now q is a vector at p in the sense that it lies in $l(p)$. The action of parallel-transport from p to $p + dp$ in the hypersurface will cause the vector q to change to $q + dq$ in the subspace $l(p + dp)$. Therefore,

$$q + dq = -\frac{(q.Il(p + dp)).l(p + dp)}{Il(p + dp).l(p + dp)} \tag{9.9}$$

is the equation for parallel-transport. Since equation (9.7) is a general projection formula it will hold for projecting q into $l(p+dp)$ and its dual $Il(p+dp)$.

$$q = \frac{Il(p + dp).(q.l(p + dp))}{Il(p + dp).l(p + dp)} - \frac{(q.Il(p + dp)).l(p + dp)}{Il(p + dp).l(p + dp)}$$

Combining this equation with (9.9) gives the following form of the equation for parallel-transport.

$$dq = -\frac{Il(p + dp).(q.l(p + dp))}{Il(p + dp).l(p + dp)}$$

Now $ql(p) = 0$ so that $ql(p + dp) = ql(p) + qdl = qdl$ and to first order we get,

$$dq = -\frac{Il.(q.dl)}{Il.l} \tag{9.10}$$

which is the canonical ordinary differential equation for parallel-transport of a vector in the hypersurface. An immediate consequence of this is that,

$$dq.Il = 0 \tag{9.11}$$

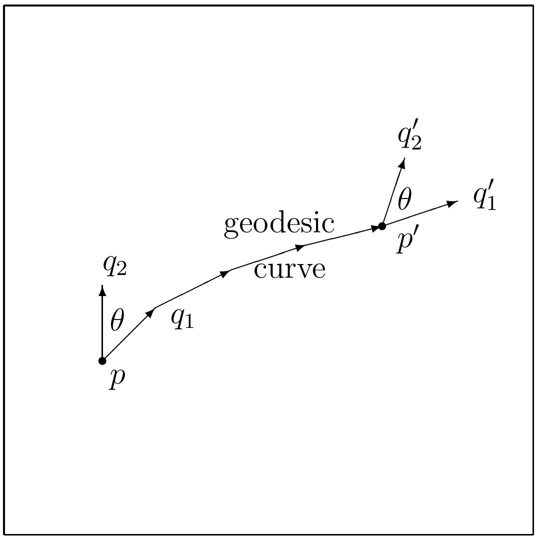

Figure 9.3: Parallel transport along a geodesic in the hypersurface

because from (9.10) dq lies in Il, so that[8] $dq.Il = 0$.

Another form of the parallel-transport equation is found by combining (9.9) with (9.8) to give,

$$\begin{aligned} dq &= (q+dq) - q = -\frac{(q.Il(p+dp)).l(p+dp)}{Il(p+dp).l(p+dp)} + \frac{(q.Il(p)).l(p)}{Il(p).l(p)} \\ &= -dp.\frac{\partial}{\partial p}\left(\frac{(q.Il(p)).l(p)}{Il(p).l(p)}\right) \end{aligned} \tag{9.12}$$

where the derivative acts only on p so that q is a constant as far as the derivative is concerned.

Let's now show that our equations for parallel-transport do sensible things. Firstly, the angle between two vectors q_1 and q_2 remains the same under

[8]For a formal proof of all arguments of this type, we can show that $(X_r.Y_s).X_r = 0$ where X_r and Y_s are subspaces of hyperplane grade r and s respectively and $r+s<n$ so that X_rY_s is the intersection of the two subspaces. If we have reference points $a_1,\ldots,a_n$ and dual hyperplanes $A_1,\ldots,A_n$ with $a_iA_j = \delta_{ij}$, then without loss of generality, since the a_i are arbitrary reference points, $X_r = A_1\ldots A_r$. Initially take $Y_s = A_{r+1}\ldots A_{r+s}$ so that $X_rY_s = A_1\ldots A_{r+s}$. If $r+s+r \le n$ then $(X_rY_s).X_r$ is an intersection which is zero by antisymmetry. Alternatively, if $r+s+r>n$ so that $(X_rY_s).X_r$ is a joining together of subspaces which is evaluated as a product of points,

$$(X_rY_s).X_r = \frac{a_{r+s+1}\ldots a_n}{a_1\ldots a_n}\frac{a_{r+1}\ldots a_n}{a_1\ldots a_n} = 0$$

by antisymmetry. For general Y_s, the same result follows by linearity. So, for the intersection of two subspaces $X_r.Y_s$ we always have $(X_r.Y_s).X_r = 0$.

parallel-transport along a path $p(\tau)$ in the hypersurface. To see this, first note that $q(p(\tau))l(p(\tau)) = 0$ for any vector q. The reason for this is that the rule (9.9) for parallel-transport has been chosen to project the vector q into the current subspace $l(p) = L.Ip.In_1(p)$ containing the tangent vectors at p. So, the current vector $q(p(\tau))$ is always in $l(p(\tau))$. Now since $q_i Ip = 0$, the angle between the lines pq_1 and pq_2 is given by equation (3.37) as[9],

$$\cos^2(\theta) = \frac{(q_1 I q_2)^2}{(q_1 I q_1)(q_2 I q_2)} \ . \tag{9.13}$$

Under parallel transport the change in the numerator is,

$$\begin{aligned} d(q_1 I q_2) &= dq_1.Iq_2 + q_1.Idq_2 = -\frac{(Il.(q_1.dl)).Iq_2}{Il.l} - \frac{(Il.(q_2.dl)).Iq_1}{Il.l} \\ &= -(-1)^{2.1}\frac{(Il.Iq_2.(q_1.dl))}{Il.l} - (-1)^{2.1}\frac{(Il.Iq_1.(q_2.dl))}{Il.l} \\ &= -(-1)^{n-3}\frac{(I(q_2 l).(q_1.dl))}{Il.l} - (-1)^{n-3}\frac{(I(q_1 l).(q_2.dl))}{Il.l} = 0 + 0 = 0 \end{aligned}$$

where we used equation (9.10) and the fact that the factors are associative and $q_i l = 0$ along the path. Therefore $q_1 I q_2$ is constant along the path in the hypersurface and by the same argument, $q_1 I q_1$ and $q_2 I q_2$ are both constant along the path. Therefore, from (9.13), the angle θ between the two vectors is preserved under parallel-transport along the path. This is the fact that shows that our equations for parallel-transport are doing what we intuitively understand by parallel-transport. For, in a homogeneous space with polarity I, figure 4.1 showed that parallel-transport of a point along a straight line preserved the angle between the point and the line. In general, straight lines are not contained within the hypersurface. However, the analog of a straight line in the hypersurface is a *geodesic* path, which is made by parallel-transporting the tangent vector to the path parallel to itself. This is shown in figure 9.3 where the tangent vector to the path is q_1. From our result that the angle between two tangent vectors is preserved under parallel-transport, we see that figure 9.3 is analogous to figure 4.1 so that equation (9.10) does capture what we intuitively mean by parallel-transport.

9.4.1 The parallel-transport equations in detail

In this section we take equation (9.10) for parallel-transport and substitute for the tangent subspace from (9.6) in order to get an explicit ordinary dif-

[9]Since we are in elliptic space, the general parameter τ in equation (3.37) can be replaced by $i\theta$ where θ is the real angle.

ferential equation. We need,

$$Il.l = (IL.p.n_1).(L.Ip.In_1) = -(-1)^{n-1}(pL)^2 \tag{9.14}$$

and,

$$\begin{aligned}
q.\frac{dl}{d\tau} &= q.\left(L.I\frac{dp}{d\tau}.In_1\right) + q.\left(L.Ip.I\frac{dn_1}{d\tau}\right) \\
&= -\left(q.I\frac{dp}{d\tau}\right)LIn_1 + \left(q.I\frac{dn_1}{d\tau}\right)LIp
\end{aligned}$$

where we used $qL = 0$, $qIp = 0$ and $qIn_1 = 0$. Furthermore,

$$\begin{aligned}
Il.\left(q.\frac{dl}{d\tau}\right) &= -\left(q.I\frac{dp}{d\tau}\right)(IL.p.n_1).LIn_1 + \left(q.I\frac{dn_1}{d\tau}\right)(IL.p.n_1).LIp \\
&= -\left(q.I\frac{dp}{d\tau}\right)(((IL.p).LIn_1)n_1 - ((IL.n_1).LIn_1)p + (pn_1.LIn_1)IL) \\
&+ \left(q.I\frac{dn_1}{d\tau}\right)(((IL.p).LIp)n_1 - ((IL.n_1).LIp)p + (pn_1.LIp)IL) \\
&= -\left(q.I\frac{dp}{d\tau}\right)(pL)IL - \left(q.I\frac{dn_1}{d\tau}\right)(-1)^{n-1}(pL)^2 n_1
\end{aligned}$$

so that,

$$\frac{dq}{d\tau} = -\left(\frac{(-1)^{n-1}}{pL}\right)\left(q.I\frac{dp}{d\tau}\right)IL - \left(q.I\frac{dn_1}{d\tau}\right)n_1 \tag{9.15}$$

is the detailed form of the equation of parallel-transport.

Although (9.15) is the equation for parallel-transport that we shall use for most of the work in this chapter, we also find a use for (9.12). This version of the parallel-transport equation is used to evaluate the multiplicative integral for parallel-transport around a small loop in the hypersurface on the quadric in section 9.9.1. In order to prepare the way for this calculation, we now obtain the expanded form of (9.12) and check that it is equivalent to (9.15). From (9.12) we need to evaluate $(q.Il).l/(Il.l)$ in the case where $p = p(\tau)$ but q is kept fixed so that it is no longer in $l[p(\tau)] = L.Ip(\tau).In_1[p(\tau)]$. So, in our calculations, we *do not* assume that $qIp = 0$ and $qIn_1 = 0$ but we still have $qL = 0$ because L is a constant and not dependent on the path $p(\tau)$.

$$\begin{aligned}
(q.Il).l &= (q.IL.p.n_1).(L.Ip.In_1) \\
&= [(q.IL.p).(L.Ip.In_1)]n_1 - [(q.IL.n_1).(L.Ip.In_1)]p \\
&\quad +[(q.p.n_1).(L.Ip.In_1)]IL - [(IL.p.n_1).(L.Ip.In_1)]q \\
&= -(qIn_1)(Lp)(pL)n_1 - 0 - (qIp)(pL)IL + (Lp)(pL)q \\
&= -(-1)^{n-1}(qIn_1)(pL)^2 n_1 - (qIp)(pL)IL + (-1)^{n-1}(pL)^2 q
\end{aligned}$$

Upon dividing the above expression by $Il.l$ from (9.14), we find that (9.12) expands to,

$$dq = dp\frac{\partial}{\partial p}\left[q - (qIn_1)n_1 - \frac{(-1)^{n-1}(qIp)IL}{pL}\right]$$

where q is constant under $\partial/\partial p$. Hence, $\partial/\partial p$ only acts on p and n_1 in the square bracket. Upon dividing by $d\tau$ we get,

$$\frac{dq}{d\tau} = -\frac{d}{d\tau}\left[(qIn_1)n_1 + \frac{(-1)^{n-1}(qIp)IL}{pL}\right] \tag{9.16}$$

and the $d/d\tau$ only acts on p and n_1 in the square bracket.

We need to check that (9.16) is equivalent to (9.15).

$$\frac{dq}{d\tau} = -\left(qI\frac{dn_1}{d\tau}\right)n_1 - (qIn_1)\frac{dn_1}{d\tau} - (-1)^{n-1}\left[qI\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right]IL$$

Now we take this as a differential equation for q. Hence q is updated along the path so now $qIn_1 = 0$ and $qIp = 0$ for all τ. Hence we get,

$$\frac{dq}{d\tau} = -\left(qI\frac{dn_1}{d\tau}\right)n_1 - \frac{(-1)^{n-1}}{pL}\left(qI\frac{dp}{d\tau}\right)IL$$

which is exactly the same as (9.15). Therefore we can be confident that (9.16) is a sensible formula for parallel-transport and is ready for service in the multiplicative integral in section 9.9.1.

9.5 Definition of curvature along a path

From duality[10], the distance between vectors q and $q + dq$, which are always directions, is the same as the angle between the directions[11]. So, from

[10] See section 3.8.

[11] Some clarification may be necessary here. A translation in the Euclidean space is produced by a parabolic congruence with invariant line $IL.q$ where q lies in L. The invariant line $IL.q$ may be thought of as the direction of the translation. In other words, a direction in Euclidean space is represented by a parabolic line $IL.q$ which lies in the hyperplane L so $qL = 0$. Now let's consider two directions represented by $IL.q_1$ and $IL.q_2$ and determine the angle between these directions. Take any point p on the quadric. The straight lines through p in the specified directions are $IL.p.q_1$ and $IL.p.q_2$. The angle between these lines is the parameter of the congruence which takes the first line into the second. If q_1 and q_2 both lie on Ip, then the invariant line of the congruence is q_1q_2. This is because the subspace in which this congruence acts like the identity is Iq_1q_2 and IL

equation (3.35) the angular change in the direction is,

$$-id\theta = \frac{\sqrt{-qdq.Iqdq}}{qIq} = \frac{\sqrt{-(qIq)(dqIdq) + (qIdq)^2}}{qIq}$$

and the factor $i = \sqrt{-1}$ is because we are in elliptic space and the minus sign has been chosen to make the curvature of a hypersphere positive[12]. Now section 9.4 has shown that qIq is invariant under parallel-transport so $0 = d(qIq) = 2qIdq$. Hence $qIdq = 0$, a fact which can also be obtained directly from equation (9.15). Therefore, the angular change in the direction under parallel-transport is,

$$d\theta = -\sqrt{\frac{dqIdq}{qIq}} \ .$$

The path in the hypersurface is $p(\tau)$ and we define the curvature k along the path as,

$$k = \frac{d\theta}{d\tau} = -\sqrt{\frac{\frac{dq}{d\tau} I \frac{dq}{d\tau}}{qIq}} \ . \tag{9.18}$$

Here, τ is an arbitrary parameter along the path in the hypersurface. Usually, curvature is defined with respect to the parameter of arc-length, but we keep τ general for a while. Using the equation of parallel-transport (9.15) in (9.18) gives,

$$k = \frac{d\theta}{d\tau} = -\frac{\left(q.I\frac{dn_1}{d\tau}\right)}{\sqrt{qIq}} \ . \tag{9.19}$$

---

and p are both in this subspace. If the parameter of the congruence is adjusted so that q_1 is moved to q_2 then this congruence takes the first line into the second line and the angle between the lines is the congruence parameter. This parameter is also the distance between q_1 and q_2. In the general case in which q_1 and q_2 do not lie on Ip, we just have to find the points at the intersection of the lines ILq_1 and ILq_2 with Ip. The angle between the lines is then the same as the distance between these intersection points. The points of intersection are,

$$ILq_i.Ip = (-1)^{n-1}(pL)q_i - (q_iIp)IL$$

and from equation (3.31), if we set $a = ILq_1.Ip$ and $b = ILq_2.Ip$, the distance/angle is,

$$\cos^2(\theta) = \frac{(q_1Iq_2)^2}{(q_1Iq_1)(q_2Iq_2)} \ . \tag{9.17}$$

This equation shows that the angle between two directions given by the parabolic lines ILq_1 and ILq_2 can be obtained by just calculating the distance between q_1 and q_2 where q_1 and q_2 are any representative points on the lines which specify the directions.

[12] See equation (9.22).

9.6 The equation of a geodesic

In general we parallel-transport a direction q along a path $p(\tau)$ in the hypersurface. However, as shown in figure 9.3, if we parallel-transport a tangent to a curve parallel to itself, the result is a geodesic curve - analog of a straight line - in the hypersurface. It is tempting to take the tangent to a curve as $q = dp/d\tau$ but this does not work because, according to section 9.4, we require $qL = 0$ and this is not the case for an arbitrary parameterization of the curve. However, if we take the tangent to the curve as,

$$q = \frac{d}{d\tau}\left(\frac{p}{pL}\right) = \frac{1}{pL}\frac{dp}{d\tau} - \frac{p}{(pL)^2}\frac{d(pL)}{d\tau} \tag{9.20}$$

then this works because $qL = 0$. We also require $qIp = 0$ and $qIn_1 = 0$. The condition $qIp = 0$ is easily seen to be satisfied by differentiating $pIp = 0$. In order to show that the condition $qIn_1 = 0$ is satisfied, start from equation (9.4),

$$qIn_1 = \frac{(-1)^{n-1}qF}{\sqrt{FIF}} - \frac{(IF.L)}{\sqrt{FIF}}\left(\frac{qIp}{pL}\right)$$

and since $qIp = 0$ we only have to demonstrate that $qF = 0$. From (9.20) and (9.1),

$$qF = \left(\frac{1}{pL}\frac{dp}{d\tau} - \frac{p}{(pL)^2}\frac{d(pL)}{d\tau}\right)\frac{\partial\phi(p)}{\partial p}\ .$$

Now using (4.14) and (4.12),

$$qF = \left(\frac{1}{pL}\frac{d\phi(p)}{d\tau} - \frac{m\phi(p)}{(pL)^2}\frac{d(pL)}{d\tau}\right)$$

and both terms are zero because $\phi(p) = 0$ on the hypersurface. So, we have shown that the tangent to the curve defined by (9.20) is in the tangent subspace $l = L.Ip.In_1$ of equation (9.6). We can now transport the tangent q parallel to itself using equation (9.15). Using (9.20), we just need

$$-\frac{(-1)^{n-1}}{pL}\left(q.I\frac{dp}{d\tau}\right) = -(-1)^{n-1}qI\left(q + \frac{p}{(pL)^2}\frac{d(pL)}{d\tau}\right) = -(-1)^{n-1}qIq$$

so that the equation of a geodesic is,

$$\begin{aligned}\frac{dq}{d\tau} &= -(-1)^{n-1}(qIq)IL - \left(q.I\frac{dn_1}{d\tau}\right)n_1 \\ &= -(-1)^{n-1}(qIq)IL + k(qIq)^{1/2}n_1 \end{aligned} \tag{9.21}$$

where the curvature k along a path has been inserted from (9.19). Note that qIq has been shown to be constant[13] in section 9.4 and that the geodesic equation is valid for an arbitrary parameterisation τ.

9.6.1 Arc-length as the parameter of the path

From equation (3.49), the squared element of arc-length along the path is,

$$\begin{aligned}
(d\tau)^2 &= -2\left(\frac{p(\tau)}{p(\tau)L}\right) I \left(\frac{p(\tau+d\tau)}{p(\tau+d\tau)L}\right) \\
&= -2\left(\frac{p}{pL}\right) I \left(\frac{p}{pL} + q d\tau + \frac{(d\tau)^2}{2}\frac{dq}{d\tau}\right) = -(d\tau)^2\left(\frac{p}{pL}\right) I \left(\frac{dq}{d\tau}\right) \\
&= -(d\tau)^2 \frac{d}{d\tau}\left(\left(\frac{p}{pL}\right) Iq\right) + (d\tau)^2(qIq) = (d\tau)^2(qIq)
\end{aligned}$$

where we expanded in a Taylor series and substituted the tangent from (9.20) and differentiated by parts. Upon dividing both sides by $(d\tau)^2$ we get $qIq = 1$ if the path parameter τ is arc-length.

9.6.2 Example: Curvature of a hypersphere

From section 7.4, the equation of a hypersphere is $\phi(p) = pS = 0$ where S is the hyperplane which represents the hypersphere. From (9.1) the hyperplane $F = S$ and it is constant independent of the point p. Equation (9.4) can be used to work out the derivative of the normal vector along a path in the hypersphere and then the curvature can be evaluated using equation (9.19). If we use arc-length to parameterise the path so that $qIq = 1$, the curvature of the hypersphere is,

$$k = -qI\frac{dn_1}{d\tau} = \frac{(IF.L)(qIq)}{\sqrt{FIF}} = \frac{(IS.L)}{\sqrt{(-1)^{n-1}(IS.S)}} \tag{9.22}$$

and from equation (7.17) this is seen to be the same as the reciprocal of the radius of the hypersphere.

9.6.3 Another formula for curvature along a path

As it stands, equation (9.19) gives the curvature along a path in terms of $dn_1/d\tau$. We can get another formula for the curvature by substituting for n_1

[13]This can also be seen by multiplying (9.21) through by Iq on the right. This gives $dq/d\tau.Iq = 0$ which shows that qIq is a constant.

from equation (9.4). To simplify the calculations, we use $F_1 = F/\sqrt{FIF}$ as the normalised version of F so that $F_1 I F_1 = 1$.

$$
\begin{aligned}
qI\frac{dn_1}{d\tau} &= (-1)^{n-1}q.\frac{dF_1}{d\tau} - (IF_1.L)(qIq) \\
&= (-1)^{n-1}q.\frac{dp}{d\tau}\frac{\partial F_1}{\partial p} - (IF_1.L)(qIq) \\
&= (-1)^{n-1}q.\left(\frac{d\ln(pL)}{d\tau}p + (pL)q\right)\frac{\partial F_1}{\partial p} - (IF_1.L)(qIq)
\end{aligned}
$$

In the above equation we used (9.20) to express $dp/d\tau$ in terms of q. Now the term $p\partial F_1/\partial p$ will be proportional to F_1 by Euler's theorem on homogeneous functions[14] and $qF_1 = 0$ so that the formula for the curvature is,

$$
\begin{aligned}
k &= (IF_1.L)(qIq)^{1/2} - \frac{(-1)^{n-1}(pL)}{(qIq)^{1/2}}q.\left(q\frac{\partial F_1}{\partial p}\right) \\
&= \frac{(IF.L)(qIq)^{1/2}}{(FIF)^{1/2}} - \frac{(-1)^{n-1}(pL)}{(FIF)^{1/2}(qIq)^{1/2}}q.\left(q\frac{\partial F}{\partial p}\right) . \qquad (9.24)
\end{aligned}
$$

The interesting thing about (9.24) is that the curvature of the hypersurface in direction q is composed of two parts,

$$k = k_{\text{sphere}} + k_{\text{bare}} \qquad (9.25)$$

where,

$$
\begin{aligned}
k_{\text{sphere}} &= \frac{(IF.L)(qIq)^{1/2}}{(FIF)^{1/2}} \qquad (9.26) \\
k_{\text{bare}} &= -\frac{(-1)^{n-1}(pL)\left(q.q\frac{\partial F}{\partial p}\right)}{(FIF)^{1/2}(qIq)^{1/2}} . \qquad (9.27)
\end{aligned}
$$

[14] Although Euler's theorem was only given for number-valued functions in equation (4.12), it is still valid for hyperplane-valued functions like $F_1 = F_1(p)$. In fact, the original argument in section 4.2.1 applies to any linear-subspace-valued function $X(p)$ because there is no physical significance to the weight of a point. Hence $X(\lambda p)$ is the same subspace as $X(p)$ and so $X(\lambda p) = \mu(\lambda, p)X(p)$ where $\mu(\lambda, p)$ is a weight-valued function. This is a slightly more general argument than the one given in section 4.2.1 where it was assumed that $\mu = \lambda^m$. Differentiating with respect to λ,

$$\frac{\partial X(\lambda p)}{\partial \lambda} = \frac{\partial \lambda p}{\partial \lambda}\frac{\partial X(\lambda p)}{\partial \lambda p} = p\frac{\partial X(\lambda p)}{\partial \lambda p} = \frac{\partial \mu(\lambda, p)}{\partial \lambda}X(p)$$

and setting $\lambda = 1$ gives a generalised version of Euler's theorem on homogeneous functions,

$$p\frac{\partial X(p)}{\partial p} = \frac{\partial \mu(\lambda = 1, p)}{\partial \lambda}X(p) . \qquad (9.23)$$

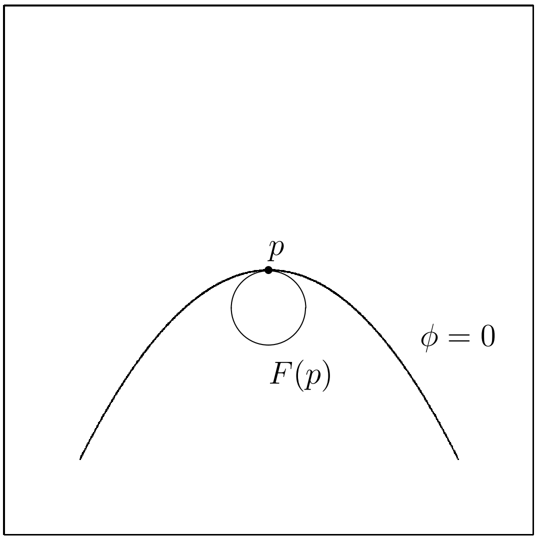

Figure 9.4: How the ϕ and F appear in Euclidean space

If we take $qIq = 1$ so that the path is parameterised by arc-length, and compare k_{sphere} with (7.17) for the radius of a hypersphere, we find that k_{sphere} is 1/radius of a hypersphere given by F. The meaning of F can be understood from figure 9.2 which shows that F is the tangent hyperplane to the hypersurface $\phi = 0$ at p. The figure shows that the hyperplane F cuts the absolute quadric. Section 7.4 shows that the (n-2)-dimensional hyperplane F cuts the quadric hypersurface. The points at which F intersects the quadric represent an (n-3)-dimensional hyperspherical surface in the Euclidean space on the quadric. This is also shown in figure 7.1 where a hyperplane S cuts the quadric and represents a hypersphere in the Euclidean space. The hyperplanes S in figure 7.1 and F in figure 9.2 play the same rôle of representing a hypersphere in the Euclidean space.

Figure 9.4 shows how things appear in Euclidean space. The equation $\phi(p) = 0$ defines a hypersurface within the Euclidean space. At each point p on the hypersurface we have a tangential hypersphere. So, for $n = 5$, we have 3-d Euclidean space. The equation $\phi(p) = 0$ defines a 2-d surface in the Euclidean space and at each point p on the surface there is a 2-d spherical surface which touches the 2-d surface and is tangential to it at the point of contact.

The term k_{bare} in equation (9.25) is the correction that has to be applied to the curvature of the tangent hypersphere F to get the true curvature k along the path. Notice that if the hypersphere F degenerates into a hyperplane (in the Euclidean space) then F goes through the point at infinity IL

and so $IL.F = IF.L = 0$. Then $k_{\text{sphere}} = 0$ and the bare curvature is the same as the true curvature $k = k_{\text{bare}}$.

9.6.4 Numerical integration of the geodesic equation

The curves in figure 9.1 are the geodesics passing through the origin IN on the saddle-shaped surface used as an example in section 9.2.1. These geodesics were obtained by numerically integrating equations (9.21,9.20) using Yorick's `rkutta.i` Runge-Kutta ODE integration package. The derivative $dy/d\tau$ needed by Yorick's `rk_integrate` function is provided by the `deriv425` function in the package `curvature.i`. The state is essentially the pair $y = (p/(pL), q)$ and so the derivative is $dy/d\tau = (q, dq/d\tau)$. The $dq/d\tau$ part is given by the geodesic equation (9.21) with the unit normal vector n_1 given by (9.4) and the curvature k given by (9.24) which was derived from (9.19).

In order to keep everything real, we work with the reference points $e_1, \dots, e_5$ and dual hyperplanes $E_1, \dots, E_5$ that were defined in section 9.2.1. These reference points and dual hyperplanes are defined inside the package `curvature.i`. The state has 10 components $y_1, \dots, y_{10}$ with,

$$\begin{aligned} p &= y_1e_1 + y_2e_2 + y_3e_3 + y_4e_4 + y_5e_5 \\ q &= y_6e_1 + y_7e_2 + y_8e_3 + y_9e_4 + y_{10}e_5 \ . \end{aligned}$$

We would like to work with $p/(pL)$ as the first part of the state, but this is awkward because we need pL in (9.24) so that $p/(pL)$ as a single entity is not compatible with a straighforward calculation of k. In order to overcome this problem we assume $pL = 1$ for all τ by setting $pL = pE_4 = y_4 = 1$. So, with $y_4 = 1$, equation (9.20) is $q = dp/d\tau$. Hence $y_9 = qL = dpL/d\tau = 0$. The state must be set so that $y_4 = 1$ and $y_9 = 0$. If these conditions are set up at $\tau = 0$ then the differential equation ensures that they continue to hold for all τ.

The following Yorick program was used to generate figure 9.1. The program uses Yorick's `rkutta.i` package together with the derivative function `deriv425` from the `curvature.i` package to integrate the geodesic equation. Geodesics are launched in directions with 10 degree increments from the origin IN. In addition the program sets up a simulated camera using the techniques of chapter 8 to project the geodesics on the saddle-shaped surface into the graphics window.

```
require, "vision.i";
require, "curvature.i";
```

```
require, "rkutta.i";
//Make up our own camera calibration data with a short focal length.
//Here the object reference plane is at z=0 in the TCP frame and is an
//square of side 20mmx20mm which is mapped onto a 512x512 pixel graphics
//world. The perspective centre is set 20mm above the object ref plane.
zero=array(complex,5);
pTCP=pGW=array(complex,4,5);
pTCP(1,)=pcon(IL,10.0*(-a1-a2),IN);pTCP(2,)=pcon(IL,10.0*(a1-a2),IN);
pTCP(3,)=pcon(IL,10.0*(a1+a2),IN);pTCP(4,)=pcon(IL,10.0*(-a1+a2),IN);
pGW(1,)=IN;pGW(2,)=pcon(IL,512.0*a1,IN);
pGW(3,)=pcon(IL,512.0*a1+512.0*a2,IN);pGW(4,)=pcon(IL,512.0*a2,IN);
cTCP=pcon(IL,20.0*a3,IN);//Adjust this to change perspective
limits, 0, 512, 0, 512;//Limits set to 512 x 512 image
fma;
//Position the camera using Euler angles roll, yaw.
roll=70.0*pi/180.0;yaw=45.0*pi/180.0;
RW_TCP=Donkin(transpose([yaw*a3,zero]),
               transpose([roll*a1,zero]));
TCP_RW=reverse(RW_TCP);TCP_RW(2,)=-10.0*a3;//Distance of camera
//----------------- Geodesics through IN -------------------
c(1,2)=-0.2;c(3,4)=1.0;//Set up the homogeneous polynomial
p=IN;F=homopoly2(p,c,E);
l=wap(L,wap(I(p),F));//Tangent space at p
q1=wap(wap(a1,I(l)),l);q1=q1/sqrt(wap(q1,I(q1)));
q2=wap(l,I(q1));q2=q2/sqrt(wap(q2,I(q2)));
t=span(0.0,10.0,41);path=array(complex,dimsof(t)(2),5);
y1=array(complex,10);
for (theta=0.0;theta<360;theta+=10.0){
  q=con(i*theta*pi/180.0,wap(q1,q2),q1);
  y1(1:5)=wap2(4,E,p)/2.0;y1(6:10)=wap2(4,E,q);
  y=rk_integrate(deriv425,y1,t,0.001,0.1);//Integrate ODE
  for (j=1;j<=dimsof(t)(2);++j) path(j,)=y(1:5,j)(+)*e(+,);
  chain=array(long,dimsof(t)(2)-1,2);
  for (j=1;j<=dimsof(chain)(2);++j) chain(j,)=[j,j+1];
  display_target,path,chain,TCP_RW,pTCP,pGW,cTCP,"black";
}
```

9.7 Extremal curvatures

The curvature k at a point p in the hypersurface along a path with tangent q is given by equation (9.24). As the next stage in the saddle-shaped surface example, we evaluate the curvature in different directions. In the following Yorick session, we set the point as the origin and check that this point is on the surface. Then we compute F, FIF, $IF.L$ and pL which are needed to evaluate (9.24).

```
> p=IN;
> homopoly1(p,c,E);
0+0i
> F=homopoly2(p,c,E);
> FIF=wap(F,I(F));
> IF_L=wap(I(F),L);
> pL=wap(p,L);
```

Next we find the tangent space $l = L.Ip.F$ which turns out to be the line $a_1 a_2$.

```
> l=wap(L,wap(I(p),F));
> print_element, l;
+4.000ia1a2
```

We set up tangent $q = (a_1 + a_2)/\sqrt{2}$ such that $qIq = 1$ so that the path parameter is arc-length and evaluate the curvature k from (9.24) in the this direction. The hyperplane $q\partial F/\partial p$ is needed for the calculation. The `curvature.i` package defines surfaces in terms of quadratic polynomials given by (9.3), so that,

$$F(p) = \frac{\partial \phi}{\partial p} = \sum_{ij} c_{ij} E_i (pE_j) + \sum_{ij} c_{ij} (pE_i) E_j \tag{9.28}$$

$$q\frac{\partial F(p)}{\partial p} = \sum_{ij} c_{ij} E_i (qE_j) + \sum_{ij} c_{ij} (qE_i) E_j = F(q) \ . \tag{9.29}$$

```
> q=(a1+a2)/sqrt(2.0);
> qdFbydp=homopoly2(q,c,E);
> k=IF_L/sqrt(FIF)-pL*wap(q,qdFbydp)/sqrt(FIF);
> k
0.2+0i
```

The curvature in the direction $q = (a_1 - a_2)/\sqrt{2}$ is also evaluated[15].

[15] These values of the curvature $k = \pm 0.2$ are the same as the values that would be obtained for the saddle-shaped surface $z = 0.2xy$ using the methods of classical differential geometry as given (say) in chapter 19 of [14].

```
> q=(a1-a2)/sqrt(2.0);
> qdFbydp=homopoly2(q,c,E);
> k=IF_L/sqrt(FIF)-pL*wap(q,qdFbydp)/sqrt(FIF);
> k
-0.2+0i
```

In the above example, the directions were chosen to be those of the maximum and minimum curvatures for the saddle-shaped surface at the given point. Instead of evaluating the curvature along different directions in the hypersurface at a given point in order to find the extremal curvatures, it is interesting to solve the problem directly. This leads to an eigenvalue problem for the extremal curvatures. The operator that occurs in the eigenvalue problem can be regarded as a *curvature operator* that summarises the curvature properties of the hypersurface at the given point[16]. However, before being able to set up the eigenvalue problem for the extremal curvatures, we make a digression to formulate the method of Lagrange multipliers in the language of Whitehead's algebra.

9.7.1 Lagrange multipliers

The problem is to find an extremum of $\alpha(p)$, subject to constraints $\phi_\mu(p) =$ constant for $\mu = 1, \ldots, m$ where $\alpha(p)$ and the $\phi_\mu(p)$ are number-valued functions of a point. In the method of Lagrange multipliers, this problem is solved by seeking an *unconstrained* extremum of,

$$\alpha(p) - \sum_{\mu=1}^{m} \lambda_\mu \phi_\mu(p)$$

where the λ_μ are numbers called Lagrange multipliers. The unconstrained extremum obeys,

$$0 = \frac{\partial}{\partial p}(\alpha - \sum_\mu \lambda_\mu \phi_\mu) = \frac{\partial \alpha}{\partial p} - \sum_\mu \lambda_\mu \frac{\partial \phi_\mu}{\partial p}$$

so that,

$$\frac{\partial \alpha}{\partial p} = \sum_\mu \lambda_\mu \frac{\partial \phi_\mu}{\partial p} \tag{9.30}$$

at the solution point p. Now consider the variation $\delta\alpha$ as $p \to p + \delta p$. From (9.30),

$$\delta\alpha = \delta p . \frac{\partial \alpha}{\partial p} = \delta p . \sum_\mu \lambda_\mu \frac{\partial \phi_\mu}{\partial p} = \sum_\mu \lambda_\mu \delta p . \frac{\partial \phi_\mu}{\partial p} = \sum_\mu \lambda_\mu \delta \phi_\mu$$

[16]In Whitehead's algebra, the curvature operator replaces the *fundamental magnitudes of the second order* that occur in classical differential geometry [14].

so that if δp is a variation that obeys the constraints $\phi_\mu = \text{const}$ then $\delta\phi_\mu = 0$ and we have $\delta\alpha = 0$ for variations $p \to p + \delta p$ that obey the constraints. So, Lagrange's method finds a point p such that $\delta\alpha = 0$ for variations that obey the constraints. In other words, it is a solution to the problem of finding an extremum of α subject to the constraints. This explains why the method of Lagrange multipliers works.

We can solve for the Lagrange multipliers in (9.30) by multiplying by the hyperplane product,

$$(-1)^{\nu-1}\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\check{\phi}_\nu}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}$$

where the "ˇ" denotes a missing factor. By antisymmetry, only one term in the summation survives.

$$(-1)^{\nu-1}\frac{\partial\alpha}{\partial p}\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\check{\phi}_\nu}{\partial p}\cdots\frac{\partial\phi_m}{\partial p} = \lambda_\nu\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}$$

The Lagrange multiplier can be obtained by using the polarity I.

$$\begin{aligned}(-1)^{\nu-1}I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)&\cdot\left(\frac{\partial\alpha}{\partial p}\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\check{\phi}_\nu}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\\ &= \lambda_\nu I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\end{aligned}$$

The Lagrange multipliers are,

$$\lambda_\mu = \frac{(-1)^{\mu-1}I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\alpha}{\partial p}\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\check{\phi}_\mu}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)}{I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)}\,. \tag{9.31}$$

Now keep (9.31) and (9.30) in mind, whilst expanding the following hyperplane formula using (1.41).

$$\begin{aligned}I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\alpha}{\partial p}\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right) &= \left[I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\right]\frac{\partial\alpha}{\partial p}\\ &-\sum_{\mu=1}^{m}(-1)^{\mu-1}\left[I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\alpha}{\partial p}\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\check{\phi}_\mu}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\right]\frac{\partial\phi_\mu}{\partial p}\\ &= \left[I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\right]\left[\frac{\partial\alpha}{\partial p}-\sum_\mu\lambda_\mu\frac{\partial\phi_\mu}{\partial p}\right] = 0\end{aligned}$$

So,

$$I\left(\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right)\cdot\left(\frac{\partial\alpha}{\partial p}\frac{\partial\phi_1}{\partial p}\cdots\frac{\partial\phi_m}{\partial p}\right) = 0 \tag{9.32}$$

is the equation that we have to solve in order to find the point p which gives an extremum of $\alpha(p)$ subject to constraints $\phi_\mu = \text{const}$. If we set $Y = \partial\alpha/\partial p$ and $H_\mu = \partial\phi_\mu/\partial p$ in the hyperplane projection formula (6.10), then it says that (9.32) is $Y_\perp = 0$ and so $Y = Y_\parallel$. In other words, at the extremum, the derivative $\partial\alpha/\partial p$ is a linear combination of the $\partial\phi_\mu/\partial p$. The Lagrange multipliers λ_μ are the coefficients in this linear combination (9.30). Equation (9.32) automatically takes care of finding the multipliers.

9.7.2 Equations for the extremal curvatures

The curvature k at a point p in the hypersurface along a path with tangent q is given by equation (9.24). We seek an extremum of k at p by varying the tangent vector q. The vector q must remain in the tangent space $l = L.Ip.In_1$ to the hypersurface at p that was introduced by equation (9.6). Hence we seek an extremum of k by varying q subject to the constraint $ql = 0$. This constraint gives the three scalar constraints $qL = 0$, $qIp = 0$ and $qIn_1 = 0$. In addition, we need to vary q subject to $qIq = \text{const}$. This is because the particular parameterisation used for the path must always be the same as we vary the direction q of the path at p. Section 9.6.1 showed that $qIq = 1$ if the path parameter τ is arc-length. So, $qIq = 1$ is the usual constraint, but our formula (9.24) for the curvature is valid for an arbitrary path parameterisation, so we preserve the generality by only requiring $qIq = \text{const}$. Putting all four constraints in equation (9.32) gives the following equation for a direction of extremal curvature.

$$I\left(\frac{\partial qIq}{\partial q}.\frac{\partial qL}{\partial q}.\frac{\partial qIp}{\partial q}.\frac{\partial qIn_1}{\partial q}\right).\left(\frac{\partial k}{\partial q}.\frac{\partial qIq}{\partial q}.\frac{\partial qL}{\partial q}.\frac{\partial qIp}{\partial q}.\frac{\partial qIn_1}{\partial q}\right) = 0$$

Evaluating the derivatives[17], simplifying, and swopping the order of some factors, gives the equation,

$$(q.IL.p.n_1).\left(Iq.\frac{\partial k}{\partial q}.L.Ip.In_1\right) = 0\ . \tag{9.33}$$

This is our canonical form of the Lagrange multiplier equation for extremal curvatures. From this point onwards, there are two tracks through the theory. If we explore the consequences of equation (9.33), we move on a track which appears to yield the most theoretically important formulae. These formulae all involve the surface normal n_1. However, computationally, it is easier to characterise the surface with the hyperplane function $F(p)$, which can easily

[17]See equations (4.16) and (4.17).

be derived from the equation $\phi(p) = 0$ for the hypersurface via equation (9.1). Of course, both tracks are equivalent because the normal n_1 is related to F by equation (9.4). Upon substituting (9.4) into (9.33), we obtain,

$$(q.IL.p.IF).\left(Iq.\frac{\partial k}{\partial q}.L.Ip.F\right) = 0 \tag{9.34}$$

using antisymmetry. This is the form of the Lagrange multiplier equation for extremal curvatures that leads to the most useful formulae for computational work.

9.7.3 Curvature operator

In this section we show that equation (9.34) for the extremal curvatures can be put into the form of an eigenvalue equation. The operator of this eigenvalue equation naturally characterises the curvature of the hypersurface and so it is called the *curvature operator*.

We start the derivation of the curvature operator by using equation (9.24) for k to evaluate the derivative $\partial k/\partial q$ so that it can be substituted into equation (9.34). The formula for the derivative $\partial k/\partial q$ contains terms proportional to Iq. However, we do not need to calculate[18] these because they are killed by antisymmetry when substituted in equation (9.34). Consequently,

$$\begin{aligned} \frac{\partial k}{\partial q} &= -\frac{(-1)^{n-1}(pL)}{(FIF)^{1/2}(qIq)^{1/2}}\frac{\partial}{\partial q}\left[q.\left(q\frac{\partial F}{\partial p}\right)\right] \\ &\qquad + \text{terms which are killed by antisymmetry in (9.34)}\ . \end{aligned} \tag{9.36}$$

Now $\partial F/\partial p$ is a correlation[19] because when it acts on an argument q to give $q\partial F/\partial p$, the result is a hyperplane. Consequently, the term $q.(q\partial F/\partial p)$ in square brackets in (9.36) is a number-valued function of the point q. Furthermore, it is a homogeneous function of degree 2 in q. Hence, by Euler's

[18] In order to evaluate a term such as $\partial(qIq)^{1/2}/\partial q$, we need the chain rule. In general, we need to evaluate $\partial\psi(\phi(p))/\partial p$ where ψ is a function of a number-valued argument ϕ. Using (4.15),

$$\delta p.\frac{\partial \psi}{\partial p} = \delta\psi = \delta\phi\frac{\partial\psi}{\partial\phi} = \delta p.\frac{\partial\phi}{\partial p}\frac{\partial\psi}{\partial\phi}\ .$$

Since this holds for all variations δp we have the chain rule,

$$\frac{\partial\psi}{\partial p} = \frac{\partial\phi}{\partial p}\frac{\partial\psi}{\partial\phi}\ . \tag{9.35}$$

[19] See section 3.3.

theorem (4.12),

$$q\frac{\partial}{\partial q}\left[q.\left(q\frac{\partial F}{\partial p}\right)\right] = 2q.\left(q\frac{\partial F}{\partial p}\right) \ .$$

Since this holds for all points q, we must have,

$$\frac{\partial}{\partial q}\left[q.\left(q\frac{\partial F}{\partial p}\right)\right] = 2q\frac{\partial F}{\partial p} + H \tag{9.37}$$

where H is some hyperplane containing q so that $qH = 0$. Now, from section 9.4, we know that q lies in the subspace $l = L.Ip.F$. Hence, the hyperplane H must be a linear combination of hyperplanes L, Ip and F. Hence, when we calculate $\partial k/\partial q$ by substituting (9.37) into (9.36), the hyperplane H can be ignored[20] because it will be killed by antisysmmetry when $\partial k/\partial q$ is substituted into the extremal curvature equation (9.34). Consequently, (9.34) becomes,

$$(q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.L.Ip.F\right) = 0 \ . \tag{9.38}$$

The next stage in the derivation of the curvature operator is to expand (9.38) using (1.41).

$$\begin{aligned}
0 \;&=\; (q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.L.Ip.F\right) \\
&=\; \left[(q.IL.p.IF).\left(q\frac{\partial F}{\partial p}.L.Ip.F\right)\right] Iq - [(q.IL.p.IF).(Iq.L.Ip.F)]q\frac{\partial F}{\partial p} \\
&+\; \left[(q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.Ip.F\right)\right] L - \left[(q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.L.F\right)\right] Ip \\
&+\; \left[(q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.L.Ip\right)\right] F
\end{aligned} \tag{9.39}$$

The numerical coefficients in the square brackets can be evaluated by formula (1.43). The coefficient of Iq is particularly simple to evaluate because only a single term in the summation in (1.43) survives. To see this, first note that the tangent vector q lies in subspace $L.Ip.F$. Hence, $qL = 0$, $qIp = 0$ and $qF = 0$ and so the only non-zero product with q is $q.q\partial F/\partial p$. Then, the only non-zero product with p is pL. This is because $pIp = 0$ since p is on the

[20] We could have invoked equation (9.1) and used coordinates and the equality of mixed partial derivatives to show that the correlation $\partial F/\partial p$ is, in fact, a polarity and that the hyperplane H vanishes. However, the argument given in the text seems more elegant.

quadric and $pF=0$ because p is on the hypersurface[21]. The coefficient of Iq then evaluates to,

$$(q.IL.p.IF).\left(q\frac{\partial F}{\partial p}.L.Ip.F\right) = -\left(q.q\frac{\partial F}{\partial p}\right)(pL)^2(FIF)\ .$$

The remaining four terms on the rhs of (9.39) can be combined into a single term. In order to do this we note that, in each of the four coefficients, the only non-zero product with Iq is qIq. If we factor this out, the four terms can be combined as,

$$\begin{aligned}
& -[(q.IL.p.IF).(Iq.L.Ip.F)]q\frac{\partial F}{\partial p} + \left[(q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.Ip.F\right)\right]L \\
& -\left[(q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.L.F\right)\right]Ip + \left[(q.IL.p.IF).\left(Iq.q\frac{\partial F}{\partial p}.L.Ip\right)\right]F \\
= & -(qIq)(IL.p.IF).\left(q\frac{\partial F}{\partial p}.L.Ip.F\right)\ .
\end{aligned}$$

All five terms on the rhs of (9.39) can now be replaced by,

$$0 = -\left(q.q\frac{\partial F}{\partial p}\right)(pL)^2(FIF)Iq - (qIq)(IL.p.IF).\left(q\frac{\partial F}{\partial p}.L.Ip.F\right)\ .$$

Now apply the polarity I to both sides of the above equation, in order to convert it from a hyperplane equation to a point equation.

$$0 = -(-1)^{n-1}\left(q.q\frac{\partial F}{\partial p}\right)(pL)^2(FIF)q - (qIq)(L.Ip.F).I\left(q\frac{\partial F}{\partial p}.L.Ip.F\right)$$

Now divide throughout by $(qIq)^{1/2}(FIF)^{3/2}(pL)$.

$$0 = -\frac{(-1)^{n-1}(pL)\left(q.q\frac{\partial F}{\partial p}\right)}{(FIF)^{1/2}(qIq)^{1/2}}q - \frac{(qIq)^{1/2}(L.Ip.F).I\left(q\frac{\partial F}{\partial p}.L.Ip.F\right)}{(FIF)^{3/2}(pL)} \tag{9.40}$$

The coefficient of q is identical to the term k_{bare} in equation (9.27) for the curvature of the hypersurface in direction q. So, with a final re-arrangement to tidy up by swopping the order of factors,

$$(L.Ip.F).I\left(q\frac{\partial F}{\partial p}.L.Ip.F\right) = (-1)^{3(n-4)}I\left(q\frac{\partial F}{\partial p}.L.Ip.F\right).(L.Ip.F)$$

[21] See the end of section 9.2.

to put the product of points to the left of the hyperplanes[22], we obtain the following equation for the directions of extremal curvature.

$$-\frac{(-1)^{n-1}(qIq)^{1/2}I\left(q\frac{\partial F}{\partial p}.L.Ip.F\right).(L.Ip.F)}{(pL)(FIF)^{3/2}} = k_{\text{bare}}q \tag{9.41}$$

Since qIq is a constant, this is an eigenvalue equation $K(q) = k_{\text{bare}}q$ where the curvature operator is defined as,

$$K(q) = -\frac{(-1)^{n-1}(qIq)^{1/2}I\left(q\frac{\partial F}{\partial p}.L.Ip.F\right).(L.Ip.F)}{(pL)(FIF)^{3/2}}\ . \tag{9.42}$$

The curvature operator characterises the hypersurface. The eigenvectors of the curvature operator are the directions of the extremal curvatures. The eigenvalues are the extremal bare curvatures k_{bare}. In classical differential geometry [14] the extremal curvatures are called the *principal curvatures*. In our case we have to keep in mind that the eigenvalues are the bare curvatures and so the principal curvatures are $k = k_{\text{sphere}} + k_{\text{bare}}$.

We now pick up the saddle-shaped surface example that we left at the beginning of section 9.7. The function `homopoly4(p,q,c,E)` in the package `curvature.i` computes the curvature operator given by equation (9.42). At the origin IN, the maximum and minimum curvatures were shown in section 9.7 to be in the directions $q = (a_1 \pm a_2)/\sqrt{2}$ with curvatures ± 0.2. We can show that these directions are eigen-directions of the curvature operator[23].

```
> print_element, homopoly4(IN,(a1+a2)/sqrt(2.0),c,E);
+0.141a1+0.141a2
> print_element, 0.2*(a1+a2)/sqrt(2.0);
+0.141a1+0.141a2
> print_element, homopoly4(IN,(a1-a2)/sqrt(2.0),c,E);
-0.141a1+0.141a2
> print_element, -0.2*(a1-a2)/sqrt(2.0);
-0.141a1+0.141a2
```

The `curvature.i` package defines surfaces using (9.3). The coefficients c_{ij} have been set up so that the polynomial ϕ contains no term incorporating pN. This ensures that $F(p)$ always represents a plane so $IF.L = 0$ and hence $k_{\text{sphere}} = 0$ by (9.26).

```
> F=homopoly2(IN,c,E);
> print_element, wap(I(F),L);
+0.000a1a2a3a4a5
```

[22]See equation (1.13). The sign $(-1)^{3(n-4)} = (-1)^{3n} = (-1)^{3n-2n} = (-1)^n$.

[23]At the time of writing Yorick does not have a canned eigenvector routine, so we can only verify the eigenvectors by application of the operator.

9.7.4 Invariants of the curvature operator

The invariants of the curvature operator are clearly important. The invariants of an operator were considered in section 4.4.3. In the case of a 2-d surface, the subspace of directions $L.Ip.F$ is two-dimensional, so there are only two invariants, the trace and the determinant. In classical differential geometry [14] these are called the *mean curvature* and the *Gaussian curvature.* In section 4.4.3 the invariants were considered from the point of view that the operator acted on hyperplanes. However, the curvature operator has been defined by (9.42) to act on points q which live in $L.Ip.F$ and have the meaning of directions at p. If we have an operator g that has been defined by its action on points, then its invariants are more naturally $g(X).\partial/\partial X$. Since, $g(X).\partial/\partial X = X.g^{-1}(\partial/\partial X)$ the theory of section 4.4.3 just needs to be applied to the inverse operator g^{-1} instead of g. The trace is,

$$\begin{aligned} \mathrm{tr}(K^{-1}) &= K(q).\frac{\partial}{\partial q} = K\left(\sum_i (qA_i)a_i\right).\frac{\partial}{\partial q} \\ &= \sum_i K(a_i).\frac{\partial(qA_i)}{\partial q} = \sum_i K(a_i).A_i \end{aligned} \tag{9.43}$$

and if the reference points a_i are also the eigenpoints of the curvature operator so that $K(a_i) = k_{\text{bare}}^{(i)} a_i$,

$$\mathrm{tr}(K^{-1}) = \sum_i K(a_i).A_i = \sum_i k_{\text{bare}}^{(i)} a_i.A_i = \sum_i k_{\text{bare}}^{(i)} \; .$$

where $k_{\text{bare}}^{(i)}$ denotes the ith eigenvalue. For a 2-d surface the mean curvature[24] is $\mathrm{tr}(K^{-1})/2$. With X a pseudonumber, the determinant or Gaussian curvature is,

$$\begin{aligned} \det(K^{-1}) &= K(X).\frac{\partial}{\partial X} = K((X.A_1 \ldots A_n)a_1 \ldots a_n).\frac{\partial}{\partial X} \\ &= K(a_1 \ldots a_n).\frac{\partial(X.A_1 \ldots A_n)}{\partial X} \\ &= K(a_1 \ldots a_n).A_1 \ldots A_n \end{aligned} \tag{9.44}$$

and if the reference points a_i are also the eigenpoints of the curvature operator $\det(K^{-1}) = \prod_i k_{\text{bare}}^{(i)}$. For a 2-d surface the Gaussian curvature is $\det(K^{-1}) = k_{\text{bare}}^{(1)} k_{\text{bare}}^{(2)}$.

[24] We emphasise that $\mathrm{tr}(K^{-1})/2$ is only the mean curvature of classical differential geometry when $k_{\text{sphere}} = 0$ so that the bare curvature is the same as the true curvature.

Let's calculate the curvature invariants using the saddle-shaped surface example. At the origin IN the direction subspace $l = L.Ip.F$ has already been shown to be the line a_1a_2 in section 9.7. From (9.43) the mean curvature at the origin turns out to be zero. The dual hyperplanes are conveniently calculated by using the elliptic polarity as $A_i = (-1)^{n-1}Ia_i$ and $n = 5$ so $A_i = Ia_i$.

```
> wap(homopoly4(IN,a1,c,E),I(a1))+wap(homopoly4(IN,a2,c,E),I(a2))
0+0i
```

From (9.44) the Gaussian curvature at the origin is -0.04 .

```
> Ka1a2=wap(homopoly4(IN,a1,c,E),homopoly4(IN,a2,c,E));
> print_element, Ka1a2
-0.040a1a2
> print_element, wap(Ka1a2,I(wap(a1,a2)));
-0.040a1a2a3a4a5
```

Let's calculate the curvature invariants at some arbitrary point on the saddle-shaped surface. Pick a point on the quadric and check that it does not lie on the saddle surface by evaluating $\phi(p)$.

```
> p=pcon(IL,a1+a2,IN);
> homopoly1(p,c,E)
-0.8+0i
```

Now send the point onto the surface and check that it has worked so $\phi(p) = 0$.

```
> p=homopoly3(p,c,E);
> homopoly1(p,c,E)
7.76831e-17+0i
```

Also check that $pIp = 0$ so that the point is also on the quadric and extract the (x, y, z) Cartesian coordinates of the point.

```
> print_element, wap(p,I(p));
-0.000a1a2a3a4a5
> print_element, q_from_p(p);
+0.963a1+0.963a2+0.185a3
```

Calculate the subspace of directions $l = L.Ip.F$.

```
> F=homopoly2(p,c,E);
> l=wap(L,wap(I(p),F));
> print_element, l;
+4.000ia1a2+0.770ia1a3-0.770ia2a3-3.995ia1a4+3.995ia2a4
+0.000a3a4-3.995a1a5+3.995a2a5+0.000a3a5
```

Now we need to set up reference points in the subspace of directions l. l is a line so we need two reference points. We start by picking a single reference point q_1 on l. One way to do this is to project a point such as a_1 (say) onto the subspace $l = L.Ip.F$ using the projection formula (6.8). In this case it is proportional to $(a_1.Il).l$. We then normalise q_1 so that $q_1 I q_1 = 1$ so that the path parameter is arc-length and check that the projected point q_1 is on l because $q_1 l = 0$.

```
> q1=wap(wap(a1,I(l)),l);
> q1=q1/sqrt(wap(q1,I(q1)));
> print_element, q1
+0.983a1-0.035a2+0.182a3-0.946a4+0.946ia5
> print_element, wap(q1,l);
+0.000a1a2a3-0.000a1a2a4+0.000a1a3a4-0.000a2a3a4+0.000a1a2a5+0.000a1a3a5
```

We can then take the second reference point $q_2 \propto l.Iq_1$ as the normalised intersection of line l with the hyperplane Iq_1.

```
> q2=wap(l,I(q1));
> q2=q2/sqrt(wap(q2,I(q2)));
> print_element, q2;
+0.000a1-0.982a2-0.189a3+0.981a4-0.981ia5
> print_element, wap(q2,l);
+0.000a1a2a3+0.000a1a3a4-0.000a2a3a4-0.000a1a3a5+0.000a2a3a5
```

We now have two normalised reference points q_1, q_2 with dual hyperplanes Iq_1, Iq_2.

```
> print_element, wap(q1,I(q1));
+1.000a1a2a3a4a5
> print_element, wap(q2,I(q2));
+1.000a1a2a3a4a5
> print_element, wap(q1,I(q2));
+0.000a1a2a3a4a5
```

From (9.43) the mean curvature is -0.0067,

```
> 0.5*(wap(homopoly4(p,q1,c,E),I(q1))+wap(homopoly4(p,q2,c,E),I(q2)))
-0.00666272+0i
```

and from (9.44) the Gaussian curvature is -0.035 .

```
> Kq1q2=wap(homopoly4(p,q1,c,E),homopoly4(p,q2,c,E));
> wap(Kq1q2,I(wap(q1,q2)));
-0.0346664+0i
```

9.7.5 Properties of the curvature operator

The curvature operator (9.42) acts on directions q that live in the subspace $L.Ip.F(p)$. It is self-adjoint. In order to prove self-adjointness, we start with the definition (3.54) of the adjoint. This shows that we just need to show that $K(q_1).Iq_2 = K(q_2).Iq_1$ for any pair of directions q_1, q_2. From equation (9.42),

$$K(q_1).Iq_2 = -\frac{(-1)^{n-1}(q_1Iq_1)^{1/2} I\left(q_1\frac{\partial F}{\partial p}.L.Ip.F\right).(L.Ip.F.Iq_2)}{(pL)(FIF)^{3/2}} .$$

Since q_2 lies in the subspace $L.Ip.F$, there is only one non-zero product involving Iq_2. so using (1.44),

$$\begin{aligned} K(q_1).Iq_2 &= \frac{(-1)^{n-1}(q_1Iq_1)^{1/2}\left[I\left(q_1\frac{\partial F}{\partial p}\right).Iq_2\right]\left[I(L.Ip.F).(L.Ip.F)\right]}{(pL)(FIF)^{3/2}} \\ &= \frac{(q_1Iq_1)^{1/2}\left[q_2.\left(q_1\frac{\partial F}{\partial p}\right)\right]\left[I(L.Ip.F).(L.Ip.F)\right]}{(pL)(FIF)^{3/2}} . \end{aligned} \tag{9.45}$$

Using (9.1) and the equality of mixed partial derivatives,

$$q_2.\left(q_1\frac{\partial F}{\partial p}\right) = q_2.\left(q_1\frac{\partial^2\phi}{\partial p\partial p}\right) = \left(q_2\frac{\partial}{\partial p}\right)\left(q_1\frac{\partial}{\partial p}\right)\phi = \left(q_1\frac{\partial}{\partial p}\right)\left(q_2\frac{\partial}{\partial p}\right)\phi .$$

The above result, together with the fact that qIq is a constant for a given path parameterisation, means that (9.45) is symmetric in variables q_1 and q_2. Hence $K(q_1).Iq_2 = K(q_2).Iq_1$ so that K is a self-adjoint operator.

Now section 3.11.3 showed that the eigenvectors of a self-adjoint operator are mutually orthogonal[25]. In the context of the curvature operator

[25] Some clarification may be needed here. The treatment of the eigenvalue problem for a self-adjoint operator B in section 3.11.3 considered the operator acting on hyperplanes. Section 3.11.1 explained that this was necessary in order to preserve the ordinary rules for matrices. However, in the definition (9.42), the curvature operator acts on points. Nevertheless, it is easy to see that the eigenvectors are orthogonal if we let the operator act on points instead of hyperplanes. Start from equation (3.63) and let the operator equation act on the reference point a_i. We get,

$$BU^{-1}a_i = U^{-1}Da_i = \lambda_i^{-1}U^{-1}a_i$$

so that the ith eigenvector is $U^{-1}a_i$. The reason that the eigenvalue is λ_i^{-1} is because the action of the diagonal operator on hyperplanes is $DA_i = \lambda_i A_i$ and since $a_iA_j = \delta_{ij}$ then $\delta_{ij} = Da_i.DA_j = \lambda_j Da_i.A_j$ so that $Da_i = \lambda_i^{-1}a_i$. Any pair of eigenvectors is orthogonal because,

$$IU^{-1}a_i.U^{-1}a_j = U^{-1}Ia_i.U^{-1}a_j = Ia_i.a_j = (-1)^{n-1}A_i.a_j = \delta_{ij}$$

where we used the fact that U is a congruence which commutes with the absolute elliptic polarity I and $IA_i = a_i$.

this statement means that if q and q' are any pair of eigendirections with eigenvalues k_{bare} and k'_{bare} so that $K(q) = k_{\text{bare}}q$ and $K(q') = k'_{\text{bare}}q'$, then $qIq' = 0$. The angle between any two directions in the hypersurface is given by equation (9.13). Hence, the condition $qIq' = 0$ means that any two directions of extremal curvature are at a right-angle to each other.

9.7.6 Extremal curvature and hypersurface normal

Sections 9.7.3, 9.7.4 and 9.7.5 have just explored the determination of the extremal curvatures at a point on a hypersurface by starting from equation (9.34). Our main result was the curvature operator (9.42) whose eigenvalues turned out to be the bare curvatures (9.27). This approach is natural from a computational point of view. However, it is not the best way to go in order to develop further the theory of hypersurfaces. This is because the important quantity is the curvature which, as we have seen in (9.25), is composed of the spherical curvature and the bare curvature. The bare curvature depends on how deeply the hypersurface $\phi(p) = 0$ cuts into the absolute quadric $pIp = 0$ since the intersection of these two hypersurfaces is the set of points which make up the hypersurface in the Euclidean space. When the hypersurface $\phi = 0$ only cuts the absolute quadric shallowly, the tangent hypersphere shown in figure 9.4 is small, so the spherical curvature k_{sphere} is large and the bare curvature k_{bare} is corresponding small. Similarly, when the hypersurface $\phi = 0$ cuts so deeply into the absolute quadric that the tangent hypersphere has degenerated into a hyperplane (hypersphere of infinite radius), the bare curvature is equal to the curvature. In other words, the bare curvature is not fundamental, because, at least locally, we can produce the same hypersurface in the Euclidean space by cutting shallowly or deeply into the absolute quadric with the hypersurface $\phi = 0$. Consequently, the route that we have been following from (9.34) seems to be a theoretical dead-end because of the rôle it gives to the bare curvature. So, we now retrace our steps and study the determination of the extremal curvatures from equation (9.33).

Starting from equation (9.33), we need $\partial k/\partial q$. This can be obtained by differentiating the formula (9.19) for the curvature along a path. We shall need the derivative,

$$\begin{aligned} \frac{dn_1}{d\tau} &= \frac{dp}{d\tau}\frac{\partial n_1}{\partial p} = \left((pL)q + \frac{p}{pL}\frac{d(pL)}{d\tau}\right)\frac{\partial n_1}{\partial p} \\ &= (pL)q\frac{\partial n_1}{\partial p} + \text{term proportional to } n_1 \end{aligned} \tag{9.46}$$

where $dp/d\tau$ has been replaced using equation (9.20) and the term $p\partial n_1/\partial p$

is proportional to n_1 by Euler's theorem (9.23). Whenever we use this result, the term proportional to n_1 will vanish by antisymmetry or because the product $qIn_1 = 0$ from section 9.6. So, substituting (9.46) into (9.19)[26],

$$k = -\frac{\left(q.I\frac{dn_1}{d\tau}\right)}{\sqrt{qIq}} = -\frac{(pL)}{\sqrt{qIq}}q\left(q\frac{\partial In_1}{\partial p}\right) \ . \tag{9.47}$$

By using the same type of argument that was deployed to obtain (9.37), we find,

$$\frac{\partial k}{\partial q} = -\frac{(pL)}{\sqrt{qIq}}\left(q\frac{\partial In_1}{\partial p}\right) + H \tag{9.48}$$

where H is some hyperplane containing q so that $qH = 0$. Now, from section 9.6, we know that q lies in the subspace $l = L.Ip.In_1$. Hence, the hyperplane H must be a linear combination of hyperplanes L, Ip and In_1. So, when we substitute (9.48) into (9.33), the result is,

$$(q.IL.p.n_1).\left(Iq.\left(q\frac{\partial In_1}{\partial p}\right).L.Ip.In_1\right) = 0 \ . \tag{9.49}$$

This equation is analogous to (9.38) except that we are now expressing everything in terms of the normal vector n_1 instead of the tangent hyperplane F. We were able to show that (9.38) was an eigenvalue equation by expanding and taking the term in Iq on the other side of the equation. A similar procedure works with (9.49). However, it is not necessary to go through all the working again because if make the substitutions $IF \to n_1$ and $F \to (-1)^{n-1}In_1$ in (9.38) it changes into (9.49). The working following (9.38) can then be short-circuited by jumping straight to (9.40) and making the above substitutions. The result is,

$$0 = -\frac{(pL)\left(q.q\frac{\partial In_1}{\partial p}\right)}{(qIq)^{1/2}}q - \frac{(-1)^{n-1}(qIq)^{1/2}(L.Ip.In_1).I\left(q\frac{\partial In_1}{\partial p}.L.Ip.In_1\right)}{(pL)} \ .$$

This works because the calculations following (9.38) make use of the facts $qF = 0$ and $pF = 0$ whilst we also have $qIn_1 = 0$ and $pIn_1 = 0$. The coefficient of q in the above equation is immediately recognised as the curvature k from (9.47). Hence we have our eigenvalue equation.

$$\frac{(-1)^{n-1}\sqrt{qIq}}{(pL)}(L.Ip.In_1).I\left(q\frac{\partial In_1}{\partial p}.L.Ip.In_1\right) = kq$$

[26]In order to compare the current theory and the classical differential geometry of 2-d surfaces, note that (9.47) is analogous to equation (19.35) of [14] which Coxeter uses to introduce the fundamental magnitudes of the second order.

A final swop to bring the lhs to standard form with the product of points on the right of the product of hyperplanes gives,

$$\frac{(-1)^{3(n-4)}\sqrt{qIq}}{(pL)}\left(q\frac{\partial n_1}{\partial p}.IL.p.n_1\right).(L.Ip.In_1) = kq\ . \tag{9.50}$$

This equation is analogous to (9.41) except that the eigenvalues are the extremal curvatures instead of the bare curvatures. We could take the operator on the lhs of (9.50) as the curvature operator instead of (9.42). However, it is more theoretically fruitful to examine (9.50) from another perspective.

Substitute (9.46) into (9.50). The extraneous term n_1 in (9.46) vanishes by antisymmetry.

$$\frac{(-1)^{3(n-4)}\sqrt{qIq}}{(pL)^2}\left(\frac{dn_1}{d\tau}.IL.p.n_1\right).(L.Ip.In_1) = kq \tag{9.51}$$

Now it turns out that $dn_1/d\tau$ is in the subspace $L.Ip.In_1$. Let's prove this fact. From (9.4) $n_1.L =$ for all points p on the hypersurface. Hence $0 = d(n_1.L) = dn_1.L$ so that dn_1 is in L. Also from (9.4) $n_1Ip = 0$ so $0 = dn_1.Ip + n_1Idp$. Using (9.20), dp is a linear combination of p and q and $n_1Ip = 0$ and also $n_1Iq = 0$ so that $0 = dn_1.Ip$ and we have shown that dn_1 is in Ip. Finally, $n_1In_1 = 1$ so that $0 = dn_1.n_1$ and dn_1 is in In_1. Therefore, dn_1 is in the subspace of directions $L.Ip.In_1$. Note that this applies to any direction in the hypersurface at p, not just the extremal directions determined by the eigenvalue equation. Now write dn_1 using the projection formula (6.6) with respect to the subspace $H_1H_2H_3 = L.Ip.In_1$.

$$dn_1 = \frac{(IL.p.n_1).(dn_1.(L.Ip.In_1))}{(IL.p.n_1).(L.Ip.In_1)} - \frac{(dn_1.IL.p.n_1).(L.Ip.In_1)}{(IL.p.n_1).(L.Ip.In_1)}$$

Any point can be decomposed in this way, but we have just shown that $dn_1.(L.Ip.In_1) = 0$ and so the projection formula collapses to,

$$dn_1 = -\frac{(dn_1.IL.p.n_1).(L.Ip.In_1)}{(IL.p.n_1).(L.Ip.In_1)} = \frac{(dn_1.IL.p.n_1).(L.Ip.In_1)}{(-1)^{n-1}(pL)^2}\ . \tag{9.52}$$

But, the rhs of the above equation is almost exactly the lhs of (9.51). Hence, our eigenvalue equation (9.51) becomes the delightfully simple equation,

$$\frac{dn_1}{d\tau} = -k\frac{q}{\sqrt{qIq}} \tag{9.53}$$

which is called *Rodrigues's formula* in classical differential geometry [14]. It is important to bear in mind that (9.53) does not hold for arbitary directions q in the hypersurface at p. It is only valid along the directions of extremal curvature. Note that if we substitute for $dn_1/d\tau$ from (9.53) in equation (9.19) we get the reassuring consistency check $k = k$.

9.8 Gauss's hyperspherical representation

Gauss saw that the normal $n_1(p)$ can be regarded as a map $p \mapsto n_1(p)$. In our terms, the map takes a point p in the hypersurface to a point n_1 in the bulk elliptic space as shown in figure 9.2. So, instead of working with a cell of the hypersurface near p, we can consider the new cell of hypersurface under the action of the map. Suppose that we have coordinates[27] for the original hypersurface so that $p = p(\theta_1, \ldots, \theta_{n-3})$. Under the normal map the points of the new hypersurface are $n_1 = n_1(p) = n_1(\theta_1, \ldots, \theta_{n-3})$. Using (4.3) the volume element of the new cell is,

$$dV' = n_1 . \frac{\partial n_1}{\partial \theta_1} \ldots \frac{\partial n_1}{\partial \theta_{n-3}} d\theta_1 \ldots d\theta_{n-3} \; . \tag{9.54}$$

We need a number to represent the volume of the cell. The logical way to get a number is,

$$\sqrt{dV'.IdV'} = d\theta_1 \ldots d\theta_{n-3} \sqrt{\frac{\partial n_1}{\partial \theta_1} \ldots \frac{\partial n_1}{\partial \theta_{n-3}} I \frac{\partial n_1}{\partial \theta_1} \ldots \frac{\partial n_1}{\partial \theta_{n-3}}} \tag{9.55}$$

where the leading factor n_1 has vanished because $n_1 I n_1 = 1$ and by differentiation $n_1 I \partial n_1 / \partial \theta_i = 0$. Now suppose we take the coordinates θ_i as arc-length parameters[28] along the extremal directions at p. Then, according to (9.53),

$$\frac{\partial n_1}{\partial \theta_i} = -k_i \frac{\partial}{\partial \theta_i} \left(\frac{p}{pL} \right) \; .$$

Here, k_i is the ith curvature eigenvalue of (9.53) or equivalently (9.50). Also,we replaced the direction q in (9.53) by its definition (9.20). Substituting these results into the (9.55) gives the volume of a cell in the new hypersurface as,

$$\sqrt{dV'.IdV'} = \left(\prod_i k_i \right) \left(\prod_i d\theta_i \right) \sqrt{ \left(\prod_i \frac{\partial}{\partial \theta_i} \left(\frac{p}{pL} \right) \right) I \left(\prod_i \frac{\partial}{\partial \theta_i} \left(\frac{p}{pL} \right) \right) } \tag{9.56}$$

where all the products contain $(n-3)$ factors. Now, equation (7.27) is the formula for the volume element $\sqrt{-dV.IdV}$ on the original hypersurface in

[27] In order to avoid confusion, we note that the $n-3$ coordinates $\theta_1, \ldots, \theta_{n-3}$ specify the set of points on the intersection of the absolute quadric $pIp = 0$ and the hypersurface $\phi(p) = 0$. In other words these $n-3$ coordinates specify the hypersurface embedded in the Euclidean space on the quadric. These coordinates do not apply to the hypersurface $\phi(p) = 0$ of the bulk elliptic space.

[28] So that $qIq = 1$ from section 9.6.1

the quadric. Comparing (9.56) with (7.27) we have,

$$\sqrt{dV'.IdV'} = \left(\prod_i k_i\right)\sqrt{-dV.IdV} \tag{9.57}$$

where the product of extremal curvatures is the curvature invariant formed from the determinant of the curvature operator[29] and known as the Gaussian curvature. Equation (9.57) says that the ratio of volume cells of the new hypersurface to the orginal one is the Gaussian curvature. This result is known as Gauss's *spherical representation* [14] in the case of a 2-d surface embedded in 3-d Euclidean space.

9.8.1 The normal map in terms of coordinates

There is an interesting relation between the volume dV of a cell of the hypersurface and the normal n_1 to the hypersurface. From (7.26) the volume element,

$$dV = d\theta_1 \ldots d\theta_{n-3} IL.\frac{p}{pL}.\frac{\partial}{\partial\theta_1}\left(\frac{p}{pL}\right)\ldots\frac{\partial}{\partial\theta_{n-3}}\left(\frac{p}{pL}\right) \tag{9.58}$$

is a product of $(n-1)$ points. Hence IdV is the point formed by the intersection of the $(n-1)$ hyperplanes,

$$L, I\left(\frac{p}{pL}\right), I\left(\frac{\partial}{\partial\theta_1}\left(\frac{p}{pL}\right)\right), \ldots, I\left(\frac{\partial}{\partial\theta_{n-3}}\left(\frac{p}{pL}\right)\right) \; .$$

Now the normal n_1 is on each of these $(n-1)$ hyperplanes. Hence IdV and n_1 are the same point up to an unimportant weight. The weight can be found from the requirement that $n_1 I n_1 = 1$. The result is,

$$IdV = in_1\sqrt{-dV.IdV} \; . \tag{9.59}$$

Originally, the normal n_1 was defined by (9.4) using the hyperplane $F = \partial\phi/\partial p$. However, equation (9.59) can be thought of as defining the normal n_1 directly in terms of the coordinates used to specify the hypersurface by $p = p(\theta_1, \ldots, \theta_{n-3})$. In this way we make contact with classical differential geometry [14] where a 2-d surface embedded in 3-d Euclidean space has points $p = p(\theta_1, \theta_2)$.

[29]Here the curvature operator has to be the operator on the lhs of equation (9.50) because we want its eigenvalues to be the extremal true curvatures and not the bare curvatures.

9.8.2 The inverse of the normal map

Currently, we have Gauss's map $p \mapsto n_1(p)$ that maps a $(n-3)$-d hypersurface in the quadric to a new $(n-3)$-d hypersurface in the bulk $(n-1)$-d elliptic space. It would be helpful if we could invert this map so that $n_1 \mapsto p$. In classical differential geometry the inverse map from the normal vectors back onto the original surface is one-to-many so it is not very useful. However, in our case, the map is drawn in figure 9.2 and it is qualitatively different from classical differential geometry because the $n_1(p)$ are just points in the bulk elliptic space.

From (9.54) the cell dV' in the bulk elliptic space is a product of $(n-2)$ points. Hence IdV' is the line formed by the intersection of $(n-2)$ hyperplanes. Furthermore, the point p and the point at infinity IL are both on each of these hyperplanes. Therefore, the line IdV' and the line $IL.p$ are the same modulo an unimportant weight. The weight can be found by calculating $dV'IdV'$ and the result is,

$$IdV' = iIL.\frac{p}{pL}\sqrt{dV'IdV'}\ . \tag{9.60}$$

The actual point p on the hypersurface in the quadric can easily be found by intersecting the line IdV' with the quadric. The function `p_from_Z` which implements equation (7.24) can be used to compute the intersection. So, just as (9.59) defined n_1 in terms of the coordinates $p = p(\theta_1, \ldots, \theta_{n-3})$ used to define the hypersurface in the quadric, we can regard (9.60) as defining p in terms of the coordinates $n_1 = n_1(\theta_1, \ldots, \theta_{n-3})$ used to define the hypersurface in the bulk elliptic space. We are now able to calculate on the hypersurface in the bulk elliptic space, where calculations might be more straightforward, and map our results back to the hypersurface in the quadric using the inverse (9.60) of Gauss's normal map.

Let's take up the saddle-shaped surface example from the point that we left it in section 9.7.4. We were at a point p on the saddle-shaped surface,

```
> homopoly1(p,c,E)
7.76831e-17+0i
```

with the following Cartesian coordinates.

```
> print_element, q_from_p(p);
+0.963a1+0.963a2+0.185a3
```

We need the normalised version of F,

```
> F=homopoly2(p,c,E);
> F=F/sqrt(wap(F,I(F)));
```

so that we can calculate n_1 from (9.4).

```
> n1=I(F)-wap(I(F),L)*p/wap(p,L);
> print_element, n1;
-0.186a1-0.186a2+0.965a3+0.179a4-0.179ia5
```

We also check the properties $n_1 I n_1 = 1$, $n_1 I p = 0$ and $n_1 L = 0$.

```
> print_element, wap(n1,I(n1));
+1.000a1a2a3a4a5
> print_element, wap(n1,I(p));
+0.000a1a2a3a4a5
> print_element, wap(n1,L);
+0.000a1a2a3a4a5
```

Now let's demonstrate the inverse map from the point n_1 on the surface in the bulk elliptic space back to the physical saddle-shaped surface embedded in 3-d Euclidean space. We would like to use equation (9.60) but we do not have a coordinate description $n_1 = n_1(\theta_1, \theta_2)$ of the surface in the bulk elliptic space which would allow us to calculate dV' from (9.54). However, in general, the direction subspace $l = L.Ip.F$ is,

$$\frac{\partial n_1}{\partial \theta_1} \cdots \frac{\partial n_1}{\partial \theta_{n-3}} d\theta_1 \dots d\theta_{n-3}$$

up to an unimportant weight. So, if we cheat a little, by calculating $l = L.Ip.F$,

```
> l=wap(L,wap(I(p),F));
```

and imagining that we obtained it from the derivatives of n_1, then dV' is the same as $n_1.l$ modulo a weight. So, from (9.60), $IL.p$ is $I(n_1.l)$ up to a weight. The line $IL.p$ is called, for historical reasons, the Z-form of a point in the `utilities.i` package. We can use the functions `p_from_Z`, `q_from_Z` or `r_from_Z` to extract various representations of a point from the line $IL.p$ as described in section 7.5.2. The function `q_from_Z` delivers the Cartesian coordinates of the point,

```
> print_element, q_from_Z(I(wap(n1,l)));
+0.963a1+0.963a2+0.185a3
```

and this is the original point on the saddle-shaped surface. This example has demonstrated how to go from the physical surface up to the surface in the bulk elliptic space and back down again.

9.9 Riemannian geometry

So far the theory of a curved hypersurface embedded in Euclidean space in Whitehead's algebra has had similarities to the theory of a 2-d surface in classical differential geometry [14]. The normal to the hypersurface plays a prominent rôle and we have Rodrigues's formula and Gauss's spherical representation - things which all appear in the classical differential geometry of 2-d surfaces. In this section we show how the current theory relates to the other approach to differential geometry - Riemannian geometry - which is the mathematical arena for general relativity [17, 29].

In Riemannian geometry the curvature at a point is characterised by the *Riemann curvature tensor*. This tensor is not closely related to the curvature operator (9.42) that characterises the curvature of a hypersurface in the treatment in this text. Instead, Riemann's curvature tensor measures the change in a vector after it has been parallel-transported around a small loop[30]. Now, in section 4.3.1 we studied parallel-transport around a small loop in elliptic space. Our central result was equation (4.33) which showed that parallel-transport around a small loop in elliptic space boils down to a rotation. The rotation angle being equal to the area of the loop. Of course, as explained at the start of section 9.4, we cannot use this result for parallel-transport around a small loop in a hypersurface on the absolute quadric because parallel-transport has to be defined in a different way. Nevertheless, section 9.4 went on to show how to treat parallel-transport on the hypersurface. Consequently, by using the theory of parallel-transport developed in section 9.4 in conjunction with the methods of section 4.3.1, we will find the formula for parallel-transport around a small loop on a hypersurface. In this way, will be able to show how the curvature tensor of Riemannian geometry shows up in Whitehead's algebra.

9.9.1 Parallel-transport on a loop in the hypersurface

If we have a path $p(\tau)$ and a vector $q(\tau)$ then parallel-transport over a small segment of path goes,

$$q(\tau + d\tau) = q(\tau) + d\tau \mathcal{A}(\tau) q(\tau) = [1 + d\tau \mathcal{A}(\tau)] q(\tau)$$

where $\mathcal{A}(\tau)$ is some linear operator acting on $q(\tau)$. If we start from $q(0)$ we have a multiplicative integral,

$$q(\tau) = f_\tau q(0) = \prod_0^\tau [1 + d\tau \mathcal{A}(\tau)] q(0)$$

[30] §91 of [29] introduces the curvature tensor in this manner.

for parallel-transport along the path. The overall operator is,

$$f_\tau = \prod_0^\tau [1 + d\tau \mathcal{A}(\tau)]$$

so that $q(\tau) = f_\tau q(0)$. Equation (4.24) also defined a similar multiplicative integral and equation (4.27) gave its expansion as an infinite series. Our task is to evaluate the first two terms in this series using the operator for parallel-transport in the hypersurface given by (9.16). Thus in the current situation,

$$\begin{aligned}
\mathcal{A}(\tau)q(0) &= -\frac{d}{d\tau}\left[(q(0)In_1)n_1 + \frac{(-1)^{n-1}(q(0)Ip)IL}{pL}\right] \qquad (9.61)\\
&= -\left(q(0)I\frac{dn_1}{d\tau}\right)n_1 - (q(0)In_1)\frac{dn_1}{d\tau} - (-1)^{n-1}\left[q(0)I\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right]IL
\end{aligned}$$

and we can write this operator as $\mathcal{A}(\tau) = d\mathcal{B}(\tau)/d\tau$ where,

$$\mathcal{B}(\tau)q(0) = -(q(0)In_1)n_1 - \frac{(-1)^{n-1}(q(0)Ip)IL}{pL}\ . \qquad (9.62)$$

Let's now work generally using $\mathcal{A}(\tau) = d\mathcal{B}(\tau)/d\tau$ in the series (4.27) for the multiplicative integral.

$$\begin{aligned}
f_\tau &= 1 + \int_0^\tau d\tau' \mathcal{A}(\tau') + \int_0^\tau d\tau' \mathcal{A}(\tau') \int_0^{\tau'} d\tau'' \mathcal{A}(\tau'') + \ldots\\
&= 1 + \int_0^\tau d\tau' \frac{\mathcal{B}(\tau')}{d\tau'} + \int_0^\tau d\tau' \mathcal{A}(\tau') \int_0^{\tau'} d\tau'' \frac{\mathcal{B}(\tau'')}{d\tau''} + \ldots\\
&= 1 + [\mathcal{B}(\tau) - \mathcal{B}(0)] + \int_0^\tau d\tau' \mathcal{A}(\tau')[\mathcal{B}(\tau') - \mathcal{B}(0)] + \ldots
\end{aligned}$$

We are integrating over a closed loop so $\mathcal{B}(\tau) = \mathcal{B}(0)$.

$$\begin{aligned}
f_\tau &= 1 + \int_0^\tau d\tau' \mathcal{A}(\tau')\mathcal{B}(\tau') - \int_0^\tau d\tau' \mathcal{A}(\tau')\mathcal{B}(0) + \ldots\\
&= 1 + \int_0^\tau d\tau' \mathcal{A}(\tau')\mathcal{B}(\tau') - 0 + \ldots
\end{aligned}$$

Now, by integrating by parts,

$$\begin{aligned}
\int_0^\tau d\tau' \mathcal{A}(\tau')\mathcal{B}(\tau') &= \int_0^\tau d\tau' \frac{\mathcal{B}(\tau')}{d\tau'}\mathcal{B}(\tau') = \int_0^\tau d\tau' \left\{\frac{d}{d\tau'}[\mathcal{B}(\tau')\mathcal{B}(\tau')] - \mathcal{B}(\tau')\frac{\mathcal{B}(\tau')}{d\tau'}\right\}\\
&= -\int_0^\tau d\tau' \mathcal{B}(\tau')\frac{\mathcal{B}(\tau')}{d\tau'} = -\int_0^\tau d\tau' \mathcal{B}(\tau')\mathcal{A}(\tau')
\end{aligned}$$

we can put the multiplicative integral into an antisymmetric form.

$$f_\tau = 1 + \frac{1}{2}\int_0^\tau d\tau'[\mathcal{A}(\tau')\mathcal{B}(\tau') - \mathcal{B}(\tau')\mathcal{A}(\tau')] + \ldots \tag{9.63}$$

The next step is to explicitly write down the form of the commutator using (9.61,9.62).

$$\begin{aligned}
\mathcal{A}(\tau')\mathcal{B}(\tau')q(0) &= -\left(\mathcal{B}q(0)I\frac{dn_1}{d\tau}\right)n_1 - (\mathcal{B}q(0)In_1)\frac{dn_1}{d\tau} - (-1)^{n-1}\left[\mathcal{B}q(0)I\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right]IL \\
&= \left\{\left[(q(0)In_1)n_1 + \frac{(-1)^{n-1}(q(0)Ip)IL}{pL}\right]I\frac{dn_1}{d\tau}\right\}n_1 \\
&\quad + \left\{\left[(q(0)In_1)n_1 + \frac{(-1)^{n-1}(q(0)Ip)IL}{pL}\right]In_1\right\}\frac{dn_1}{d\tau} \\
&\quad +(-1)^{n-1}\left\{\left[(q(0)In_1)n_1 + \frac{(-1)^{n-1}(q(0)Ip)IL}{pL}\right]I\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right\}IL \\
&= (q(0)In_1)\frac{dn_1}{d\tau}
\end{aligned}$$

In obtaining this result we have used $n_1L = 0$, $dn_1L = 0$, $n_1In_1 = 1$, $dn_1In_1 = 0$, $d(p/pL).L = 0$ and $d(p/pL)In_1 = 0$. Similarly,

$$\begin{aligned}
\mathcal{B}(\tau')\mathcal{A}(\tau')q(0) &= -(\mathcal{A}q(0)In_1)n_1 - \frac{(-1)^{n-1}(\mathcal{A}q(0)Ip)IL}{pL} \\
&= \left\{\left[\left(q(0)I\frac{dn_1}{d\tau}\right)n_1 + (q(0)In_1)\frac{dn_1}{d\tau} + (-1)^{n-1}\left(q(0)I\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right)IL\right]In_1\right\}n_1 \\
&\quad +(-1)^{n-1}(pL)^{-1}\left\{\left[\left(q(0)I\frac{dn_1}{d\tau}\right)n_1 + (q(0)In_1)\frac{dn_1}{d\tau} + (-1)^{n-1}\left(q(0)I\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right)IL\right]Ip\right\}IL \\
&= \left(q(0)I\frac{dn_1}{d\tau}\right)n_1 + (-1)^{n-1}\left(q(0)I\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right)IL\ .
\end{aligned}$$

Upon substituting these results for the commutator $\mathcal{AB} - \mathcal{BA}$ in the multiplicative integral (9.63), we obtain,

$$\begin{aligned}
q(\tau) &= f_\tau q(0) \\
&= q(0) + \frac{1}{2}\oint d\tau\left[(q(0)In_1)\frac{dn_1}{d\tau} - \left(q(0)I\frac{dn_1}{d\tau}\right)n_1 - (-1)^{n-1}\left(q(0)I\frac{d}{d\tau}\left(\frac{p}{pL}\right)\right)IL\right] + \ldots \\
&= q(0) + \frac{1}{2}\oint n_1 dn_1.Iq(0) + \ldots
\end{aligned}$$

At this stage in the derivation, the multiplicative integral is,

$$f_\tau = 1 + \frac{1}{2}\oint n_1 dn_1.I + \ldots\ . \tag{9.64}$$

Fortunately, this integral has already been evaluated in equation (4.29). The result is therefore,

$$f_\tau = 1 + \frac{1}{2}\oint n_1 dn_1.I + \ldots = 1 + \int_{S'} (dS'.In_1).I + \ldots \tag{9.65}$$

where,

$$dS' = n_1 \frac{\partial n_1}{\partial \theta_1}\frac{\partial n_1}{\partial \theta_2} d\theta_1 d\theta_2 \ . \tag{9.66}$$

Notice that the integral in (9.65) is written in terms of the points n_1 of the hypersurface in the bulk elliptic space and *not* over the points p of the original hypersurface in the quadric. So, the integral on the extreme right of (9.65) is over the 2-d area S' enclosed by the loop on the hypersurface in the bulk elliptic space. Furthermore, equation (9.65) is the same as (4.33) which was derived for parallel-transport around a loop in the bulk elliptic space. So, as far as the curvature is concerned, it seems that we can forget about the original hypersurface in the quadric and instead think of the hypersurface as existing in the bulk elliptic space. The results of section 4.3.1 apply so that parallel-transport around a loop of area dS' at point n_1 is produced by the action of a congruence with invariant line $dS'.In_1$.

9.9.2 Riemann tensor

Suppose we start with a direction q and parallel-transport it around a small loop. According to equation (9.65), the direction gets rotated by an amount dq given by,

$$dq = (dS'.In_1).Iq = \left(n_1 \frac{\partial n_1}{\partial \theta_1}\frac{\partial n_1}{\partial \theta_2} d\theta_1 d\theta_2 .In_1\right).Iq = \left(\frac{\partial n_1}{\partial \theta_1}\frac{\partial n_1}{\partial \theta_2} d\theta_1 d\theta_2\right).Iq \ . \tag{9.67}$$

As things stand, this simple formula lives in the bulk elliptic space. However, the Riemann tensor lives on the physical hypersurface - the one on the quadric. So, we have to work out how the rotation looks from the point of view of the physical hypersurface. We would like to just replace the derivatives by,

$$\frac{\partial n_1}{\partial \theta_i} = \frac{\partial p}{\partial \theta_i}\frac{\partial n_1}{\partial p} \tag{9.68}$$

in order to introduce directions on the physical hypersurface. This does not work out as intended because $\partial p/\partial \theta_i$ does not live in the subspace $l = L.Ip.In_1$ of directions[31], a fact that we have already noticed at the start of

[31]Notice that the subspace $l = L.Ip.In_1$ is the common subspace of directions for the physical hypersurface on the quadric *and* for the hypersurface in the bulk elliptic space.

section 9.6. Instead we have to use $\partial(p/pL)/\partial\theta_i$. However, using (9.20) to introduce the directions into (9.68) to get,

$$\frac{\partial n_1}{\partial \theta_i} = (pL)\frac{\partial}{\partial \theta_i}\left(\frac{p}{pL}\right)\frac{\partial n_1}{\partial p} + \frac{\partial \ln(pL)}{\partial \theta_i} p \frac{\partial n_1}{\partial p} \tag{9.69}$$

does not work either because the term $p\partial n_1/\partial p$ has appeared. The problem with this term is that p is not in the subspace l of directions and so we cannot make the link to the Riemann tensor which only has slots for directions. The solution is to return to (9.52) which shows that the change in the normal along any path always lives in the direction subspace $L.Ip.In_1$. In other words, $\partial n_1/\partial\theta_i$ is in $l = L.Ip.In_1$, but our problem boils down to the fact that $q\partial n_1/\partial p$, where q is a direction, *does not* lie in l. The way forward is now clear, whenever the renegade term $q\partial n_1/\partial p$ appeared in a formula for the curvature operator, for example (9.50), it was projected into the direction subspace l. So, from (9.52) and (9.69),

$$\begin{aligned}\frac{\partial n_1}{\partial \theta_i} &= \frac{(-1)^{n-1}}{(pL)^2}\left(\frac{\partial n_1}{\partial \theta_i}.IL.p.n_1\right).(L.Ip.In_1) \\ &= \frac{(-1)^{n-1}}{(pL)}\left(\frac{\partial}{\partial \theta_i}\left(\frac{p}{pL}\right)\frac{\partial n_1}{\partial p}.IL.p.n_1\right).(L.Ip.In_1) \end{aligned}\tag{9.70}$$

where the second term on the rhs of (9.69) is proportional to n_1 by Euler's theorem (9.23) and hence it vanishes in formula (9.70) by antisymmetry. The coordinates θ_1, θ_2 are just any arbitrary coordinates used to parameterise the 2-d patch enclosed by the loop. We can therefore introduce arbitrary direction vectors of the form,

$$dq_i = d\theta_i \frac{\partial}{\partial \theta_i}\left(\frac{p}{pL}\right)$$

and substitute them into (9.70) to get the terms $d\theta_i \partial n_1/\partial\theta_i$. Then we can substitute these terms into (9.67) for the rotation caused by parallel-transport around the loop,

$$dq = \frac{1}{(pL)^2}\left\{\left[\left(dq_1\frac{\partial n_1}{\partial p}.Il\right).l\right]\left[\left(dq_2\frac{\partial n_1}{\partial p}.Il\right).l\right]\right\}.Iq \tag{9.71}$$

where $l = L.Ip.In_1$ is the subspace of directions. Equation (9.71) gives the rotation due to parallel-transport around a small loop in terms of quantities defined on the physical hypersurface.

In order to get the Riemann tensor, we just have to introduce coordinates. Let's describe the points of the hypersurface in terms of coordinates as $p = p(\theta^1, \ldots, \theta^{n-3})$ as in section 9.8, except that now we label the coordinates with superscripts instead of subscripts because our quantities have to match up with the contravariant and covariant indices used in Riemannian geometry. We can now introduce a set of reference points[32],

$$e_i = \frac{\partial}{\partial \theta^i}\left(\frac{p}{pL}\right) \text{ for } i = 1, \ldots, n-3 \tag{9.72}$$

which all lie in the direction subspace $L.Ip.In_1$. The dual hyperplanes are[33],

$$E^i = (pL)\frac{\partial \theta^i}{\partial p} \text{ for } i = 1, \ldots, n-3 \; . \tag{9.73}$$

These hyperplanes are dual to the reference points because,

$$e_i E^j = (pL)\frac{\partial}{\partial \theta^i}\left(\frac{p}{pL}\right).\frac{\partial \theta^j}{\partial p} = \frac{\partial p}{\partial \theta^i}\frac{\partial \theta^j}{\partial p} - \frac{\partial \ln(pL)}{\partial \theta_i}p\frac{\partial \theta^j}{\partial p} = \delta_i^j - 0 = \delta_i^j \; .$$

In obtaining the above result, we made use of (4.14) to deal with the first term on the rhs and (4.13) to zero the second term. The dual hyperplanes (9.73) can also be written as,

$$E^i = \frac{(-1)^{i-1} e_1 \ldots \check{e}_i \ldots e_{n-3} I e_1 \ldots e_{n-3}}{e_1 \ldots e_{n-3} I e_1 \ldots e_{n-3}} \; . \tag{9.74}$$

It is easily seen that this expression for the E^i makes them dual to the reference points by calculating $e_i E^j$. For $i = j$ the result is unity and for $i \neq j$ the result is zero by antisymmetry.

Having established a system of coordinates, reference points and dual hyperplanes for the direction subspace $l = L.Ip.In_1$ we can resume the quest for the Riemann tensor. The points dq, q_1, q_2 and q in equation (9.71) all lie in the direction subspace $l = L.Ip.In_1$ and so they can all be expanded in terms of the reference points e_i. For example, $q = (q.E^a)e_a$ where the summation convention is on force on pairs of covariant (subscript) and contravariant (superscript) indices. The contravariant components of q are the numbers $(q.E^a)$. In terms of components, (9.71) becomes,

$$dq.E^a = \frac{1}{(pL)^2}\left\{\left[\left(e_b\frac{\partial n_1}{\partial p}.Il\right).l\right]\left[\left(e_c\frac{\partial n_1}{\partial p}.Il\right).l\right]\right\}.Ie_d.E^a(dq_1.E^b)(dq_2.E^c)(q.E^d)$$

and the Riemann tensor is defined in relation to parallel-transport around an infinitesimal loop by,

$$dq.E^a = R^a_{dcb}(dq_1.E^b)(dq_2.E^c)(q.E^d)$$

[32]See section 4.2.3.

[33]See equation (4.10).

and so the relation between the Riemann curvature tensor and the derivative $\partial n_1/\partial p$ is,

$$R^a_{dcb} = \frac{1}{(pL)^2}\left\{\left[\left(e_b\frac{\partial n_1}{\partial p}.Il\right).l\right]\left[\left(e_c\frac{\partial n_1}{\partial p}.Il\right).l\right]\right\}.Ie_d.E^a \ . \tag{9.75}$$

9.9.3 Riemann tensor of a hypersphere

We now compute the Riemann tensor of a hypersphere using equation (9.75) as a check that we have established the correct link between the methods of this chapter and Riemannian geometry.

Equation (7.18) has shown that a hypersphere with centre IN is $S = L + N$ and the radius of the hypersphere depends on the relative weights of the hyperplanes L, N. We can set up a hypersphere of radius ρ with centre IN by weighting L as follows,

$$S = \frac{(-1)^{n-1}(IN.L)\rho^2 L}{2} + N \tag{9.76}$$

and it is easily checked using (7.17) that the radius is ρ. The hypersphere has equation $pS = 0$ so that, from (9.4), the normal is,

$$\begin{aligned} n_1 &= \frac{IS}{\sqrt{SIS}} - \frac{(IS.L)}{\sqrt{SIS}}\left(\frac{p}{pL}\right) \\ &= \frac{(-1)^{n-1}\rho IL}{2} + \frac{IN}{(IN.L)\rho} - \frac{p}{(pL)\rho} \ . \end{aligned} \tag{9.77}$$

The derivative needed for (9.75) is,

$$e_b\frac{\partial n_1}{\partial p} = -\frac{e_a}{(pL)\rho}$$

where we used the fact that $e_bL = 0$ since e_b is in $l = L.Ip.In_1$. In this case the derivative of the normal is automatically in the direction subspace l so that the projection required by (9.75) just changes the weight on e_b.

$$\left[\left(e_b\frac{\partial n_1}{\partial p}.Il\right).l\right] = -\frac{(-1)^{n-1}(pL)e_b}{\rho}$$

Substituting in (9.75) gives,

$$\begin{aligned} R^a_{dcb} &= \rho^{-2}e_be_c.(Ie_d.E^a) \\ &= \rho^{-2}[(e_bIe_d)\delta^a_c - (e_cIe_d)\delta^a_b] = \rho^{-2}[g_{bd}\delta^a_c - g_{cd}\delta^a_b] \end{aligned} \tag{9.78}$$

where we have introduced the components of the *metric tensor* of Riemannian geometry as $g_{ab} = e_a I e_b$. The canonical form of the Riemann tensor for a hypersphere is obtained by lowering the contravariant index with the metric. From (9.78)

$$R_{edcb} = g_{ea} R^a_{dcb} = \rho^{-2}[g_{ec}g_{bd} - g_{eb}g_{cd}]$$

which is in agreement with the Riemann tensor for a hypersphere given in the curvature chapter of the lecture notes [30].

9.9.4 The metric tensor

The metric $g_{ab} = e_a I e_b$ is not important to the theory of a hypersurface in Whitehead's algebra. Nevertheless, it is central in the development of Riemannian geometry, so we develop its properties here.

In section 9.6.1 the condition that a path $p = p(\tau)$ was parameterised by arc-length was shown to be $1 = qIq$ where the direction vector q is given by (9.20). Re-arranging this condition gives,

$$(d\tau)^2 = d\left(\frac{p}{pL}\right) I d\left(\frac{p}{pL}\right) .$$

We write this equation in components by expanding in terms of the coordinates θ^a using the reference points e_a in (9.72),

$$d\left(\frac{p}{pL}\right) = e_a d\theta^a$$

with the summation convention in force. Note that the coordinates θ^a are just arbitrary, they do not have to parameterise arc-length themselves. Then,

$$(d\tau)^2 = d\left(\frac{p}{pL}\right) I d\left(\frac{p}{pL}\right) = (e_a I e_b) d\theta^a d\theta^b = g_{ab} d\theta^a d\theta^b \qquad (9.79)$$

shows how the metric tensor measures arc-length on the physical hypersurface.

In Riemannian geometry, the inverse of the metric tensor, the symmetric tensor g^{ab} also shows up. In our terms this is,

$$g^{ab} = E^a I E b \qquad (9.80)$$

where the E^a are the dual hyperplanes (9.73). IE^a is a point. Using (9.74) we could expand IE^a explicitly in terms of the reference points e_a which means that IE^a also lies in the direction subspace $L.Ip.In_1$. Instead of using

(9.74) directly, we can just rely on the fact that IE^a can be expanded in terms of the e_b and write (summation convention in force),

$$IE^a = (IE^a.E^b)e_b$$

and so,

$$IE^a.Ie_c = e_b = (IE^a.E^b)(e_bIe_c) = (-1)^{n-1}(E^bIE^a)(e_bIe_c) = (-1)^{n-1}g^{ba}g_{bc}$$

and the lhs is $(-1)^{n-1}\delta^a_c$ so that,

$$\delta^a_c = g^{ba}g_{bc}$$

which shows that g^{ab} defined by (9.80) is the inverse of the metric $g_{ab} = e_aIe_b$.

Chapter 10

Application of hypersurface theory: Equations of an empty space-time

10.1 Introduction

Chapter 9 has given a detailed treatment of a curved hypersurface embedded in a flat space. The hypersurface theory looks a bit like the classical differential geometry of a 2-d surface embedded in 3-d space, but it works in higher dimensions. As in the classical differential geometry of a surface, the curvature is characterised by the derivative of the (hyper)surface normal. Now, the standard way to study curvature of higher dimensional manifolds is to use Riemannian geometry, which is the mathematical arena for Einstein's theory of gravity, general relativity. In general relativity, the gravitational field is represented by the curvature of the space-time manifold, and curvature is represented by the Riemann curvature tensor R^a_{dcb}. Nevertheless, in section 9.9, we found the precise relation between the Riemann curvature tensor and the derivative of the hypersurface normal. Therefore, it is possible to translate back and forth between the language of Riemannian geometry and the hypersurface theory in Whitehead's algebra. In these circumstances, it is interesting to try to use the hypersurface theory to set up a theory[1] of the free gravitational field (i.e. a universe without any matter) by assuming that the physical space-time is a curved 4-d hypersurface embedded in a flat 5-d Minkowski space. Gravity would be a manifestation of the curvature of the

[1] In order to avoid any confusion, it should be noted that the theory described in the current chapter has nothing do with A. N. Whitehead's pure particle theory of gravitation [31].

space-time just as in general relativity. An interesting question is whether or not the free field equations of the hypersurface theory would be equivalent to the Einstein equations? This chapter answers this question.

10.2 Einstein's action for the free gravitational field

The field equations of general relativity, like all the field equations in physics, are obtained from the *principle of least action* [29]. The action is an integral S over the space-time of the field quantities. The actual values of the fields are assumed to be such as to produce an extremum of the action. In other words, a variation of the fields produces a change δS in the action and the actual values of the fields are found by requiring $\delta S = 0$. In setting up an action integral, there is not much freedom because one must make the integrand out of fields which result in S being a scalar.

In general relativity the fields themselves are the components of the metric tensor g_{ab}, and the action for the free gravitational field is taken as,

$$S = \int R dV = \int R\sqrt{-g}d\theta^1 \dots d\theta^4 \tag{10.1}$$

where g is the determinant of the metric tensor g_{ab} written in the (arbitrary) coordinates $\theta^1, \dots, \theta^4$ and R is the scalar curvature formed by contracting the Riemann tensor,

$$R = g^{db}R_{db} = g^{db}R^c_{dcb} \ . \tag{10.2}$$

The first contraction of the Riemann tensor is called the *Ricci* tensor $R_{db} = R^c_{dcb}$. The condition $\delta S = 0$ implies the Einstein equations for the free gravitational field,

$$R_{db} = 0 \ . \tag{10.3}$$

The Ricci tensor can be writttten out in terms of the fields g_{ab} and their first and second partial derivatives. Consequently, the field equations $R_{db} = 0$ are in fact a set of non-linear, second order, coupled partial differential equations.

10.3 The action for an empty space-time

The key to setting up the action in the hypersurface theory is the observation made following equation (9.66) that curvature looks particularly simple when studied on the hypersurface in the bulk elliptic space. Recall that this new hypersurface in the bulk elliptic space is the locus of the normal n_1 to the physical space-time. It is the generalisation of Gauss's spherical representation of a 2-d surface. However, Gauss's spherical representation only works for a small patch of the physical surface because the inverse of the normal map is one-to-many. In our case, the normal map (9.4) is one-to-one and section 9.8.2 has shown how to move up to the hypersurface in the bulk elliptic space and back down to the physical hypersurface starting from any point on the physical hypersurface.

So, the mathematics is telling us that the action S should be an integral over the hypersurface in the bulk elliptic space instead of an integral over the physical space-time. The simplest invariant integral over the hypersurface in the bulk elliptic space is its volume. So from (9.55) the action is[2],

$$S = \int \sqrt{dV'.IdV'} = \int d\theta_1 \ldots d\theta_{n-3} \sqrt{\frac{\partial n_1}{\partial \theta_1} \cdots \frac{\partial n_1}{\partial \theta_{n-3}} I \frac{\partial n_1}{\partial \theta_1} \cdots \frac{\partial n_1}{\partial \theta_{n-3}}} \ . \quad (10.4)$$

In our theory, the fields are $n_1 = n_1(\theta_1, \ldots, \theta_{n-3})$ and its derivatives $\partial n_1/\partial \theta_i$. In the case of 4-d space-time, the action integral is over a 4-d hypersurface in 6-d elliptic space so that $n = 7$.

In order to obtain the field equations, we have to vary the hypersurface in the bulk elliptic space and apply the condition $\delta S = 0$. Now, we cannot vary this hypersurface in a completely arbitrary way because the points of the hypersurface must always obey the conditions $n_1 L = 0$ and $n_1 I n_1 = 1$. These conditions were established immediately following the definition of the normal map in equation (9.4). Since the action is really a number-valued function of the hypersurface, the two conditions need to be put in as constraints on the hypersurface,

$$\int (n_1 L)\sqrt{dV'.IdV'} \ = \ 0 \quad (10.5)$$

$$\int (n_1 I n_1 - 1)\sqrt{dV'.IdV'} \ = \ 0 \ . \quad (10.6)$$

We can now introduce two Lagrange multipliers λ and μ and obtain the field equations from an unconstrained extremum of,

$$\begin{aligned} S \ &= \ \int \sqrt{dV'.IdV'} - \lambda \int (n_1 L)\sqrt{dV'.IdV'} - \mu \int (n_1 I n_1 - 1)\sqrt{dV'.IdV'} \\ &= \ \int [1 + \mu - \lambda(n_1 L) - \mu(n_1 I n_1)]\sqrt{dV'.IdV'} \ . \end{aligned} \quad (10.7)$$

[2] It is interesting to note that Schrödinger, writing about Einstein's action integral in chapter XI of his book *Space-Time Structure* [32], first sets up the action as the volume of the physical space-time, and then shows that the resulting equation $g^{ab} = 0$ cannot serve as a field equation. After this failure, he sets up Einstein's action (10.1).

10.4 Euler-Lagrange equations

The field equations are the Euler-Lagrange equations that follow from the condition $\delta S = 0$. Let's see how these look in general. The action is something like,

$$S = \int d\theta_1 \ldots d\theta_{n-3} \mathcal{L}\left(n_1, \frac{\partial n_1}{\partial \theta_1}, \ldots, \frac{\partial n_1}{\partial \theta_{n-3}}\right) \tag{10.8}$$

where $\mathcal{L}$ is known as the *Lagrangian*. We now vary $n_1 \to n_1 + \delta n_1$ so that the derivatives change by,

$$\delta\left(\frac{\partial n_1}{\partial \theta_i}\right) = \frac{\partial \delta n_1}{\partial \theta_i} \ .$$

The variation in the action is,

$$\delta S = \int d\theta_1 \ldots d\theta_{n-3} \left[\delta n_1 \frac{\partial \mathcal{L}}{\partial n_1} + \sum_i \frac{\partial \delta n_1}{\partial \theta_i} \frac{\partial \mathcal{L}}{\partial\left(\frac{\partial n_1}{\partial \theta_i}\right)}\right]$$

where the partial derivatives of $\mathcal{L}$, being derivatives respect to points, are hyperplane-valued as in equation (4.6). Integrating by parts,

$$\delta S = \int d\theta_1 \ldots d\theta_{n-3} \left[\delta n_1 \frac{\partial \mathcal{L}}{\partial n_1} + \sum_i \frac{\partial}{\partial \theta_i}\left\{\delta n_1 \frac{\partial \mathcal{L}}{\partial\left(\frac{\partial n_1}{\partial \theta_i}\right)}\right\} - \sum_i \delta n_1 \frac{\partial}{\partial \theta_i}\left\{\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial n_1}{\partial \theta_i}\right)}\right\}\right]$$

and assuming that the contribution from the boundary term - the first summation on the rhs - vanishes, then,

$$\delta S = \int d\theta_1 \ldots d\theta_{n-3} \delta n_1 \left[\frac{\partial \mathcal{L}}{\partial n_1} - \sum_i \frac{\partial}{\partial \theta_i}\left\{\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial n_1}{\partial \theta_i}\right)}\right\}\right] \ .$$

The only way that $\delta S = 0$ for arbitrary variations δn_1 is if the fields obey the Euler-Lagrange equations,

$$\frac{\partial \mathcal{L}}{\partial n_1} = \sum_i \frac{\partial}{\partial \theta_i}\left\{\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial n_1}{\partial \theta_i}\right)}\right\} \ . \tag{10.9}$$

10.5 The equations of empty space-time

The derivatives $\partial n_1/\partial \theta_i$ are in fact the reference points for the direction subspace[3] at n_1. Using the notation of equation set (4.8), we write these

[3]See footnote 31 of chapter 9.

reference points[4] as,

$$e_i = \frac{\partial n_1}{\partial \theta_i} \text{ for } i = 1, \ldots, n-3 \ . \tag{10.10}$$

Let us set up hyperplanes dual to these reference points. We have already done something very similar from equation (9.72) onwards. The difference was only that we were working on the physical space-time and now we are working on the hypersurface in the bulk elliptic space. So, the dual hyperplanes are,

$$E_i = \frac{\partial \theta_i}{\partial n_1} \text{ for } i = 1, \ldots, n-3$$

and $e_i E_j = \delta_{ij}$. Furthermore, by a similar argument to the one that led to (9.74), the dual hyperplanes can also be written,

$$E_i = \frac{\partial \theta_i}{\partial n_1} = \frac{(-1)^{i-1} e_1 \ldots \check{e}_i \ldots e_{n-3} I e_1 \ldots e_{n-3}}{e_1 \ldots e_{n-3} I e_1 \ldots e_{n-3}} \ . \tag{10.11}$$

The Euler-Lagrange equations (10.9) now take the form,

$$\frac{\partial \mathcal{L}}{\partial n_1} = \sum_i \frac{\partial}{\partial \theta_i} \left\{ \frac{\partial \mathcal{L}}{\partial e_i} \right\} \ . \tag{10.12}$$

From (10.4), the volume element can now be written,

$$\sqrt{dV'.IdV'} = d\theta_1 \ldots d\theta_{n-3} \sqrt{e_1 \ldots e_{n-3} I e_1 \ldots e_{n-3}}$$

and by comparing the unconstrained action (10.7) with the general form (10.8), the Lagrangian for empty space-time is,

$$\mathcal{L} = [1 + \mu - \lambda(n_1 L) - \mu(n_1 I n_1)] \sqrt{e_1 \ldots e_{n-3} I e_1 \ldots e_{n-3}} \ . \tag{10.13}$$

We can now substitute the Lagrangian into the Euler-Lagrange equations (10.12). The most problematical derivative is,

$$\frac{\partial}{\partial e_i} \sqrt{e_1 \ldots e_{n-3} I e_1 \ldots e_{n-3}} = \frac{(-1)^{i-1} e_1 \ldots \check{e}_i \ldots e_{n-3} I e_1 \ldots e_{n-3}}{\sqrt{e_1 \ldots e_{n-3} I e_1 \ldots e_{n-3}}} \ . \tag{10.14}$$

As a check on this expression, we note that,

$$\sqrt{e_1 \ldots e_{n-3} I e_1 \ldots e_{n-3}}$$

[4]Notice that n_1 is just e_0 in the notation of equation (4.8).

is a homogeneous function of degree 1 in the variable point e_i. Hence, from Euler's theorem (4.12), we must get,

$$e_i \frac{\partial}{\partial e_i} \sqrt{e_1 \dots e_{n-3} I e_1 \dots e_{n-3}} = \sqrt{e_1 \dots e_{n-3} I e_1 \dots e_{n-3}}$$

with no summation. This result is also obtained by multiplying (10.14) on the left by e_i so that (10.14) makes sense. The Euler-Lagrange equations are now,

$$\begin{aligned} &[-\lambda L - 2\mu I n_1] \sqrt{e_1 \dots e_{n-3} I e_1 \dots e_{n-3}} \\ = \sum_i \frac{\partial}{\partial \theta_i} &\left\{ [1 + \mu - \lambda(n_1 L) - \mu(n_1 I n_1)] E_i \sqrt{e_1 \dots e_{n-3} I e_1 \dots e_{n-3}} \right\} \end{aligned}$$

where we used (10.11). The factor between the square brackets on the rhs simplifies to unity when we employ the constraints $n_1 L = 0$ and $n_1 I n_1 = 1$. Hence, the field equations for an empty space-time are,

$$[-\lambda L - 2\mu I n_1] \sqrt{e_1 \dots e_{n-3} I e_1 \dots e_{n-3}} = \sum_i \frac{\partial}{\partial \theta_i} \left\{ E_i \sqrt{e_1 \dots e_{n-3} I e_1 \dots e_{n-3}} \right\} \tag{10.15}$$

where the Lagrange multipliers λ, ν are to be found from the constraints (10.5,10.6).

10.6 Solution of the field equations

The field equations (10.15) are a set of coupled, non-linear second-order partial differential equations and so any systematic attempt to find solutions seems out of the question. The only way forward seems to set up a trial solution based on physical intuition and substitute it into the field equations in order to check whether or not the trial solution is, in fact, a solution. Now in chapter 5 it was argued that the absence of dilatations in our world implies that space-time cannot be Minkowskian. It was then shown that the simplest space-time consistent with the absence of dilatations is elliptic de Sitter space. So, based on this argument, we will check whether or not de Sitter space is a solution of the field equations.

10.6.1 The trial solution

In chapter 5, we studied de Sitter space as the ultra-infinite points of hyperbolic geometry. However, in physics [17, 18, 19, 20], 4-d de Sitter space is always regarded as being a 4-d hypersphere embedded in a 5-d Minkowski space. This happens to be the viewpoint used in deriving the field equations, the 4-d curved space-time is regarded as being embedded in a 5-d flat Minkowski space.

We start with the reference points $a_1, \ldots, a_7$ of the bulk 6-d elliptic space. The dual hyperplanes are $A_1, \ldots, A_7$ with $a_i A_j = \delta_{ij}$. The elliptic polarity is $Ia_i = A_i$. The origin is $IN = a_6 + ia_7$ and the point at infinity is $IL = a_6 - ia_7$.

The next task is to set up the 5-d Minkowski space on the absolute quadric of the 6-d bulk elliptic space. From (3.50), a general point on the quadric is,

$$p = \exp(ILM.I)IN = IN + (IN.L)(IM - \frac{1}{2}(M.IM)IL) \tag{10.16}$$

where the generator IM lies in the subspace $LN = 2ia_1a_2a_3a_4a_5$. We write the generator in the form,

$$IM = x_1a_1 + x_2a_2 + x_3a_3 + x_4a_4 + x_5ia_5 \tag{10.17}$$

where the last reference point is given the imaginary weight. This has the effect of making the flat 5-d space on the quadric into a Minkowski space. To see this, we use equation (3.52). This equation gives the squared distance from the origin IN to point p as MIM. From (10.17),

$$MIM = x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_5^2 \tag{10.18}$$

is clearly the formula for distance in a 5-d Minkowski space.

The physical de Sitter space-time will be a 4-d hypersphere of radius ρ and centre IN. Equation (9.76) can be used without modification.

$$S = \frac{(IN.L)\rho^2 L}{2} + N \tag{10.19}$$

The equation of the de Sitter space is $pS = 0$. Substituting for p and S using (10.16) and (10.19) gives,

$$\begin{aligned} 0 &= pS = [IN + (IN.L)(IM - \frac{1}{2}(M.IM)IL)] \left[\frac{(IN.L)\rho^2 L}{2} + N\right] \\ &= \frac{(IN.L)^2}{2}(\rho^2 - MIM) \end{aligned}$$

where we used $IL.L = 0$, $IN.N = 0$, $IM.L = 0$, $IM.N = 0$. Hence the equation of the de Sitter space is $\rho^2 = MIM$ or,

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_5^2 = \rho^2 \tag{10.20}$$

using (10.18). This is the equation used in [17, 18, 19, 20] to embed 4-d de Sitter space in a 5-d Minkowski space. From equation (5.13), the natural coordinates for de Sitter space are τ, χ, θ, ϕ where,

$$\begin{aligned}
x_1 &= \rho \cosh\tau \cos\chi \\
x_2 &= \rho \cosh\tau \sin\chi \sin\theta \cos\phi \\
x_3 &= \rho \cosh\tau \sin\chi \sin\theta \sin\phi \\
x_4 &= \rho \cosh\tau \sin\chi \cos\theta \\
x_5 &= \rho \sinh\tau \ .
\end{aligned} \tag{10.21}$$

The x_i defined by (10.21) automatically satisfy (10.20).

The normal to the hypersphere is given by (9.77).

$$\begin{aligned}
n_1 &= \frac{IS}{\sqrt{SIS}} - \frac{(IS.L)}{\sqrt{SIS}}\left(\frac{p}{pL}\right) \\
&= \frac{\rho IL}{2} + \frac{IN}{(IN.L)\rho} - \frac{p}{(pL)\rho} \ .
\end{aligned}$$

From (10.16), for p on the hypersphere so that $MIM = \rho^2$,

$$\frac{p}{pL} = \frac{IN}{IN.L} + IM - \frac{\rho^2}{2}IL$$

and so the normal to the hypershere is,

$$n_1 = \frac{\rho IL}{2} + \frac{IN}{(IN.L)\rho} - \frac{IN}{(IN.L)\rho} - \frac{IM}{\rho} + \frac{\rho}{2}IL = \rho IL - \frac{IM}{\rho} \ .$$

The trial solution $n_1 = n_1(\tau, \chi, \theta, \phi)$ is obtained by substituting for IM using (10.17) and the coordinates (10.21). It is,

$$\begin{aligned}
n_1 = \ & \rho IL - a_1 \cosh\tau \cos\chi - a_2 \cosh\tau \sin\chi \sin\theta \cos\phi \\
& - a_3 \cosh\tau \sin\chi \sin\theta \sin\phi - a_4 \cosh\tau \sin\chi \cos\theta \\
& - i a_5 \sinh\tau \ .
\end{aligned} \tag{10.22}$$

10.6.2 Reference points and dual hyperplanes

In order to check whether or not (10.22) is a solution of the field equations (10.15) we need to evaluate the reference points (10.10) for the direction subspace at n_1 and their dual hyperplanes (10.11).

The reference points are obtained by differentiating (10.22).

$$\begin{aligned}
e_1 &= \frac{\partial n_1}{\partial \tau} = -a_1 \sinh\tau\cos\chi - a_2\sinh\tau\sin\chi\sin\theta\cos\phi \\
&\quad -a_3\sinh\tau\sin\chi\sin\theta\sin\phi - a_4\sinh\tau\sin\chi\cos\theta \\
&\quad -ia_5\cosh\tau \\
e_2 &= \frac{\partial n_1}{\partial \chi} = a_1\cosh\tau\sin\chi - a_2\cosh\tau\cos\chi\sin\theta\cos\phi \\
&\quad -a_3\cosh\tau\cos\chi\sin\theta\sin\phi - a_4\cosh\tau\cos\chi\cos\theta \\
e_3 &= \frac{\partial n_1}{\partial \theta} = -a_2\cosh\tau\sin\chi\cos\theta\cos\phi \\
&\quad -a_3\cosh\tau\sin\chi\cos\theta\sin\phi + a_4\cosh\tau\sin\chi\sin\theta \\
e_4 &= \frac{\partial n_1}{\partial \phi} = a_2\cosh\tau\sin\chi\sin\theta\sin\phi \\
&\quad -a_3\cosh\tau\sin\chi\sin\theta\cos\phi
\end{aligned} \tag{10.23}$$

In order to obtain the dual hyperplanes E_i it helps to know the matrix e_iIe_j. Upon evaluating the matrix using the set (10.23), it is found to be diagonal. The non-zero elements are given below.

$$\begin{aligned}
e_1Ie_1 &= -1 \\
e_2Ie_2 &= \cosh^2\tau \\
e_3Ie_3 &= \cosh^2\tau\sin^2\chi \\
e_4Ie_4 &= \cosh^2\tau\sin^2\chi\sin^2\theta
\end{aligned} \tag{10.24}$$

The ubiquitous factor is therefore,

$$e_1e_2e_3e_4Ie_1e_2e_3e_4 = -\cosh^6\tau\sin^4\chi\sin^2\theta\ . \tag{10.25}$$

The dual hyperplanes can now be obtained from (10.11).

$$\begin{aligned}
E_1 &= \frac{e_2e_3e_4Ie_1e_2e_3e_4}{e_1e_2e_3e_4Ie_1e_2e_3e_4} = \frac{Ie_1}{e_1Ie_1} = -Ie_1 \\
E_2 &= -\frac{e_1e_3e_4Ie_1e_2e_3e_4}{e_1e_2e_3e_4Ie_1e_2e_3e_4} = \frac{Ie_2}{e_2Ie_2} = \frac{Ie_2}{\cosh^2\tau} \\
E_3 &= \frac{e_1e_2e_4Ie_1e_2e_3e_4}{e_1e_2e_3e_4Ie_1e_2e_3e_4} = \frac{Ie_3}{e_3Ie_3} = \frac{Ie_3}{\cosh^2\tau\sin^2\chi} \\
E_4 &= -\frac{e_1e_2e_3Ie_1e_2e_3e_4}{e_1e_2e_3e_4Ie_1e_2e_3e_4} = \frac{Ie_4}{e_4Ie_4} = \frac{Ie_4}{\cosh^2\tau\sin^2\chi\sin^2\theta}
\end{aligned} \tag{10.26}$$

10.6.3 Testing the trial solution

The field equations (10.15) are now,

$$
\begin{aligned}
[-\lambda L - 2\mu In_1]\cosh^3\tau\sin^2\chi\sin\theta &= -\frac{\partial}{\partial\tau}[(\cosh^3\tau\sin^2\chi\sin\theta)Ie_1] \\
+\frac{\partial}{\partial\chi}[(\cosh\tau\sin^2\chi\sin\theta)Ie_2] &+ \frac{\partial}{\partial\theta}[(\cosh\tau\sin\theta)Ie_3] \\
&+ \frac{\partial}{\partial\phi}\left[\frac{\cosh\tau}{\sin\theta}Ie_4\right] \qquad (10.27)
\end{aligned}
$$

so we need the derivatives of the reference points. These derivatives can be obtained in terms of the points $a_1,\ldots,a_5$ by differentiating set (10.23). However, it is more instructive to express these derivatives in terms of the reference points $IL, n_1, e_1, \ldots, e_4$ which are the physically relevant ones for the hypersurface in the bulk elliptic space. In order to do this, we have to be able to express the a_i in terms of the reference points for the hypersurface in the bulk elliptic space. We start by expressing a_1 in these terms. Write,

$$a_1 = \eta IL + \xi_0 n_1 + \xi_1 e_1 + \xi_2 e_2 + \xi_3 e_3 + \xi_4 e_4 \ . \qquad (10.28)$$

The coefficient ξ_0 is found by evaluating $a_1 In_1$. From (10.28), $a_1 In_1 = \xi_0$ since $n_1 L = 0$ and the e_i are in the direction subspace $L.Ip.In_1$ so $e_i In_1 = 0$. So, the coefficient $\xi_0 = n_1 Ia_1$ can be evaluated from equation (10.22). The result is $\xi_0 = -\cosh\tau\cos\chi$. The coefficient ξ_1 can be found by evaluating $a_1 Ie_1$ using (10.28). Thus, $a_1 Ie_1 = \xi_1(e_1 Ie_1)$ using the fact that the matrix $e_i Ie_j$ is diagonal. Hence, $\xi_1 = e_1 Ia_1/e_1 Ie_1$ and $e_1 Ia_1$ is evaluated from the set (10.23). The result is $\xi_1 = \sinh\tau\cos\chi$. The coefficients ξ_2, ξ_3 and ξ_4 are found in the same way. The points $a_1,\ldots,a_5$ all lie in the subspace LN. Consequently, $e_i N = 0$ using set (10.23). The remaining coefficient η can now be found by evaluating $a_1 N$ using (10.28). Clearly, $0 = \eta IL.N + \xi_0 n_1 N$, and $n_1 N = \rho IL.N$ from (10.22). Hence $\eta = -\rho\xi_0$ ξ_0 is already known so that $\eta = \rho\cosh\tau\cos\chi$. Thus,

$$a_1 = (IL\rho - n_1)\cosh\tau\cos\chi + e_1\sinh\tau\cos\chi + e_2\frac{\sin\chi}{\cosh\tau} \ . \qquad (10.29)$$

The expansions of the other points can be worked out by the same method and the results are as follows.

$$
\begin{aligned}
a_2 &= (IL\rho - n_1)\cosh\tau\sin\chi\sin\theta\cos\phi + e_1\sinh\tau\sin\chi\sin\theta\cos\phi \\
&\quad -e_2\frac{\cos\chi\sin\theta\cos\phi}{\cosh\tau} - e_3\frac{\cos\theta\cos\phi}{\cosh\tau\sin\chi} + e_4\frac{\sin\phi}{\cosh\tau\sin\chi\sin\theta}
\end{aligned}
$$

$$
\begin{aligned}
a_3 &= (IL\rho - n_1)\cosh\tau\sin\chi\sin\theta\sin\phi + e_1\sinh\tau\sin\chi\sin\theta\sin\phi \\
&\quad -e_2\frac{\cos\chi\sin\theta\sin\phi}{\cosh\tau} - e_3\frac{\cos\theta\sin\phi}{\cosh\tau\sin\chi} - e_4\frac{\cos\phi}{\cosh\tau\sin\chi\sin\theta} \\
a_4 &= (IL\rho - n_1)\cosh\tau\sin\chi\cos\theta + e_1\sinh\tau\sin\chi\cos\theta \\
&\quad -e_2\frac{\cos\chi\cos\theta}{\cosh\tau} + e_3\frac{\sin\theta}{\cosh\tau\sin\chi}
\end{aligned}
\tag{10.30}
$$

We can now calculate the derivatives from set (10.23) and substitute for the a_i using (10.29, 10.30). After some tedious working[5],

$$
\begin{aligned}
\frac{\partial e_1}{\partial\tau} &= -(IL\rho - n_1) \\
\frac{\partial e_2}{\partial\chi} &= (IL\rho - n_1)\cosh^2\tau + e_1\sinh\tau\cosh\tau \\
\frac{\partial e_3}{\partial\theta} &= (IL\rho - n_1)\cosh^2\tau\sin^2\chi + e_1\sinh\tau\cosh\tau\sin^2\chi - e_2\sin\chi\cos\chi \\
\frac{\partial e_4}{\partial\phi} &= (IL\rho - n_1)\cosh^2\tau\sin^2\chi\sin^2\theta + e_1\sinh\tau\cosh\tau\sin^2\chi\sin^2\theta \\
&\quad -e_2\sin\chi\cos\chi\sin^2\theta - e_3\sin\theta\cos\theta\ .
\end{aligned}
\tag{10.31}
$$

The partial derivatives commute with the polarity, so that, for example,

$$
\frac{\partial Ie_1}{\partial\tau} = I\left(\frac{\partial e_1}{\partial\tau}\right)\ .
$$

Using this rule and the formulae (10.31), most of the terms on the rhs of the field equations (10.27) cancel out to leave,

$$
(-\lambda L - 2\mu I n_1)\cosh^3\tau\sin^2\chi\sin\theta = (L\rho - In_1)4\cosh^3\tau\sin^2\chi\sin\theta\ . \tag{10.32}
$$

If we had solved the field equations (10.15) in a systematic way, the solution n_1 would have depended on the two Lagrange multipliers λ, μ as parameters of the solution. These parameters would then be found from substituting the solution into the contraints (10.5,10.6). However, in our case the trial solution (10.22) already satisfies the constraints, so that the constraints have to be chosen to ensure that the trial solution satisfies the field equations (10.32); fortunately this can be done by choosing $\lambda = -4\rho$ and $\mu = 2$. Thus our physical intuition was correct and it turns out that de Sitter space is a solution of the field equations (10.15) of empty space. The points p of the physical de Sitter space can be recovered from the solution (10.22) using equation (9.60).

[5] All the formulae in this section were checked numerically using the Scilab package in appendix A.

10.7 Conclusion

De Sitter space is not a solution of the Einstein equations $R_{bd} = 0$ for empty space. Therefore, the equations (10.15) that have been set up on the basis of the hypersurface theory of chapter 9 are not equivalent to the Einstein equations for a universe with no matter. In order to get de Sitter space as a solution, the Einstein action (10.1) has to be modified to include a cosmological term proportional to $\int dV$. Since the universe is now thought [20] to be asymptotic to de Sitter space into the future, the cosmological term must presumably be included in the Einstein action.

In reference [29], Landau and Lifshitz make the following remark.

> The general theory of relativity . . . represents probably the most beautiful of all existing physical theories.

When the writer of this text first read about general relativity, it was easy to agree wholeheartedly with this statement. However, recently, the writer has tended to view the Einstein action as a fairly arbitrary creation which is not particularly compelling. For example, the scalar curvature (10.2) has the following formula in Whitehead's algebra,

$$R = g^{db} R^c_{dcb} = \frac{(-1)^{n-1}}{(pL)^2} \left\{ \left[\left(I \frac{\partial}{\partial q_2} \frac{\partial n_1}{\partial p} .Il \right) .l \right] \left[\left(q_1 \frac{\partial n_1}{\partial p} .Il \right) .l \right] \right\} .Iq_2 . \frac{\partial}{\partial q_1}$$

where we used the trace operation from section 4.4.3 and the inverse of the metric tensor from (9.80). It is difficult to envisage how this sort of expression could arise on physical grounds. In contrast, the action (10.4) in the hypersurface theory is the simplest one possible.

The foregoing remarks are not meant to suggest that the field equations (10.15) are particularly important. The writer's view is that this chapter serves a heuristic purpose. To repeat a part of the introduction, given a theory of a hypersurface embedded in a flat space, it is natural to try to set up a theory - along the lines of general relativity - in which gravity is a manifestation of space-time curvature. This chapter has hopefully shown that this can be done in a very natural way with the result that the empty space-time is de Sitter. Now earlier, in chapter 5, it had already been argued that the absence of dilatations in our experience implies that space-time is de Sitter, and now we find that if we try to set up a theory of a curved space-time, we are led straight back to de Sitter. In other words, the lesson of this chapter supports the general feeling that one gets from studying space-time from the Cayley-Klein approach to geometry; physics should be studied on a background of a de Sitter space. This seems an appropriate place to end this text on Whitehead's geometric algebra.

Appendix A

Scilab function package for Whitehead's algebra

A.1 Introduction

Throughout the main text, the theoretical material has been augmented by numerical examples using the implementation of Whitehead's algebra in Yorick. The Yorick functions are based on fixed arrays that implement Whitehead's algebra in 4-d elliptic space with complex weights[1]. This fixed structure has been set up to handle 3-d Euclidean geometry by modelling it on the absolute quadric hypersurface of the 4-d elliptic space as explained in chapter 7. The aim with the Yorick functions has been to make an efficient means for programming engineering applications in computational geometry and computer vision. Nevertheless, the ability to use complex weights in 4-d elliptic space has meant that we have also been able to do calculations in de Sitter space[2] with the Yorick functions. In spite of this, it is clear that a general numerical implementation of Whitehead's algebra for (n-1)-dimensional elliptic space with complex weights would be a useful tool. So, this appendix describes some functions that do this written for the scientific interpreter *Scilab* [26]. They can be downloaded from the site [21]. These functions have similar names to the Yorick functions, except that they mostly have an argument `n` for the number of reference points needed for the projective space. The functions use Scilab's `list` construct to store the elements of the algebra. Consequently, the Scilab functions are much slower than the Yorick functions, but that is the trade-off for being able to work in higher dimensions.

[1]See section 7.3.

[2]See section 7.3.4.

A.2 Representation of simple elements of Whitehead's algebra

In this appendix, the examples are from the diary of a Scilab session. So, we first load the Scilab function packages.

```
-->clear;
-->getf('Whitehead.sci');
-->getf('Elliptic.sci');
```

A simple element is a list. The first slot in the list is a string which is either `'point'` or `'hyperplane'`. It says whether the simple element is represented by products of points or intersections of hyperplanes as in section 1.9. The next slot in the list is the weight. This is a real or complex number. The remaining slots are filled with the labels of the reference points or dual hyperplanes that form the element. In the following example we create the element $X = 2a_{10}a_6a_5a_2a_9a_3$ in point form using a list and then print it out to the terminal.

```
-->X=list('point',2.0,10,6,5,2,9,3);
-->cprint_element(X);
+2.000a10a6a5a2a9a3
```

In the next example we create an element $Y = 3A_3A_6A_7$ as the intersection of hyperplanes.

```
-->Y=list('hyperplane',3.0,3,6,7);
-->cprint_element(Y);
+3.000A3A6A7
```

Although X was defined as a product of points, we can change its representation to an intersection of hyperplanes with the function `hyperplane_form`.

```
-->n=10;
-->cprint_element(hyperplane_form(n,X));
-2.000A1A4A7A8
```

Notice that function `hyperplane_form` needs to know the number of reference points needed for the projective space. In the above examples, we assume that there are $n = 10$ reference points $a_1, \ldots, a_{10}$ so that the projective space is 9-dimensional. Similarly, Y, which was defined as an intersection of hyperplanes, can be represented as a product of points.

```
-->cprint_element(point_form(n,Y));
+3.000a1a2a4a5a8a9a10
```

A.3 Representation of compound elements

A compound element is a linear combination of elements of the form,

```
X=list('sum',w,E1,E2,...,Er)
```

which represents $X = w\sum_i E_i$ where the E_i are elements which are always the same grade in Whitehead's algebra. The first slot in the list is the string `'sum'`. The second slot is for an overall weight. The remaining slots are filled with lists representing other elements of the algebra. So, suppose we need to set up the point $p = 2a_1 + 3a_2$. The way to do this is to first create some variables that represent the reference points a_1 and a_2 and then to form the point with a compound list.

```
-->a1=list('point',1.0,1);
-->a2=list('point',1.0,2);
-->p=list('sum',1.0,list('sum',2.0,a1),list('sum',3.0,a2));
-->cprint_element(p);
-->cprint_element(p);
+2.000a1+3.000a2
```

A.4 The antisymmetric product

The Scilab package has a function `wap(n,X,Y)` that returns the antisymmetric product XY of two elements X and Y. This function does the same thing as the `wap` function for the Yorick package described in section 7.3.3, except that the Scilab function needs the argument `n` for the grade of the projective space. There are examples of the use of the `wap` function later in this appendix.

A.5 The elliptic polarity

The Scilab package has a function `I(n,X)` that implements the elliptic polarity I. This function does the same thing as the `I` function for the Yorick package described in section 7.3.2. except that the Scilab function needs the argument `n` for the grade of the space. The action of Scilab's elliptic polarity on reference points is,

$$Ia_j = A_j$$

and this differs from the formula (3.11) that we gave for the elliptic polarity by an unimportant weight $(-1)^{n-1}$. The reason for this trivial difference is that the Whitehead algebra package for Scilab was developed when the theory in this text was in its infancy and before the argument in section 3.4 was written to motivate the introduction of the elliptic polarity. In order to check this, we first make reference point a_1 and the dual hyperplane A_1.

```
-->a1=list('point',1.0,1);
-->cprint_element(a1);
+1.000a1
-->A1=list('hyperplane',1.0,1);
-->cprint_element(A1);
+1.000A1
```

Notice that $a_1 A_1 = 1$ as expected.

```
-->n=6;
-->cprint_element(wap(n,a1,A1));
+1.000
```

Now check the action of the elliptic polarity.

```
-->cprint_element(I(n,a1));
+1.000A1
-->cprint_element(I(n,A1));
-1.000a1
```

The above example was for $n = 6$, but for $n = 7$ we get the following.

```
-->n=7;
-->cprint_element(I(n,a1));
+1.000A1
-->cprint_element(I(n,A1));
+1.000a1
```

A.6 Elliptic congruence

The Scilab package has a function `congruence(n,theta,ILM,X)` that returns the action of the canonical form (3.26) of a simple congruence. In other words, the Scilab function calculates,

$$f_\tau X = f_{i\theta} X = \exp\left(\frac{\tau ILM.I}{\sqrt{-ILM.LM}}\right) X = \exp\left(\frac{\theta ILM.I}{\sqrt{ILM.LM}}\right) X$$

in which the congruence parameter $\tau = i\theta$ where $i = \sqrt{-1}$, ILM is the invariant line of the congruence, and X is an arbitrary element of the algebra.

This function does the same thing as the `con` function for the Yorick package described in section 7.3.4, except that the Scilab function needs the argument `n` for the grade of the space and the congruence parameter is θ instead of τ. The change in the definition of the congruence parameter is because the `congruence` function was written in the early development of the theory.

A.7 Distance in elliptic space

The Scilab package also has the function `theta=distance(n,a,b)` which returns the distance given by equation (3.32). The distance returned by the Scilab function is the number θ where $\tau = i\theta$ in equation (3.32).

A.8 Other functions

We have already seen that the function `cprint_element` will print an element of the algebra to the terminal. There are also functions `print_element` and `fprint_element`, but they only handle elements with real-valued weights, whereas `cprint_element` handles complex weights. The function `fprint_element` prints to a file.

Sometimes it may be necessary to use the function `tidy_up(n,X)`. In the following example, we form an element $IS_1 + IS_2$ using variables IS_1 and IS_2 that appear later in this appendix.

```
-->cprint_element(list('sum',1.0,IS1,IS2));
-0.816ia1+0.246a2-0.067a3+1.154a4-0.095a5-0.510a6-0.816ia1-0.613a2
-0.545a3-0.558a4-0.257a5-0.785a6
```

Notice that the expression has not been simplified. This can be cured by the use of `tidy_up`.

```
-->cprint_element(tidy_up(n,list('sum',1.0,IS1,IS2)));
-1.633ia1-0.367a2-0.612a3+0.597a4-0.351a5-1.295a6
```

It is not often necessary to use this function, because most of the functions in the packages `Whitehead.sci` and `Elliptic.sci` call `tidy_up` prior to returning.

A.9 Example: Calculation of touching hyperspheres in 4-d Euclidean space

In this section we give an extended example of the use of the Scilab functions in higher dimensional geometry. The example involves numerical calculations of mutually touching hyperspheres in 4-d Euclidean space. The example builds on the theoretical material of section 7.6.1 on Soddy's formula.

A.9.1 Setting up the mutually touching hyperspheres

In this section, we work generally at first by considering n mutually touching hyperspheres $S_1, \ldots, S_n$ in (n-2)-d Euclidean space. As in chapter 7, the Euclidean space is modelled as the absolute quadric hypersurface of (n-1)-d elliptic space. We start with a set of reference points $a_1, \ldots, a_n$ for the elliptic space. The dual hyperplanes are $A_1, \ldots, A_n$ so that $a_i A_j = \delta_{ij}$. In the Scilab package, the elliptic polarity is $Ia_i = A_i$. Note that this is trivially different from the formula (3.11) as explained in section A.5. In the model of Euclidean geometry on the absolute quadric of elliptic space, the hyperspheres S_i are in fact hyperplanes in the elliptic space[3]. The IS_i are therefore points. They can be taken as a new set of reference points for the elliptic space. We can set them up by transforming the a_i using some collineation[4] B as $IS_i = Ba_i$. The matrix elements are,

$$\begin{aligned} S_i I S_j &= (-1)^{n-1} I^2 S_i . I S_j = (-1)^{n-1} I B a_i . B a_j = B a_j . I B a_i \\ &= B a_i . I B a_j = a_i . B^{-1} I B a_j = a_i J a_j \end{aligned} \tag{A.1}$$

where the new polarity $J = B^{-1} I B$. Now, in section 7.6.1 it is shown that the condition for the hyperspheres to be mutually touching is that the diagonal matrix elements can be taken as $S_i I S_i = 1$ and the off-diagonal elements as $S_i I S_j = -1$. This means that the polarity J is fixed because its matrix elements $a_i J a_j$ are known from equation (A.1). From equation (3.66), we can write the collineation as $B = V^{-1} D^{1/2} U$ where U and V are congruences[5] and D is diagonal so that $DA_i = \lambda_i A_i$. Substituting for B in $J = B^{-1} I B$ gives,

$$J = U^{-1} D^{-1/2} V I V^{-1} D^{1/2} U = U^{-1} D^{-1/2} I D^{1/2} U \ .$$

Multiplying by U^{-1} on the right gives,

$$J U^{-1} = U^{-1} D^{-1/2} I D^{1/2} \ . \tag{A.2}$$

[3] See section 7.4.

[4] B is not a congruence because it does not commute with the absolute polarity I.

[5] In other words U and V both commute with the absolute polarity I.

This is an eigenvalue equation. We can see this by writing it as a matrix equation.

$$a_i.JU^{-1}a_j = a_i.U^{-1}D^{-1/2}ID^{1/2}a_j \ .$$

It is easy to see that $Da_i = \lambda_i^{-1}a_i$ since,

$$1 = a_iA_i = D(a_iA_i) = Da_i.DA_i = \lambda_i^{-1}a_i.\lambda_iA_i = a_iA_i$$

with the summation convention suspended. The rhs of the matrix equation is now,

$$a_i.U^{-1}D^{-1/2}ID^{1/2}a_j = \lambda_j^{-1/2}a_i.U^{-1}D^{-1/2}Ia_j = \lambda_j^{-1/2}a_i.U^{-1}D^{-1/2}A_j = \lambda_j^{-1}a_i.U^{-1}A_j$$

and the lhs is,

$$\begin{aligned} a_i.JU^{-1}a_j &= \sum_k(a_iJa_k)(U^{-1}a_j.A_k) = \sum_k(a_iJa_k)(U^{-1}a_j.Ia_k) \\ &= \sum_k(a_iJa_k)(a_k.IU^{-1}a_j) = \sum_k(a_iJa_k)(a_k.U^{-1}Ia_j) = \sum_k(a_iJa_k)(a_k.U^{-1}A_j) \ . \end{aligned}$$

Putting both sides together gives the matrix form of equation (A.2).

$$\sum_k(a_iJa_k)(a_k.U^{-1}A_j) = \lambda_j^{-1}(a_i.U^{-1}A_j) \ . \qquad \text{(A.3)}$$

This is a matrix eigenvalue equation. The eigenvectors are the columns of matrix $a_i.U^{-1}A_j$ with eigenvalues λ_j^{-1}. So, by setting up the matrix a_iJa_j with the known values S_iIS_j from (A.1), we can solve the matrix eigenvalue equation (A.3) and make the diagonal matrix D and the congruence U from the eigenvalues and eigenvectors respectively. Then, the new reference points are $IS_i = Ba_i = V^{-1}D^{1/2}Ua_i$. Notice that they are only defined modulo an arbitrary congruence V. In section 7.6.1 it was pointed out that the radii of the hyperspheres are left unspecified until the point at infinity IL is fixed with respect to the reference points IS_i determined by the hyperspheres. This is just another way of saying that the points IS_i are only defined up to an arbitrary congruence V. The reason is that the signed curvatures of the hyperspheres are shown in section 7.6.1 to be $IS_i.L$. We can therefore write the curvatures as,

$$IS_i.L = V^{-1}D^{1/2}Ua_i.L = D^{1/2}Ua_i.VL \ .$$

This equation can be interpreted as meaning that the hyperspheres are fixed by $IS_i = D^{1/2}Ua_i$, but now the point at infinity IL has been moved to an arbitrary position $IVL = VIL$ on the quadric by the congruence[6] V. Setting $V = 1$, since we can always move IL later, the reference points fixed by the hyperspheres are,

$$IS_i = D^{1/2}Ua_i = \sum_k D^{1/2}a_k(Ua_i.A_k) = \sum_k \lambda_k^{-1/2}(a_i.U^{-1}A_k)a_k \ . \qquad (A.4)$$

We now give a Scilab program named `Soddy.sci` which solves the matrix eigenvalue equation (A.3) and calculates the reference points IS_i using (A.4) in 4-d Euclidean space. The presentation of the program is broken up into chunks of code in order to illustrate the use of the Scilab packages. The list construct is a little cumbersome, so the first task is to set up some variables for the reference points $a_1, \ldots, a_6$ of the 5-d elliptic space. The 4-d Euclidean geometry will take place on the 4-d absolute quadric hypersurface in the elliptic space.

```
getf('Whitehead.sci');
getf('Elliptic.sci');
//Define the reference points for 5-d elliptic space
n=6;
sgn=(-1)^(n-1);//Ubiquitous sign factor
a1=list('point',1,1);
a2=list('point',1,2);
a3=list('point',1,3);
a4=list('point',1,4);
a5=list('point',1,5);
a6=list('point',1,6);
```

After running the program to this point, Scilab understands the reference points.

```
-->cprint_element(a6);
+1.000a6
```

The next part of the program solves matrix eigenvalue equation (A.3). The Scilab function `spec` returns the matrix elements $a_i.U^{-1}A_j$ in the matrix `Uinv` and the eigenvalues λ^{-1} are along the diagonal of matrix `Dinv`. The matrix elements `c(i,k)` are the coefficients $\lambda_k^{-1/2}(a_i.U^{-1}A_k)$ in equation (A.4).

[6]Of course, V is a congruence, since it commutes with I, but it is not an element of the Euclidean group of congruences that endow the quadric with a Euclidean geometry in a Kleinian sense because V does not leave the point at infinity IL invariant.

```
//Set up the J matrix.
J=[[1,-1,-1,-1,-1,-1];
   [-1,1,-1,-1,-1,-1];
   [-1,-1,1,-1,-1,-1];
   [-1,-1,-1,1,-1,-1];
   [-1,-1,-1,-1,1,-1];
   [-1,-1,-1,-1,-1,1]];
[Uinv,Dinv]=spec(J);//Solve matrix eigenvalue problem
c=Uinv*sqrt(Dinv);//Coefficients
IS1=list('sum',1.0,list('sum',c(1,1),a1),list('sum',c(1,2),a2),
list('sum',c(1,3),a3),list('sum',c(1,4),a4),list('sum',c(1,5),a5),
list('sum',c(1,6),a6));
IS2=list('sum',1.0,list('sum',c(2,1),a1),list('sum',c(2,2),a2),
list('sum',c(2,3),a3),list('sum',c(2,4),a4),list('sum',c(2,5),a5),
list('sum',c(2,6),a6));
IS3=list('sum',1.0,list('sum',c(3,1),a1),list('sum',c(3,2),a2),
list('sum',c(3,3),a3),list('sum',c(3,4),a4),list('sum',c(3,5),a5),
list('sum',c(3,6),a6));
IS4=list('sum',1.0,list('sum',c(4,1),a1),list('sum',c(4,2),a2),
list('sum',c(4,3),a3),list('sum',c(4,4),a4),list('sum',c(4,5),a5),
list('sum',c(4,6),a6));
IS5=list('sum',1.0,list('sum',c(5,1),a1),list('sum',c(5,2),a2),
list('sum',c(5,3),a3),list('sum',c(5,4),a4),list('sum',c(5,5),a5),
list('sum',c(5,6),a6));
IS6=list('sum',1.0,list('sum',c(6,1),a1),list('sum',c(6,2),a2),
list('sum',c(6,3),a3),list('sum',c(6,4),a4),list('sum',c(6,5),a5),
list('sum',c(6,6),a6));
S1=list('sum',sgn,I(n,IS1));//Get the touching hyperspheres
S2=list('sum',sgn,I(n,IS2));
S3=list('sum',sgn,I(n,IS3));
S4=list('sum',sgn,I(n,IS4));
S5=list('sum',sgn,I(n,IS5));
S6=list('sum',sgn,I(n,IS6));
```

After running the program to this point, we can evaluate any of the points IS_i,

```
-->cprint_element(IS1);
-0.816ia1+0.246a2-0.067a3+1.154a4-0.095a5-0.510a6
```

and check that lines IS_iS_j are indeed parabolic lines[7] so that $IS_iS_j.S_iS_j = 0$.

[7]See sections 3.10 and 7.6.1.

```
-->S1S2=wap(n,S1,S2);
-->cprint_element(wap(n,I(n,S1S2),S1S2));
+0.000
```

A.9.2 The radius function

The radius of a hypersphere is given by formula (7.17) which contains the hyperplane at infinity L. Following the argument at the end of section 7.6.1, we initially take the point at infinity as $IL = IS_1S_2.S_1$. The program sets this up as follows.

```
//Define a point at infinity IL.
IL=wap(n,wap(n,IS1,IS2),S1);
L=list('sum',sgn,I(n,IL));
```

The next part of the example program is a function to compute the radius of a hypersphere using equation (7.17). The formula for the radius is,

$$r = \sqrt{\frac{(SIS)}{(IS.L)^2}}\ .$$

The thing to notice about this formula is that the numerator SIS and the denominator $IS.L$ are both numbers. So, if we run the program and then calculate the number $IS_3.L$ (say) we get,

```
\index{length}
-->IS3_L=wap(n,I(n,S3),L);
-->cprint_element(IS3_L);
-2.000
-->length(IS3_L)
 ans = 2.
-->IS3_L(1)
 ans = point
-->IS3_L(2)
 ans  = -2.
```

which shows that the number $IS_3.L$ is stored as a point list with two slots. In fact it is `IS3_L=list('point',-2.0)`. The number itself is in the second slot of the list which is used to store an overall weight. Consequently, if an expression in Whitehead's algebra evaluates to a number, the numerical value resides in the second slot of the list that stores the expression. The following code shows how the function to calculate the radius is coded in the example program. The operations of multiplication, division, and square rooting, do not work on lists, so we have to extract the weight part of the lists.

```
function [r]=radius(n,S);
//Radius of a hypersphere S
  SIS=wap(n,S,I(n,S));
  IS_L=wap(n,I(n,S),L);
  r=sqrt(SIS(2)/IS_L(2)^2);
endfunction;
```

With our initial choice of the point at infinity as $IL = IS_1S_2.S_1$, the radius of S_1 is infinite as is that of S_2. In other words S_1 and S_2 are hyperplanes within the 4-d Euclidean space. The remaining hyperspheres $S_3, \ldots, S_6$ all have equal radii as the following example of calling the radius function shows.

```
-->radius(n,S3)
 ans = 0.5
```

The situation in which S_1 and S_2 are hyperplanes, whilst the remaining hyperspheres are of equal size is the 4-d analogue of the situation show in figure 7.5.

A.9.3 Moving the point at infinity

In order to get different configurations of touching hyperspheres, it is necessary to change the position of the point at infinity IL. Following section 7.2, one might think that $IL = a_5 - ia_6$ would be a suitable choice. In the following Scilab session, we re-define IL and check that it is on the quadric so that $IL.L = 0$. Then we evaluate the radius of a hypersphere, and find that our choice of IL makes the hyperspheres have complex radii.

```
-->IL=list('sum',1.0,a5,list('sum',-%i,a6));
-->cprint_element(IL);
+1.000a5-1.000ia6
-->L=list('sum',sgn,I(n,IL));
-->cprint_element(wap(n,IL,L));
+0.000
-->radius(n,S1)
 ans = 0.3516057 + 1.8961531i
```

In order to make the radii real-valued, it is necessary to make sure that IL is given by a linear combination of the new reference points IS_i with real coefficients. This was the case with our initial choice,

$$IL = IS_1S_2.S_1 = (IS_1.S_1)IS_2 - (IS_2.S_1)IS_1 = -(IS_1 + IS_2)$$

since[8] $IS_1.S_1 = -1$ and $IS_1.S_2 = 1$.

A convenient way of moving the point at infinity IL around on the quadric hypersurface is to act on it with a parabolic congruence of the form $\exp(\tau IS_iS_j.I)IL$. Equation (3.43) gives the formula for the action of a parabolic congruence on a point. It is repeated here,

$$f_\tau p = \exp(\tau ILM.I)p = p + \tau(ILM.Ip) + \frac{\tau^2}{2}(ILM.I(ILM.Ip))$$

for comparison with the following Scilab code.

```
function [fp]=pcon(n,tau,ILM,p);
//Action of a parabolic congruence with invariant line ILM on a point p.
  p1=list('sum',tau,wap(n,ILM,I(n,p)));
  p2=list('sum',0.5,wap(n,ILM,I(n,p1)));
  fp=list('sum',1.0,p,p1,p2);
endfunction;
```

So, we move the point at infinity IL with a parabolic congruence using (say) the invariant line IS_3S_4 and check that IL is still on the quadric so that $IL.L = 0$ and then evaluate the radii of the hyperspheres.

```
-->IL=pcon(n,1.0,wap(n,IS3,IS4),IL);
-->L=list('sum',sgn,I(n,IL));
-->cprint_element(wap(n,IL,L));
-0.000
-->radius(n,S1)
 ans = 0.25
-->radius(n,S2)
 ans = 0.25
-->radius(n,S3)
 ans = 0.5
-->radius(n,S4)
 ans = 0.1666667
-->radius(n,S5)
 ans = 0.1666667
-->radius(n,S6)
 ans = 0.1666667
```

We could move IL to arbitrary positions on the quadric by applying compound parabolic congruences as shown in section 3.6.2.

[8] See section 7.6.1.

A.9.4 Centres of the hyperspheres

Formula (7.15) gives the centre of a hypersphere. It is repeated here,

$$IN = \frac{2(IL.S)IS}{IS.S} - IL$$

for comparison with the code for a Scilab function to evaluate the point IN at the centre of a hypersphere S.

```
function [IN]=centre(n,S);
//Centre of a hypersphere by reflecting IL in S.
  IS_S=wap(n,I(n,S),S);
  IS_L=wap(n,I(n,S),L);
  IN=list('sum',1.,list('sum',2.0*IS_L(2)/IS_S(2),I(n,S)),list('sum',-1.,IL));
endfunction;
```

An application of this function is given in the next section.

A.9.5 Distances between the centres

Formula (3.49) gives the distance between two points on the quadric hypersurface. It is repeated here,

$$\tau = (-1)^{n-1}\sqrt{\frac{-2(pIq)}{(pL)(qL)}}$$

for comparison with the code of the corresponding Scilab function.

```
function [t]=quadric_dist(n,p,q);
//Distance between two points on the quadric
  pIq=wap(n,p,I(n,q));pL=wap(n,p,L);qL=wap(n,q,L);
  t=sqrt(-2.0*pIq(2)/(pL(2)*qL(2)));
endfunction;
```

We can now calculate the distances between the centres of pairs of hyperspheres. To make things clearer, we first calculate the signed curvatures[9] $\eta_i = IS_i.L$ of each hypersphere.

```
-->cprint_element(wap(n,IS1,L));
-4.000
-->cprint_element(wap(n,IS2,L));
```

[9]See section 7.6.1.

```
-4.000
-->cprint_element(wap(n,IS3,L));
+2.000
-->cprint_element(wap(n,IS4,L));
-6.000
-->cprint_element(wap(n,IS5,L));
-6.000
-->cprint_element(wap(n,IS6,L));
-6.000
```

Notice that the signed curvatures all have the same sign except for S_3 where the signed has changed. The sign change means that hypersphere S_3 encloses the other hyperspheres. So, if we calculate the distance between the centres of two hyperspheres with the same sign, say S_1 and S_4, we get,

```
-->quadric_dist(n,centre(n,S1),centre(n,S4))
 ans = 0.4166667
-->radius(n,S1)+radius(n,S4)
 ans = 0.4166667
```

where the sum of the radii confirms that the hyperspheres S_1 and S_4 are touching each other on their outsides. However, if we calculate the distance between the centres of S_3 and S_4,

```
-->quadric_dist(n,centre(n,S3),centre(n,S4))
 ans = 0.3333333
-->radius(n,S3)-radius(n,S4)
 ans = 0.3333333
```

then the difference of the radii confirms that S_4 is touching the inside of S_3.

Appendix B

Axioms for Whitehead's Algebra

B.1 Introduction

This appendix covers the same material as chapter 1 of the main text. Chapter 1 developed the theory heuristically, guided by geometrical intuition. This appendix is an attempt to develop the theory in a more formal way. In both versions, the product of any pair of basis elements of the algebra is evaluated using the rule of the middle factor. The difference is that chapter 1 derives the rule of the middle factor from the product of points and the product of hyperplanes, whereas this appendix postulates the rule of the middle factor.

B.2 Grassmann algebra

Let $a_1, \ldots, a_n$ be a set of anti-commuting generators,

$$a_i a_j = -a_j a_i \text{ and } (a_i)^2 = 0 \text{ for all } i,j. \tag{B.1}$$

for the basis elements of an algebra. The algebra is called a *Grassmann algebra.* The basis elements are of the form $a_i, a_i a_j, a_i a_j a_k, \ldots$. Anti-commutativity means that a generator cannot occur twice in a basis element. Furthermore, anti-commutativity ensures that all basis elements containing the same set of generators, but differing only by a permutation of the order of the generators, differ only by a sign change and are thus not independent. For example, if σ is a permutation of the indices $1, 2, \ldots, r$ as $\sigma(1), \sigma(2), \ldots, \sigma(r)$ then,

$$a_{\sigma(1)} \ldots a_{\sigma(r)} = \operatorname{sgn}(\sigma) a_1 a_2 \ldots a_r \ . \tag{B.2}$$

There are ${}^nC_r$ ways of picking r generators, all different, without regard to order, from the set of n generators. Hence there are ${}^nC_r$ independent basis elements of grade r of the form $a_{j_1}a_{j_2}\ldots a_{j_r}$ with $1 \leq j_1 < j_2 < \ldots < j_r \leq n$. The total number of independent basis elements for the Grassmann algebra is $2^n - 1$. The sequence of basis elements terminates after the single element $a_1 \ldots a_n$ of grade n.

The basis elements can be multiplied by numerical weights. Thus if $a_1 \ldots a_r$ is a basis element of grade r, then we can multiply it by a numerical weight ξ to form the grade r element $\xi a_1 \ldots a_r$. We assume that multiplication by a weight is such that whatever rule is used for multiplying two elements X and Y to form XY then,

$$\xi(XY) = (\xi X)Y = (X\xi)Y = X(\xi Y) = X(Y\xi) = (XY)\xi \ . \tag{B.3}$$

The grade n element $a_1 \ldots a_n$ is assumed to evaluate to a weight,

$$\omega = a_1 \ldots a_n \ . \tag{B.4}$$

For the moment, ω may be thought of as a numerical weight. However, it does not transform in quite the same way as a number, and for this reason ω is called a *pseudonumber*. Section B.10.5 explains how a pseudonumber transforms. In spite of this, ω is assumed to commute with everything as required by equation (B.3), and so qualifies as a weight.

The process of building basis elements of increasing grade by multiplying by new generators is cyclic because the grade n element is a weight. The sequence, $a_1, a_1a_2, \ldots$ eventually gets to $a_1 \ldots a_n = \omega$ and so attempts to make basis elements of grades higher than n simply result in the cycle of basis elements being repeated. Thus, $(a_1 \ldots a_n)a_1 = \omega a_1$. Therefore we say that the generators multiply *modulo* $a_1 \ldots a_n$.[1]

[1]Some clarification may be in order here. When we generate a basis element such as $a_1a_2a_3a_4$ we are really forming it as (say) $((a_1a_2)a_3)a_4$ by combining an extra generator at each stage. The generators freely associate so that we could equally well have formed this basis element by $a_1((a_2a_3)a_4)$ or some other order of combining the generators. This is why there is no confusion when we omit the brackets in $a_1a_2a_3a_4$. Suppose that $n = 4$ so that our element is the pseudonumber $\omega = a_1a_2a_3a_4$. When we try to combine another generator, say a_3 we just get the element ωa_3 which is proportional to a_3. In this way, the use of the generators to form the basis elements is cyclic, and we say that generator multiplication is modulo ω. The scheme of generator multiplication is a means of defining the basis elements of the algebra. It is not intended to define the product in the algebra. Although generator multiplication defines the product of some elements such as $a_1(a_2a_3) = a_1a_2a_3$ it does not define the product in all cases. For example, there is no meaning within the scheme of generator multiplication to a product such as $(a_1a_2a_3)(a_4a_3)$ for $n = 4$ because there is no way of combinatorially reaching an element $(a_1a_2a_3)(a_4a_3)$

The general element of the algebra of grade r is the linear combination,

$$X = \sum_{1 \le j_1 \ldots j_r \le n} \xi_{j_1 \ldots j_r} a_{j_1} \ldots a_{j_r} \tag{B.5}$$

where the $\xi_{j_1 \ldots j_r}$ are weights.

B.3 Dual generators for the algebra

In section B.2 we used the generators $a_1, \ldots, a_n$ to define the Grassmann algebra. We now assume that the same Grassmann algebra is also generated by a set of dual generators $A_1, \ldots, A_n$. A space and its dual space often occur in branches of mathematics. For example, in the theory of differential forms [11], vectors and 1-forms live in different spaces. Similarly, in the Hilbert space of quantum mechanics, bras and kets live in different spaces. However, in our case there is an important difference, the Grassmann algebra of the generators $a_1, \ldots, a_n$ and the Grassmann algebra made by the dual generators $A_1, \ldots, A_n$ are the same algebra. In other words, the set of basis elements $A_i, A_i A_j, \ldots$ are just an alternative set of basis elements for the Grassmann algebra. Everything in section B.2 can just be repeated by replacing the generators by their duals. In particular, we have the anti-commutation relations amongst the dual generators,

$$A_i A_j = -A_j A_i \text{ and } (A_i)^2 = 0 \text{ for all } i,j. \tag{B.6}$$

and the single basis element formed by the product of n dual elements is assumed to evaluate to the weight,

$$\Omega = A_1 \ldots A_n \ . \tag{B.7}$$

This equation is simply the dual statement to (B.4). The relation between the basis elements and the dual basis elements will soon appear in the theory.

B.4 The rule of the middle factor

We now define the rule which will multiply any pair X, Y of basis elements in the algebra to form the product XY. The first thing to do is to write the

---

because of the modulo ω property of the generators. The modulo ω property of generator multiplication breaks the free association of the generators. An expression such as $(a_1 a_2 a_3)(a_4 a_3)$ no longer freely associates for $n = 4$. Free association of the generators only works as long as the contents of the term does not exceed n generators.

basis element on the left X as a product of generators and the basis element on the right Y as a product of dual generators. Suppose that the grade of X, when expressed in terms of generators, is greater than the grade of Y expressed in terms of dual generators. So,

$$\begin{aligned} X &= a_1 \dots a_r a_{r+1} \dots a_{r+s} \\ Y &= A_1 \dots A_r \end{aligned}$$

and the rule says that the product is,

$$XY = (a_1 \dots a_r a_{r+1} \dots a_{r+s})(A_1 \dots A_r) = a_{r+1} \dots a_{r+s} \; . \tag{B.8}$$

In other words, the product is the middle factor $a_{r+1} \dots a_{r+s}$. The same rule applies if the grade of X, when expressed in terms of generators, is less than the grade of Y expressed in terms of dual generators. So, if,

$$\begin{aligned} X &= a_{r+1} \dots a_{r+s} \\ Y &= A_1 \dots A_r A_{r+1} \dots A_{r+s} \end{aligned}$$

the rule of the middle factor gives,

$$XY = (a_{r+1} \dots a_{r+s})(A_1 \dots A_r A_{r+1} \dots A_{r+s}) = A_1 \dots A_r \; . \tag{B.9}$$

In an informal way, the generators and dual generators, which match up label-wise at the ends of the expression, cancel out leaving the middle factor. If the generators do not match properly at first, then permutations are applied until the labels match. For example, take $n = 5$ and $X = a_5 a_4 a_2$ and $Y = A_2 A_4$.

$$XY = (a_5 a_4 a_2)(A_2 A_4) = (a_4 a_2 a_5)(A_2 A_4) = -(a_2 a_4 a_5)(A_2 A_4) = -a_5$$

Notice that if the basis elements are subjected to permutations which leave the labels on the generators in ascending order, then the basis elements are automatically ready for the application of the rule of the middle factor. The previous example illustrates this point.

If the generators and dual generators exactly match up and there are no generators left over in the middle factor, then the middle factor is regarded as being unity and so the result is,

$$(a_1 \dots a_r)(A_1 \dots A_r) = 1 \; . \tag{B.10}$$

If the generators and dual generators do not match up at all, then the result is zero. Thus,

$$(a_1 \dots a_r)(A_{r+1} \dots A_{r+s}) = 0 \; . \tag{B.11}$$

The rule of the middle factor defines the product XY of the algebra when the factors are basis elements. The distributive rule is used to extend the product to linear combinations of the basis elements,

$$(\xi X + \eta Y)Z = \xi(XY) + \eta(XZ) \tag{B.12}$$

where X, Y, Z are basis elements and ξ, η are weights.

B.5 Basis elements and their duals

Set up a basis element $X = a_{r+1} \dots a_n$ and a dual basis element $Y = A_1 \dots A_r$ and evaluate ωY and $X\Omega$ by the rule of the middle factor:

$$\begin{aligned} \omega Y &= a_1 \dots a_n . A_1 \dots A_r = a_{r+1} \dots a_n = X \\ X\Omega &= a_{r+1} \dots a_n . A_1 \dots A_n = A_1 \dots A_r = Y \ . \end{aligned}$$

Since ω and Ω are weights, these equations show that X and Y are proportional to one another. Furthermore, by setting $r = n$ in (B.10) we obtain $\omega\Omega = 1$. Therefore,

$$X = a_{r+1} \dots a_n = \omega Y = \Omega^{-1} Y = \frac{A_1 \dots A_r}{A_1 \dots A_n} \tag{B.13}$$

$$Y = A_1 \dots A_r = X\Omega = X\omega^{-1} = \frac{a_{r+1} \dots a_n}{a_1 \dots a_n} \tag{B.14}$$

show how the basis elements and the dual basis elements are related.

B.6 Consistency of the product

The principal difficulty in the exposition of Whitehead's algebra is that the product XY is defined in terms of its actions on basis elements. So, before we can define the product we have to introduce basis elements like $a_1 \dots a_r$ and $A_1 \dots A_r$. Now one needs to introduce anti-commutativity (B.1,B.6) in order to produce the correct number of independent basis elements. However, anti-commutativity and the way the generators build up basis elements, for example,

$$(a_1 a_2) a_3 = a_1 a_2 a_3$$

means that we already know how to evaluate XY in some cases before we get to the definition of the product. Now there is only one product in the algebra. This single product is evaluated by the rule of the middle factor. So, we have to check that the rule of the middle factor is fully consistent with the way the generators work. So, we have to check, for example, that if we set $X = a_1 a_2$ and $Y = a_3$, the rule of the middle factor also produces $XY = a_1 a_2 a_3$.

B.6.1 Consistency of building basis elements

Using the generators, we can multiply in the following fashion,

$$a_r(a_{r+1} \dots a_n) = a_r \dots a_n \ .$$

We could also do this with the rule of the middle factor. Since there is only one product in the algebra, both ways need to give the same result. In order to do the calculation by the rule of the middle factor, we write the second factor in terms of dual generators using (B.13),

$$\begin{aligned} a_r.a_{r+1} \dots a_n &= a_r.\frac{A_1 \dots A_r}{A_1 \dots A_n} = \frac{a_r.A_1 \dots A_r}{A_1 \dots A_n} \\ &= \frac{A_1 \dots A_{r-1}}{A_1 \dots A_n} = a_r \dots a_n \end{aligned}$$

and the last line follows from another application of (B.13).

B.6.2 Consistency of free association of generators

The fact that we can write the basis elements as products of the generators without any brackets means that the products of the generators freely associate,

$$a_{r+1} \dots a_n = (a_{r+1} \dots a_{r+s})(a_{r+s+1} \dots a_n) \ .$$

For consistency, the rule of the middle factor must also re-produce this result.

$$\begin{aligned} (a_{r+1} \dots a_{r+s})(a_{r+s+1} \dots a_n) &= a_{r+1} \dots a_{r+s}.\frac{A_1 \dots A_{r+s}}{A_1 \dots A_n} \\ = \frac{a_{r+1} \dots a_{r+s}.A_1 \dots A_{r+s}}{A_1 \dots A_n} &= \frac{A_1 \dots A_r}{A_1 \dots A_n} \\ &= a_{r+1} \dots a_n \end{aligned} \tag{B.15}$$

Dually, the fact that the dual basis elements can be written without brackets means that they freely associate,

$$A_1 \dots A_{r+s} = (A_1 \dots A_r)(A_{r+1} \dots A_{r+s})$$

and so the rule of the middle factor should also reproduce this behaviour.

$$\begin{aligned} (A_1 \dots A_r)(A_{r+1} \dots A_{r+s}) &= \frac{a_{r+1} \dots a_n}{a_1 \dots a_n}.A_{r+1} \dots A_{r+s} \\ = \frac{a_{r+s+1} \dots a_n}{a_1 \dots a_n} &= A_1 \dots A_{r+s} \end{aligned} \tag{B.16}$$

B.6.3 Consistency of modulo multiplication

Let us evaluate $\omega a_{r+1} \dots a_n$ by the rule of the middle factor. The generators multiply modulo ω so that the result is proportional to $a_{r+1} \dots a_n$. The rule of the middle factor,

$$a_1 \dots a_n . a_{r+1} \dots a_n = \frac{a_1 \dots a_n . A_1 \dots A_r}{A_1 \dots A_n} = \frac{a_{r+1} \dots a_n}{A_1 \dots A_n}$$

reproduces this behaviour. Dually, $a_{r+1} \dots a_n \Omega$ is proportional to $a_{r+1} \dots a_n$. The rule of the middle factor,

$$a_{r+1} \dots a_n . A_1 \dots A_n = A_1 \dots A_r = \frac{a_{r+1} \dots a_n}{a_1 \dots a_n} \ .$$

also reproduces this result.

B.6.4 Consistency of generator permutations

Using the anti-commutation relations of the generators (B.1), we can obtain the following formula

$$\begin{aligned} YX &= (a_{r+1} \dots a_{r+s})(a_1 \dots a_r) = a_{r+1} \dots a_{r+s} a_1 \dots a_r \\ &= (-1)^s a_1 a_{r+1} \dots a_{r+s} a_2 \dots a_r = (-1)^{rs} a_1 \dots a_r a_{r+1} \dots a_{r+s} \\ &= (-1)^{rs} XY \qquad \text{(B.17)} \end{aligned}$$

for swopping the order of factors in a product. The dual formula,

$$\begin{aligned} YX &= (A_{r+1} \dots A_{r+s})(A_1 \dots A_r) = (-1)^{rs} A_1 \dots A_r A_{r+1} \dots A_{r+s} \\ &= (-1)^{rs} XY \qquad \text{(B.18)} \end{aligned}$$

can also be obtained in the same way using (B.6) for the case in which the factors X, Y are basis elements formed by products of the dual generators. For consistency, these two formulae must also be obtained by swopping the order of factors in the rule of the middle factor. For brevity, we only show consistency with (B.17).

$$\begin{aligned} (a_{r+1} \dots a_{r+s})(a_1 \dots a_r) &= a_{r+1} \dots a_{r+s} . \frac{A_{r+1} \dots A_n}{A_{r+1} \dots A_n A_1 \dots A_r} \\ &= a_{r+1} \dots a_{r+s} . \frac{A_{r+1} \dots A_n}{(-1)^{r(n-r)} A_1 \dots A_n} \end{aligned}$$

The factor $(-1)^{r(n-r)}$ in the denominator of the above equation has appeared by swopping the factors using anti-commutativity (B.6) which is equivalent to an application of (B.18). Continuing, we use anti-commutativity again in the numerator to set things up for an application of the rule of the middle factor.

$$
\begin{aligned}
&(a_{r+1} \dots a_{r+s})(a_1 \dots a_r) \\
= a_{r+1} \dots a_{r+s} \cdot &\frac{(-1)^{s(n-r-s)} A_{r+s+1} \dots A_n A_{r+1} \dots A_{r+s}}{(-1)^{r(n-r)} A_1 \dots A_n} \\
&= \frac{(-1)^{s(n-r-s)} A_{r+s+1} \dots A_n}{(-1)^{r(n-r)} A_1 \dots A_n}
\end{aligned}
$$

Finally, we use anti-commutativity again to set things up for an application of (B.13).

$$
\begin{aligned}
&(a_{r+1} \dots a_{r+s})(a_1 \dots a_r) \\
&= \frac{(-1)^{s(n-r-s)} A_{r+s+1} \dots A_n}{(-1)^{r(n-r)}(-1)^{(r+s)(n-r-s)} A_{r+s+1} \dots A_n A_1 \dots A_{r+s}} \\
&= (-1)^{r(n-r-s)-r(n-r)} a_1 \dots a_{r+s} = (-1)^{rs} a_1 \dots a_{r+s}
\end{aligned}
$$

The result agrees with (B.17).

B.7 Generators and their duals

By setting $r = n - 1$ in (B.13),

$$a_n = \frac{A_1 \dots A_{n-1}}{A_1 \dots A_n} \ .$$

Since the generators are arbitrary elements of grade 1, we can re-label to get the formula for a general one.

$$
\begin{aligned}
a_j &= \frac{A_1 \dots \breve{A}_j \dots A_n}{A_1 \dots \breve{A}_j \dots A_n A_j} = \frac{A_1 \dots \breve{A}_j \dots A_n}{(-1)^{n-j} A_1 \dots A_n} \\
&= \frac{(-1)^{n-j} A_1 \dots \breve{A}_j \dots A_n}{A_1 \dots A_n}
\end{aligned} \tag{B.19}
$$

In the above formula, the sign changes have been brought about using (B.6).

Similarly, by setting $r = 1$ in (B.14),

$$A_1 = \frac{a_2 \dots a_n}{a_1 \dots a_n}$$

and by re-labelling,

$$A_j = \frac{a_1 \dots \check{a}_j \dots a_n}{a_j a_1 \dots \check{a}_j \dots a_n} = \frac{a_1 \dots \check{a}_j \dots a_n}{(-1)^{j-1} a_1 \dots a_n} = \frac{(-1)^{j-1} a_1 \dots \check{a}_j \dots a_n}{a_1 \dots a_n} . \tag{B.20}$$

B.8 Physical interpretation

The generators $a_1, \dots, a_n$ are called *reference points*. A basis element $a_1 \dots a_r$ is assumed to represent the $(r-1)$-dimensional linear subspace containing the r points. On account of this assumption and (B.20) it follows that the dual generators are *reference hyperplanes*. Since a dual generator is a hyperplane, and (B.19) shows that a point is a product of $n-1$ hyperplanes, it follows that a dual basis element must represent the intersection of its hyperplane factors.

Consider the linear subspace,

$$X = a_{r+1} \dots a_n = \omega A_1 \dots A_r .$$

A linear combination of the reference points,

$$p = \sum_{i=r+1}^{n} \xi_i a_i$$

represents a point belonging to the linear subspace X. From the rule of the middle factor,

$$pX = \omega \sum_{i=r+1}^{n} \xi_i a_i . A_1 \dots A_r = 0$$

because there are no middle factors. Hence, the equation for the locus of points p that lie in a linear subspace X is $pX = 0$. If we multiply the subspace X by a weight ξ to give the element $Y = \xi X$ then $pY = p.\xi X = \xi(pX) = 0$ and so p is also in the subspace Y. Hence, X and ξX represent the same linear subspace. In other words, the algebraic elements ξX where ξ is an arbitrary weight, correspond to the same linear subspace.

B.9 Meaning of the product

Equation (B.15) can be written,

$$X = a_{r+1} \dots a_{r+s}$$

$$
\begin{aligned}
Y &= a_{r+s+1} \dots a_n = \omega A_1 \dots A_{r+s} \\
XY &= (a_{r+1} \dots a_{r+s})(a_{r+s+1} \dots a_n) = (a_{r+1} \dots a_{r+s})(\omega A_1 \dots A_{r+s}) \\
&= a_{r+1} \dots a_n \ . \qquad \text{(B.21)}
\end{aligned}
$$

Here, the point grade of X is s and the point grade of Y is $n-r-s$. The sum of the point grades of X and Y is $n-r \le n$ and the in the rule of the middle factor, the hyperplane factor Y is the larger product since $r+s \ge s$. Equation (B.21) shows that the product evaluates to the basis element formed by just multiplying the generators from each factor. In other words, in this case the product XY represents the joining together of the factor subspaces.

Equation (B.16) can be written,

$$
\begin{aligned}
X &= A_1 \dots A_r = \Omega a_{r+1} \dots a_n \\
Y &= A_{r+1} \dots A_{r+s} \\
XY &= (A_1 \dots A_r)(A_{r+1} \dots A_{r+s}) = (\Omega a_{r+1} \dots a_n)(A_{r+1} \dots A_{r+s}) \\
&= A_1 \dots A_{r+s} \qquad \text{(B.22)}
\end{aligned}
$$

Here, the hyperplane grade of X is r and the hyperplane grade of Y is s. Furthermore, Y is the basis element containing the hyperplane A_{r+s} so we know that $r+s \le n$. In the rule of the middle factor, the point factor X is the larger product since $n-r \ge s$. Equation (B.22) shows that the product evaluates to the basis element formed by just multiplying the dual generators from each factor. In other words, in this case the product XY represents the intersection of the factor subspaces.

When we evaluate the product XY using the rule of the middle factor by writing X as a product of points and Y as a product of hyperplanes, then either X is the largest factor or Y is the largest factor, or both factors are the same size. All these cases are handled by (B.21,B.22).

B.10 Transformations

B.10.1 Transformation of hyperplanes

We assume that the world-views of two observers Alice and Bob are related by a linear transformation f.

$$
\begin{aligned}
&\text{Bob} \xrightarrow{f} \text{Alice} \\
&fA_j \qquad A_j \\
&= \sum_{j=1}^{n} (a_i.fA_j)A_i
\end{aligned}
$$

The above diagram means that if Alice sees a hyperplane A_j then Bob sees this hyperplane as fA_j where the $(a_i.fA_j)$ are the numerical matrix elements of the transformation[2]

B.10.2 Transformation of numbers

Let ξ be a number. Since f is linear,

$$\begin{array}{rcl} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ \xi f(A_i) & & \xi A_i \\ \text{alternatively,} & & \\ f(\xi)f(A_i) & & \xi A_i \end{array}$$

so for consistency a number must transform as a scalar,

$$f(\xi) = \xi \ . \tag{B.23}$$

B.10.3 Transformation of points

Consider the matrix elements $a_i.fA_j$. They are numbers so they transform transform as scalars. Hence,

$$a_i.fA_j = f^{-1}(a_i.fA_j) = f^{-1}a_i.f^{-1}fA_j = f^{-1}a_i.A_j$$

and since f is any linear transformation, we replace f by f^{-1} to obtain,

$$a_i.f^{-1}A_j = fa_i.A_j \ . \tag{B.24}$$

Hence,

$$fa_i = \sum_{j=1}^{n}(fa_i.A_j)a_j = \sum_{j=1}^{n}(a_i.f^{-1}A_j)a_j \tag{B.25}$$

the reference points transform using the inverse matrix to the transformation of the reference hyperplanes.

[2]An arbitrary hyperplane is a linear combination of reference hyperplanes $X = \sum_j \xi_j A_j$. The numerical weights in this expansion are found to be $\xi_i = a_i X$ by multiplying by a_i. For,

$$a_i X = \sum_j \xi_j a_i A_j = \sum_j \xi_j \delta_{ij} = \xi_i$$

since $a_i A_j = \delta_{ij}$ by the rule of the middle factor. Hence, the weights in the exansion of fA_j are $a_i.fA_j$.

B.10.4 Transformation of basis elements

The basis elements must transform as follows.

$$\begin{array}{rl} \text{Bob} \xrightarrow{f} & \text{Alice} \\ fA_i & A_i \\ f(A_1)\ldots f(A_r) & A_1\ldots A_r \\ \text{Alternatively,} & \\ f(A_1\ldots A_r) & A_1\ldots A_r \end{array}$$

therefore, for consistency we must have,

$$f(A_1\ldots A_r) = f(A_1)\ldots f(A_r) \tag{B.26}$$

which amounts to a definition of how to transform the higher grade elements. Dually, we also have,

$$f(a_1\ldots a_r) = f(a_1)\ldots f(a_r)\ . \tag{B.27}$$

B.10.5 Transformation of pseudonumbers

If we set $r = n$ in (B.26) and expand using the matrix elements, we find,

$$\begin{aligned} f(A_1\ldots A_n) &= fA_1\ldots fA_n \\ &= \sum_{i_1=1}^{n}\ldots\sum_{i_n=1}^{n}(a_{i_1}.fA_1)\ldots(a_{i_n}.fA_n)A_{i_1}\ldots A_{i_n}\ . \end{aligned}$$

Most of the terms $A_{i_1}\ldots A_{i_n}$ will vanish by anti-commutativity (B.6). The only non-zero terms will be the ones for which all of the n labels $i_1,\ldots,i_n$ are all different. Thus, the n-fold summation collapses to a sum over permutations σ of the sequence $1,2,\ldots,n$. Hence,

$$\begin{aligned} f(A_1\ldots A_n) &= \sum_{\sigma}(a_{\sigma(1)}.fA_1)\ldots(a_{\sigma(n)}.fA_n)A_{\sigma(1)}\ldots A_{\sigma(n)} \\ &= \sum_{\sigma}\mathrm{sgn}(\sigma)(a_{\sigma(1)}.fA_1)\ldots(a_{\sigma(n)}.fA_n)A_1\ldots A_n \end{aligned}$$

and the last line follows from an application of the dual version of (B.2). The summation over the matrix elements $a_i.fA_j$ is defined as the determinant $\det f$ in the theory of matrices. Hence, a pseudonumber transforms as,

$$f\Omega = f(A_1\ldots A_n) = (\det f)A_1\ldots A_n = (\det f)\Omega\ . \tag{B.28}$$

We can also obtain work out how $\omega = a_1 \dots a_n$ transforms. Since $\omega\Omega = 1$,

$$1 = \omega\Omega = f(\omega\Omega) = f\omega . f\Omega = (\det f) f\omega . \Omega$$

and upon dividing throughout by Ω we obtain,

$$f\omega = \frac{\omega}{\det f} = (\det f^{-1})\omega \ . \tag{B.29}$$

The fact that $\det f^{-1} = 1/\det f$ can be proved as follows. Applying the composite transformation gf gives $gf\Omega = \det(gf)\Omega$. Applying the transformations in succession gives,

$$gf\Omega = g(\det f)\Omega = (\det g)(\det f)\Omega$$

so that $\det gf = \det g \det f$ and by setting $g = f^{-1}$ the result follows.

B.11 Physical consistency of the product

Consider the product XY which we evaluate by the rule of the middle factor.

$$\begin{array}{rcl} \text{Bob} & \xrightarrow{f} & \text{Alice} \\ fX & & X \\ fY & & Y \\ fX.fY & & XY \\ \text{Alternatively,} & & \\ f(XY) & & XY \end{array}$$

For consistency we must have,

$$f(XY) = fX.fY \ . \tag{B.30}$$

We cannot just invent a product because (B.30) is a non-trivial constraint on any physically-meaningful algebra. Fortunately, the rule of the middle factor satisfies this constraint. For, an application of the rule of the middle factor either evaluates like a product of points (B.21) or as a product of hyperplanes (B.22). A product of points transforms as (B.27) and a product of hyperplanes transforms as (B.26). Both cases are fully compliant with condition (B.30). For, suppose that XY evaluates as a product of hyperplanes.

$$\begin{aligned} X &= A_1 \dots A_r \\ Y &= A_{r+1} \dots A_{r+s} \end{aligned}$$

$$
\begin{aligned}
XY &= (A_1 \dots A_r)(A_{r+1} \dots A_{r+s}) = A_1 \dots A_{r+s} \\
f(XY) &= f(A_1 \dots A_{r+s}) = fA_1 \dots fA_{r+s} \\
&= (fA_1 \dots fA_r)(fA_{r+1} \dots fA_{r+s}) = f(A_1 \dots A_r)f(A_{r+1} \dots A_{r+s}) \\
&= fX.fY
\end{aligned}
$$

When XY evaluates as a product of points the proof is the dual one.

This is an appropriate place to end this axiomatic development of Whitehead's algebra. Having studied this appendix, the reader can begin studying chapter 1 at section 1.14.

Bibliography

[1] Murray R. Spiegel. *Theory and Problems of Vector Analysis.* Schaum's Outline Series, McGraw-Hill Book Company, 1959.

[2] David Hestenes, Garret Sobczyk. *Clifford Algebra to Geometric Calculus, A Unified Language for Mathematics and Physics.* D.Reidel, Dordrecht, 1984.

[3] Chris Doran and Anthony Lasenby. *Geometric Algebra for Physicists.* Cambridge University Press, 2003.

[4] Ron Goldman. *On the Algebraic and Geometric Foundations of Computer Graphics.* ACM Transactions on Graphics, Vol. 21, No. 1, January 2002, Pages 52-86.

[5] Marilena Barnabei, Andrea Brini and Gian-Carlo Rota. *On the Exterior Calculus of Invariant Theory.* Journal of Algebra, Volume 96, 1985, Pages 120-160.

[6] Olivier Faugeras and Quang-Tuan Luong. *The Geometry of Multiple Images.* The MIT Press, 2001.

[7] Hermann Grassmann. *The Ausdehnungslehre of 1844, and Other Works.* Translated by L. C. Kannenberg, Open Court 1995.

[8] Félix Klein. *Le Programme D'Erlangen.* Gauthier-Villars, 1974.

[9] H. S. M. Coxeter. *Non-Euclidean Geometry.* Mathematical Association of America, sixth edition 1998.

[10] A. N. Whitehead. *A Treatise on Universal Algebra.* Cambridge Press 1898.

[11] Bernard Schutz. *Geometrical Methods of Mathematical Physics.* Cambridge University Press, 1980.

[12] *Yorick* http://wuarchive.wustl.edu/languages/yorick/doc/index.html

[13] M. Göckeler and T. Schücker. *Differential geometry, gauge theories, and gravity.* Cambridge University Press, 1987.

[14] H. S. M. Coxeter. *Introduction to Geometry.* John Wiley & Sons, Inc., second edition, 1989.

[15] Sir R. S. Ball. *A Treatise on the Theory of Screws.* Cambridge University Press, first published 1900, paperback edition 1998.

[16] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry.* Oxford University Press 1952.

[17] S. W. Hawking & G. F. R. Eliis. *The large scale structure of space-time.* Cambridge University Press, 1973.

[18] E. Schrödinger. *Expanding Universes.* Cambridge University Press 1956.

[19] M. Parikh, I. Savonije and E. Verlinde. *Elliptic de Sitter Space dS/Z_2.* arXiv:hep-th/0209120.

[20] Han-Ying Guo, Chao-Guang Huang and Bin Zhou. *On the Horizon in de Sitter Invariant Special Relativity.* arXiv:hep-th/0404010.

[21] http://homepage.ntlworld.com/stebla/index.html

[22] H. Li, D. Hestenes and A. Rockwood. *Generalized Homogeneous Coordinates for Computational Geometry* in G. Sommer, editor, Geometric Computing with Clifford Algebras, Springer 2000.

[23] R. Mohr, L. Morin and E. Grosso. *Relative Positioning with Uncalibrated Cameras* in J. Mundy and A. Zisserman editors, Geometric Invariance in Computer Vision, The MIT Press, Cambridge, Massachusetts, 1992.

[24] Eric W. Weisstein. *Soddy's formula or Descartes circle theorem.* http://mathworld.wolfram.com/DescartesCircleTheorem.html

[25] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* John Hopkins University Press, 1983.

[26] *Scilab* http://scilabsoft.inria.fr/

[27] Henry George Forder. *Calculus of Extension.* Chelsea Publishing Company, New York, 1960.

[28] Garrett Birkhoff and Saunders MacLane. *A Survey of Modern Algebra*. A K Peters, Wellesley, Massachusetts, fourth edition, 1997.

[29] L. D. Landau and E. M. Lifshitz translated by Morton Hamermesh. *The Classical Theory of Fields*. Pergamon Press, third English edition, 1971.

[30] Sean M. Carroll. *Lecture Notes on General Relativity*. arXiv:gr-qc/9712019.

[31] D. K. Sen. *Fields and/or Particles*. Academic Press, 1968.

[32] E. Schrödinger. *Space-Time Structure*. Cambridge University Press, 1985 paperback edition.

Index