Elmer G. Rees

Notes on Geometry

Elmer G. Rees

Notes on Geometry

With 99 Figures

Springer-Verlag
Berlin Heidelberg New York 1983

Professor Elmer G. Rees

University of Edinburgh, Department of Mathematics,
James Clerk Maxwell Building
The King's Buildings, Mayfield Road
Edinburgh, EH9 3JZ, Scotland

ISBN 3-540-12053-X Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-12053-X Springer-Verlag New York Heidelberg Berlin

Library of Congress Cataloging in Publication Data. Main entry under Title: Rees, Elmer G.
Notes on Geometry. (Universitext) Bibliography: p. Includes index.
1. Geometry. I. Title. QA445.R43 1983 516 82-19299

Printing and Binding: fotokop Wilhelm Weihert KG, Darmstadt
2141/3140-543210

Acknowledgements

I am very grateful to the very many people who have given me encouragement, help and criticism both whilst giving the course and whilst writing these notes. I would like to single out

Christopher Zeeman who, in the early days at Warwick, inspired my interest in geometry of this kind,

Michael Atiyah and Graeme Segal who encouraged me to give this course at Oxford,

Graeme Segal who made numerous suggestions and after my initial attempts gave a much better course than mine. He was very willing for me to incorporate his improvements into these notes. The approach to hyperbolic trigonometry is entirely his,

Wilson Sutherland who saved me from error many times, and finally

Pam Armstrong and David Stewart-Robinson who prepared the text and diagrams.

Preface

In recent years, geometry has played a lesser role in undergraduate courses than it has ever done. Nevertheless, it still plays a leading role in mathematics at a higher level. Its central role in the history of mathematics has never been disputed. It is important, therefore, to introduce some geometry into university syllabuses. There are several ways of doing this, it can be incorporated into existing courses that are primarily devoted to other topics, it can be taught at a first year level or it can be taught in higher level courses devoted to differential geometry or to more classical topics.

These notes are intended to fill a rather obvious gap in the literature. It treats the classical topics of Euclidean, projective and hyperbolic geometry but uses the material commonly taught to undergraduates: linear algebra, group theory, metric spaces and complex analysis. The notes are based on a course whose aim was two fold, firstly, to introduce the students to some geometry and secondly to deepen their understanding of topics that they have already met. What is required from the earlier material is a familiarity with the main ideas, specific topics that are used are usually redone.

The style of the course was informal and I hope some of the associated good aspects have survived into this version. In line with this, I have taken a concrete viewpoint rather than an axiomatic one. The view that I take is that mathematical objects exist and should be studied, they are not arbitrarily defined as the axiomatic approach might suggest. This is the view of the vast majority of mathematicians in their own work and it is a pity that this does not come across in more undergraduate courses.

There are a large number of exercises throughout the notes, many of these are very staightforward and are meant to test the reader's understanding. Problems, some of them of interest in their own right are given at the end of the three parts. Some are straightforward and some are more like small projects. The more difficult ones are marked with an asterisk.

Contents

Introduction

In Euclidean geometry, two triangles are **congruent** if one of them can be moved rigidly onto the other. Definitions such as that of congruence, which tell us when two objects should be regarded as being the same, are basic in geometry and are often used to characterize a particular geometry. Two sets A, B are defined to be equivalent if there is an 'allowed transformation' f such that $fA = B$. For Euclidean geometry the allowed transformations are the **rigid motions**. In his Erlanger programme of 1872, Felix Klein formulated the principle that a geometry is defined by its allowed transformations. The force of this principle is to make a close connection between geometry and group theory.

If S is a set (an example to bear in mind is the Euclidean plane $\mathbf{R}^2$), consider the group Bij(S) consisting of all bijections $f: S \to S$. (If S is a finite set with n elements this is the (familiar) symmetric group S_n.) To impose a geometry on S is to consider a subgroup G of Bij(S); two subsets A, B being equivalent for the geometry if there is an $f \in G$ such that $fA = B$. For Euclidean geometry, S is $\mathbf{R}^2$ and G is the group of all rigid motions. Klein's Erlanger programme not only says that the geometry on S and the subgroup G determine each other but that they are, as a matter of definition, one and the same thing. To obtain a worthwhile geometry, the subgroup G has to be chosen with some care after considerable experience. Usually the set S has some structure and the group G preserves this structure, examples are

i) S may be a topological space and the elements of G are homeomorphisms of S.

ii) S may have certain subsets (for example, lines) that are mapped to each other by the elements of G.

There are many other types of examples; in these notes we study the three 'classical' geometries, Euclidean, projective and hyperbolic, but the approach is guided by Klein's Erlanger programme.

Part I

Euclidean Geometry

We start by studying the linear groups. These are probably already familiar to the reader. They play an important role in the study of geometry.

The Linear Groups

The ring $M(n,\mathbf{R})$ of all $n \times n$ matrices over the field $\mathbf{R}$ of real numbers has the **general linear group** $GL(n,\mathbf{R})$ as its group of units, that is, $GL(n,\mathbf{R})$ consists of all the invertible real $n \times n$ matrices. We will often identify $M(n,\mathbf{R})$ with the space of all linear transformations $T: \mathbf{R}^n \to \mathbf{R}^n$. Note that $M(n,\mathbf{R})$ is a real vector space of dimension n^2, and so can be regarded as the metric space $\mathbf{R}^{n^2}$. The determinant defines a continuous map

$$\det: M(n,\mathbf{R}) \to \mathbf{R}$$

(continuous because it is given by a polynomial in the coefficients of a matrix), and $GL(n,\mathbf{R})$ is $\det^{-1}(\mathbf{R} \setminus \{0\})$, so as $\mathbf{R} \setminus \{0\}$ is an open subset of $\mathbf{R}$ we see that

$$GL(n,\mathbf{R}) \text{ is an open subset of } M(n,\mathbf{R}) = \mathbf{R}^{n^2}.$$

The determinant is multiplicative and so defines a homomorphism of groups

$$\det: GL(n,\mathbf{R}) \to \mathbf{R} \setminus \{0\}.$$

Its kernel is the **special linear group** $SL(n,\mathbf{R})$ consisting of matrices with determinant 1. The subset $SL(n,\mathbf{R})$ is closed in $GL(n,\mathbf{R})$ and has dimension $n^2 - 1$ (but is hard to visualize – try to do so for $n = 2$).

Euclidean space $\mathbf{R}^n$ will always be considered with an **inner product** $x.y$ defined on it, this satisfies

i) $(x+y).z = x.z + y.z$ for all $x,y,z \in \mathbf{R}^n$,
ii) $(\lambda x).y = \lambda(x.y)$ for all $x,y \in \mathbf{R}^n$, $\lambda \in \mathbf{R}$,
iii) $x.y = y.x$ for all $x, y \in \mathbf{R}^n$, and
iv) $x.x = 0 \Leftrightarrow x = 0$.

The inner product defines a **norm** $\| \,\|$ on $\mathbf{R}^n$ by $\|x\|^2 = x.x$ and a **metric** d by $d(x,y) = \|x-y\|$. Note that $d(x+a,y+a) = d(x,y)$ so that distance is translation invariant. From the viewpoint of Euclidean geometry the most important transformations are those that preserve distances. We will now study such linear transformations.

A linear transformation $T: \mathbf{R}^n \to \mathbf{R}^n$ is called **orthogonal** if $Tx.Ty = x.y$ for all $x, y \in \mathbf{R}^n$. A basis $\{e_1, e_2, \ldots, e_n\}$ for $\mathbf{R}^n$ is **orthogonal** if

$$e_i.e_j = 0 \text{ if } i \neq j$$
$$= 1 \text{ if } i = j.$$

If we write $x = \sum_1^n x_i e_i$ and $y = \sum_1^n y_i e_i$, then

$$x.y = \underline{x}^t \underline{y},$$

where on the right hand side $\underline{x}$, $\underline{y}$ denote the column vectors with entries x_i, y_i respectively. If T is an orthogonal transformation and A is the matrix of T with respect to an orthonormal basis, an easy calculation shows that

$$\underline{x}^t A^t A \underline{y} = \underline{x}^t \underline{y}$$

and by choosing various suitable x, y one sees that $A^tA = I$. Check this by using the following result.

Exercise Note that $a_{ij} = e_i{}^t A e_j$ and hence show that if $x^tAy = x^tBy$ for all $x,y \in \mathbf{R}^n$ then $A = B$.

The matrix of an orthogonal transformation with respect to an orthogonal basis is therefore orthogonal in the usual sense for matrices. Note that a little care is needed in handling orthogonal matrices because a matrix is orthogonal if and only if its columns (or rows) form an **orthonormal** set of vectors.

If X is a metric space, a map $f: X \to X$ is an **isometry** if it is onto and distance preserving, that is, $d(fx,fy) = d(x,y)$ for all $x,y \in X$.

Exercise i) Show that a distance preserving map is one to one.

ii) Show that a distance preserving map $f: X \to X$ is not necessarily onto by considering the map $f: \mathbf{R}_+ \to \mathbf{R}_+$ defined by $f(x) = x+1$.

Exercise If X is a metric space, verify that the set of all isometries $f: X \to X$ forms a group under composition.

The properties of the isometries of a metric space X are intimately connected with the properties of X itself. The importance of orthogonal transformations in Euclidean geometry arises because they are isometries of $\mathbf{R}^n$, moreover, apart from translations, they are in a sense all the isometries.

Lemma i) If $T: \mathbf{R}^n \to \mathbf{R}^n$ is a linear isometry then T is orthogonal.

ii) If $T: \mathbf{R}^n \to \mathbf{R}^n$ is linear and norm-preserving then T is orthogonal.

Proof Notice that any linear isometry is norm preserving because any such isometry satisfies

$$T(x-y).T(x-y) = (Tx-Ty).(Tx-Ty) = d(Tx,Ty)^2 = d(x,y)^2 = (x-y).(x-y)$$

so by putting $y = 0$ one gets

$$\|Tx\|^2 = \|x\|^2.$$

Hence it suffices to prove ii). The map T preserves the norm of $x-y$ so

$$T(x-y).T(x-y) = (x-y).(x-y).$$

Expanding these expressions using linearity gives

$$Tx.Tx - 2Tx.Ty + Ty.Ty = x.x - 2x.y + y.y.$$

But T also preserves the norms of x and y, so

$$Tx.Ty = x.y \text{ for all } x, y \in \mathbf{R}^n,$$

and so T is orthogonal.

Of course there are isometries that are not linear, for example, the translations $T_a: \mathbf{R}^n \to \mathbf{R}^n$ defined by $T_a(x) = a + x$ are not linear unless $a = 0$. Later we will show that any isometry $f: \mathbf{R}^n \to \mathbf{R}^n$ such that $f0 = 0$ is linear.

The set of all orthogonal $n \times n$ matrices form the **orthogonal group**

$$O(n) = \{A \in GL(n,\mathbf{R}) | A^tA = I\}.$$

If A is orthogonal then $\det A = \pm 1$ because $\det A^t = \det A$ and so

$$1 = \det I = \det(A^tA) = \det A^t.\det A = (\det A)^2.$$

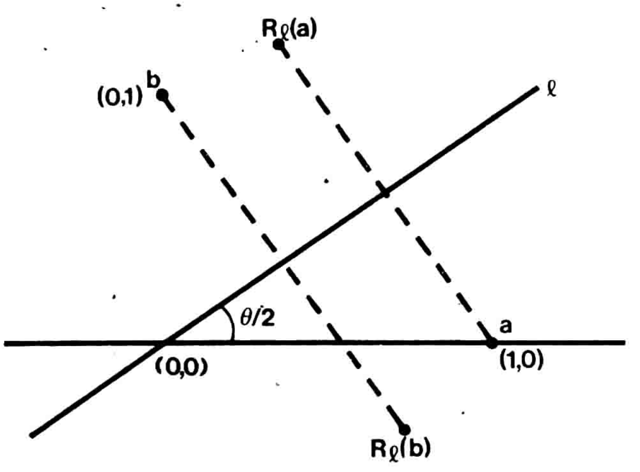
The group $O(n)$ has a normal subgroup, the **special orthogonal group** $SO(n) = O(n) \cap SL(n,\mathbf{R})$ consisting of the orthogonal matrices whose determinant is $+1$. This subgroup of $O(n)$ has index 2.

Examples $O(1) = \{\pm 1\}$, $SO(1) = \{+1\}$.

The group $O(2)$ consists of 2×2 matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ whose columns are orthonormal. An elementary calculation shows that there is a θ such that

$$\begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \text{ and } \begin{bmatrix} b \\ d \end{bmatrix} = \pm \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix},$$

SO(2) consists of the matrices $\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$. This matrix represents a rotation through the angle θ about the origin. As SO(2) has index 2 in O(2), it has two cosets, one is SO(2) itself and the other is SO(2). $\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$, which consists of the matrices $\begin{bmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{bmatrix}$. This matrix represents a reflection R_ℓ in the line ℓ which makes the angle $\theta/2$ with the x-axis.



The relationship between O(n) and GL(n,R)

A matrix in GL(n,**R**) has independent columns and a matrix in O(n) has orthonormal columns. The **Gram-Schmidt process** transforms an independent set of vectors into an orthonormal set, so it can be used to define a mapping GL(n,**R**) $\to$ O(n). To make this precise it is convenient to introduce the group $T_+(n)$ consisting of the set of **upper triangular** $n \times n$ matrices whose diagonal entries are positive (T for triangular, + for positive).

Proposition $T_+(n)$ is a subgroup of GL(n,**R**).

Proof If $A \in T_+(n)$, then $\det A = a_{11}\, a_{22} \dots a_{nn} > 0$. Hence $T_+(n)$ is a subset of GL(n,**R**). The matrix A lies in $T_+(n)$ if and only if

$$a_{ij} = 0 \text{ for } i > j$$
$$\text{and} \qquad a_{ii} > 0.$$

If A, B $\in T_+(n)$ then

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

If $i > j$, then either $i > k$ or $k \geq i > j$; in the first case $a_{ik} = 0$ and in the second $b_{kj} = 0$, so $(AB)_{ij} = 0$ in both cases. If $i = j$, then $(AB)_{ii} = a_{ii}b_{ii} > 0$. So $AB \in T_+(n)$ if A, B $\in T_+(n)$.

It remains to check that if $A \in T_+(n)$ then so is A^{-1}. Suppose $A \in T_+(n)$ and $AB = BA = I$. We show that $b_{ij} = 0$ for $i > j$ by downward induction on i. First consider $i = n$, then

$$0 = a_{nn}b_{nj} \text{ for } j < n, \text{ so } b_{nj} = 0.$$

Suppose that $b_{kj} = 0$ for all $j < k$ if $k > i$, then

$$0 = a_{ii}b_{ij} \text{ for } i > j, \text{ so } b_{ij} = 0.$$

Given that $b_{ij} = 0$ for $i > j$, one gets that

$$a_{ii}b_{ii} = 1 \text{ for all } i.$$

Hence as $a_{ii} > 0$ one sees that $b_{ii} > 0$. Hence $B \in T_+(n)$.

(The reader may prefer to go through this proof explicitly in the case $n = 2$.)

Theorem 1 For a given $A \in GL(n,\mathbf{R})$, there are unique matrices $B \in O(n)$, $C \in T_+(n)$ such that $A = BC$.

Proof As suggested above, we use the Gram-Schmidt process to construct B from A and then observe that they are related by $A = BC$ with $C \in T_+(n)$. In detail: let $a_1, a_2, \ldots, a_n$ be the columns of A. The first stage of the Gram-Schmidt process is to find an orthogonal set $f_1, f_2, \ldots f_n$. This is constructed by induction as follows.

$$f_1 = a_1,$$
$$f_k = a_k - \sum_{i=1}^{k-1} \{(a_k.f_i)/(f_i.f_i)\} f_i.$$

If F is the matrix with columns $f_1, f_2, \ldots, f_n$ then $F = AT_1$, where T_1 is in $T_+(n)$, and in fact T_1 has ones along the diagonal. Note that the matrix F is obtained from A by a sequence of elementary column operations, each new column involving only earlier columns. The second stage of the Gram-Schmidt process is to normalise the set $f_1, f_2, \ldots, f_n$, that is, let $b_i = f_i / \|f_i\|$. If B is the matrix whose columns are $b_1, b_2, \ldots, b_n$ then $B = FT_2$ where T_2 is a diagonal matrix with positive entries $(1/\|f_i\|)$ on the diagonal, hence $T_2 \in T_+(n)$. If $C = (T_1T_2)^{-1}$ we have $A = BC$ with $B \in O(n)$ and $C \in T_+(n)$. Moreover, it is clear from the formulae that the matrices B, C depend continuously on the original matrix A.

It remains to check the uniqueness of this decomposition. Suppose A has two such decompositions, B_1C_1 and B_2C_2 say. Then $D = B_2^{-1}B_1 = C_2C_1^{-1}$ is in $O(n) \cap T_+(n)$. But we will show that $O(n) \cap T_+(n) = \{I\}$ and so the decomposition is unique. Let $D \in O(n) \cap T_+(n)$ then $D^t = D^{-1}$ and as $T_+(n)$ is a subgroup we have $D^t \in T_+(n)$. But D^t is lower triangular so D must be diagonal, and therefore $D = D^t$, so using orthogonality, $D^2 = I$. So D has diagonal entries ± 1. As $D \in T_+(n)$ it has positive entries on the diagonal therefore $D = I$ as required.

Corollary $GL(n,\mathbf{R})$ is homeomorphic to $O(n) \times T_+(n)$.

Proof The homeomorphisms are constructed as follows: $A \in GL(n,\mathbf{R})$ is mapped to (B,C) and $(B,C) \in O(n) \times T_+(n)$ is mapped to BC. These are clearly mutual inverses. The map $(B,C) \to BC$ is continuous because matrix multiplication is continuous – the entries of BC are polynomials in the entries of B and C. The map $A \to (B,C)$ is also continuous for a similar reason.

Note $T_+(n)$ is homeomorphic to $\mathbf{R}^{n(n+1)/2}$. A matrix in $T_+(n)$ has $n(n-1)/2$ entries off the diagonal and each of these can be an arbitrary element in $\mathbf{R}$. There are n entries on the diagonal and each of these is an arbitrary element of $\mathbf{R}_+$, so that $T_+(n) \cong \mathbf{R}^{n(n+1)/2} \times (\mathbf{R}_+)^n$. But $\mathbf{R}_+$ is homeomorphic to $\mathbf{R}$ (under log and exp as inverse homeomorphisms).

So $GL(n,\mathbf{R})$ is homeomorphic to $O(n) \times \mathbf{R}^{n(n+1)/2}$.

Exercise The space $SL(n,\mathbf{R})$ is homeomorphic to $SO(n) \times \mathbf{R}^{(n^2+n-2)/2}$.

Examples The space $GL(1,\mathbf{R})$ is $\mathbf{R} \setminus \{0\}$, $O(1)$ is $\{\pm 1\}$ and so one can see directly that $GL(1,\mathbf{R})$ is homeomorphic to $O(1) \times \mathbf{R}$.

The group $O(2)$ is the union of $SO(2)$ and another coset of $SO(2)$ but $SO(2)$ is homeomorphic to the **circle** $S^1 = \{z \in \mathbf{C} \mid \|z\| = 1\}$, so $O(2)$ is homeomorphic to the union of two disjoint copies of S^1 and $GL(2,\mathbf{R})$ is homeomorphic to the union of two disjoint copies of $S^1 \times \mathbf{R}^3$.

Exercise Show that the group $GL(n,\mathbf{R})$ for $n > 1$ is not the direct product of its subgroups $O(n)$ and $T_+(n)$.

Exercise Show that GL(2,**R**) has many subgroups of order three but that $O(2) \times T_+(2)$ has only one such subgroup. Deduce that there is no isomorphism between GL(2,**R**) and $O(2) \times T_+(2)$. [It is true that GL(n,**R**) is not isomorphic to $O(n) \times T_+(n)$ for any $n \geq 2$, but the proof is more difficult for $n \geq 3$. For $n \geq 3$, show that the centre Z of G = GL(n,**R**) consists of the scalar matrices and that G/Z has no proper normal subgroups for n odd and only one such for n even. The group $O(n) \times T_+(n)$ modulo its centre has several proper normal subgroups.]

Affine Subspaces and Affine Independence

It is often necessary to consider lines, planes, etc. that do not pass through the origin. Linear subspaces always contain the origin but their cosets (in the additive group) do not and they are called affine subspaces. However the affine subspaces have an intrinsic definition.

A subset A of $\mathbf{R}^n$ is an **affine subspace** if $\lambda a + \mu b \in A$ for all $a, b \in A$ and all $\lambda, \mu \in \mathbf{R}$ such that $\lambda + \mu = 1$. A straightforward induction shows that the following is an equivalent condition:

$$\sum_{i=1}^{k} \lambda_i a_i \in A \text{ for all } a_i \in A \text{ and } \sum_{i=1}^{k} \lambda_i = 1.$$

If $V \subset \mathbf{R}^n$ is a linear subspace then it is easy to check that the set $V + x = \{v+x \mid v \in V\}$ is an affine subspace of $\mathbf{R}^n$ for any (fixed) $x \in \mathbf{R}^n$. Every affine subspace A is of this form, because if $a \in A$ and $V = A - a = \{x-a \mid x \in A\}$ then V is a linear subspace of $\mathbf{R}^n$. Let $\lambda \in \mathbf{R}$ and $x - a \in V$ then to check that $\lambda(x-a) \in V$ we must check that $\lambda(x-a) + a \in A$ but

$$\lambda(x-a) + a = \lambda x + (1-\lambda)a$$

and $x, a \in A$. If $x-a, y-a \in V$ then $(x-a) + (y-a) \in V$ because

$$(x-a) + (y-a) + a = x + y - a$$

which is a linear combination of elements of A, the sum of the coefficients being $1 + 1 - 1 = 1$ so $x + y - a \in A$.

Exercise If $a, b \in A$ and A is an affine subspace, show that $A - a = A - b$.

If A is an affine subspace, its **dimension** is the dimension of the linear subspace $A - a$ of $\mathbf{R}^n$.

If $X \subset \mathbf{R}^n$ is any subset, its **affine span** Aff(X) is defined as

$$\text{Aff}(x) = \{\sum_{i=1}^{k} \lambda_i x_i \mid x_i \in X, \sum_{i=1}^{k} \lambda_i = 1\}.$$

It is easy to check that Aff(x) is an affine subspace of $\mathbf{R}^n$ and that it is the smallest affine subspace containing X.

A set $X = \{x_0, x_1, \ldots, x_r\}$ is **affinely independent** if $\sum_{i=0}^{r} \lambda_i x_i = 0$ holds with $\sum_{i=0}^{r} \lambda_i = 0$ only if $\lambda_0 = \lambda_1 = \ldots = \lambda_r = 0$. It is easily checked that $\{x_0, x_1, \ldots, x_r\}$ is affinely independent if and only if $\{x_1-x_0, x_2-x_0, \ldots, x_r-x_0\}$ is a linearly independent set.

If $X = \{x_0, x_1, \ldots, x_r\}$ is affinely independent then Aff(X) has dimension r and X is called an **affine basis** for Aff(X). Note that an affine basis for an r-dimensional affine subspace has $r + 1$ elements. If $\{e_1, e_2, \ldots, e_r\}$ is a basis of a linear subspace V then an affine basis for V is $\{0, e_1, e_2, \ldots, e_r\}$.

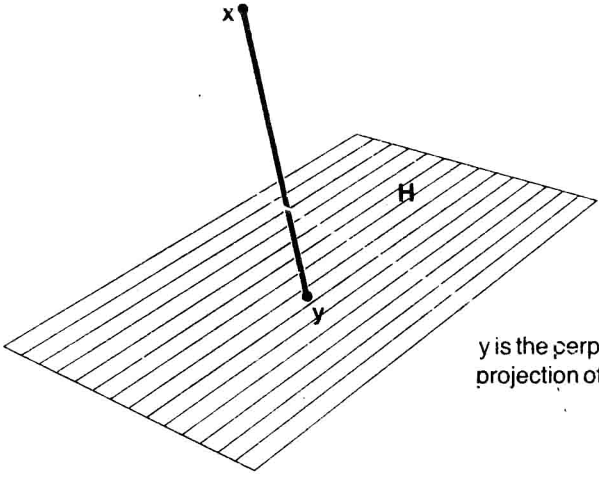
An affine subspace H of $\mathbf{R}^n$ whose dimension is $n - 1$ is called a **hyperplane**. If H is a linear hyperplane of $\mathbf{R}^n$, then there is a non-zero $x \in \mathbf{R}^n$ such that $H = \{x\}^{\perp}$. This is because one can choose an orthonormal basis for H and extend it (by a vector x) to an orthonormal basis for $\mathbf{R}^n$; it is then easy to check that $H = \{x\}^{\perp}$. Hyperplanes arise as the perpendicular bisectors of line segments.

Lemma If $a, b \in \mathbf{R}^n$ with $a \neq b$, then $B = \{x \mid d(x,a) = d(x,b)\}$ is a hyperplane in $\mathbf{R}^n$.

Proof It is clear that $(a+b)/2 \in B$ so we need to show that $H = B - (a+b)/2$ is an $(n-1)$ dimensional linear subspace. If $c = (a-b)/2$, it is easily checked using the translation invariance of distance that H is the set $\{x \mid d(x,c) = d(x,-c)\}$. If $c, e_2, e_3, \ldots, e_n$ is an orthogonal basis for $\mathbf{R}^n$, then

$e_2, e_3, \ldots, e_n$ is a basis for H.

If H is any hyperplane in $\mathbf{R}^n$ and $x \in \mathbf{R}^n$, then x can be written uniquely in the form $y + z$ where $y \in H$ and $z \perp H$.



y is the perpendicular projection of x onto H.

Mc e algebraically, let $a \in H$, then $H - a$ is a linear hyperplane, so $H - a = \{b\}^\perp$ for some b. There is a unique expression

$$x - a = \lambda b + c \text{ where } c \in H - a.$$

Let $y = c + a$, $z = \lambda b$, then $y \in H$ and $z \perp H$.

It remains to check the uniqueness. Suppose

$$y_1 + z_1 = y_2 + z_2 \quad \text{with} \quad y_1, y_2 \in H, z_1, z_2 \in H^\perp$$

Then $z_2 - z_1 = y_1 - y_2 \in H - y_2$ and $z_2 - z_1 \in H - y_2$, hence $z_1 = z_2$ and so $y_1 = y_2$.

If $A \subset \mathbf{R}^n$, $B \subset \mathbf{R}^m$ are both affine subspaces, a map $f: A \to B$ is an **affine map** if

$$f(\lambda a + \mu b) = \lambda f(a) + \mu f(b) \text{ for } a, b \in A \text{ and } \lambda + \mu = 1.$$

An affine map is therefore one that takes straight lines to straight lines because the straight line through the points a, b is the set $\{\lambda a + \mu b | \lambda, \mu \in \mathbf{R}, \lambda + \mu = 1\}$. If $a \in A$, $b \in B$ then $A-a$ and $B-b$ are linear spaces; if $L: A-a \to B-b$ is a linear map then the map $A_L: A \to B$ defined by

$$A_L(x) = L(x-a) + b$$

is an affine map. When checking this note carefully that L is only defined on $A - a$. All affine maps arise in this manner as we now show.

Lemma If $f: A \to B$ is an affine map then the map $L_f: A-a \to B-f(a)$ defined by

$$L_f(x) = f(x+a) - f(a)$$

is a linear map. The map f is obtained from L_f by the previous construction.

Proof We need to check that $L_f(x+y) = L_f(x) + L_f(y)$ and that $L_f(\lambda x) = \lambda L_f(x)$. To check the first we note that $x + a$, $y + a$, $a \in A$ and that $x + y + a = (x+a) + (y+a) - a$ is a combination of them, the sum of whose coefficients is 1. So

$$\begin{aligned} L_f(x+y) &= f(x+y+a) - f(a) \\ &= f(x+a) + f(y+a) - f(a) - f(a) \\ &= L_f(x) + L_f(y). \end{aligned}$$

$L_f(\lambda x) = f(\lambda x+a) - f(a)$
$= f(\lambda(x+a)+(1-\lambda)a) - f(a)$
$= \lambda f(x+a) + (1-\lambda)f(a) - f(a)$
$= \lambda L_f(x)$.

If $L = L_f$, it is easy to check that if one takes $b = f(a)$ then $f = A_L$.

An important special case is the following.

Corollary If f: $\mathbf{R}^n \to \mathbf{R}^n$ is an affine map then there exists $a \in \mathbf{R}^n$ such that the map L: $\mathbf{R}^n \to \mathbf{R}^n$ defined by $L(x) = f(x) - a$ is linear, so $f(x) = L(x) + a$.

Isometries of $\mathbf{R}^n$

We have already seen that translations T_a: $\mathbf{R}^n \to \mathbf{R}^n$ defined by $T_a(x) = x + a$ and orthogonal transformations T: $\mathbf{R}^n \to \mathbf{R}^n$ which are linear and satisfy $Tx.Ty = x.y$ are both examples of isometries of $\mathbf{R}^n$. We will show that all isometries are combinations of these two basic kinds. In fact isometries are affine maps. The first step is

Theorem 2 An isometry f: $\mathbf{R}^n \to \mathbf{R}^n$ is uniquely determined by the images $fa_0, fa_1, \ldots fa_n$ of a set $a_0, a_1, \ldots a_n$ of $(n+1)$ (affinely) independent points.

Proof Let f, g be isometries with $fa_i = ga_i$ for $0 \leq i \leq n$. Then $g^{-1}f$ is an isometry with $g^{-1}fa_i = a_i$. Let T be the translation defined by $Tx = x - a_0$ and let $b_i = T(a_i)$ for $0 \leq i \leq n$. Clearly, $b_0 = 0$, and the set $\{b_1, b_2, \ldots, b_n\}$ forms a basis for $\mathbf{R}^n$. We will show that $h = Tg^{-1}fT^{-1}$ is the identity, and this shows that $f = g$ as required.

Clearly $hb_i = b_i$ for $0 \leq i \leq n$, so if $y = hx$ we have that $d(x,0) = d(y,0)$ and $d(x,b_i) = d(y,b_i)$ for $1 \leq i \leq n$ because h is an isometry. Hence $x.x = y.y$ and $(x-b_i).(x-b_i) = (y-b_i).(y-b_i)$ for $1 \leq i \leq n$. By expanding these last n equations and manipulating one gets that $x.b_i = y.b_i$ for $1 \leq i \leq n$. As $b_1, b_2, \ldots, b_n$ is a basis, one has $x.z = y.z$ for every $z \in \mathbf{R}^n$, hence $x = y$, proving that h is the identity.

This proof shows that a point in $\mathbf{R}^n$ is uniquely determined by its distances from $n + 1$ independent points. Note that, in general, a point x is not uniquely determined by its distances from n independent points.

Theorem 3 If $\{a_0, a_1, \ldots a_n\}$ and $\{b_0, b_1, \ldots b_n\}$ are two sets of $(n+1)$ independent points in $\mathbf{R}^n$ with $d(a_i,a_j) = d(b_i,b_j)$ for $0 \leq i, j \leq n$ then there is an isometry f: $\mathbf{R}^n \to \mathbf{R}^n$ with $fa_i = fb_i$ for $0 \leq i \leq n$.

Proof Using translations we can clearly assume that $a_0 = b_0 = 0$. Then $\{a_1, a_2, \ldots a_n\}$ and $\{b_1, b_2, \ldots b_n\}$ are bases for $\mathbf{R}^n$, and it is easy to see that the hypotheses imply that $a_i.a_j = b_i.b_j$ for all i, j. Let g be the unique (non-singular) linear transformation such that $ga_i = gb_i$ for $1 \leq i \leq n$. Let $x - y = \Sigma \lambda_i a_i$, then $gx - gy = g(x-y) = \Sigma \lambda_i b_i$ by the linearity of g. So

$$d(gx,gy)^2 = \Sigma \lambda_i\lambda_j b_i.b_j = \Sigma \lambda_i\lambda_j a_i.a_j = d(x,y)^2.$$

Hence g is a linear isometry. The required f is the composition of g with a translation, so it is affine.

We have already proved on page 4 that every linear isometry is orthogonal. Theorems 2 and 3 therefore combine to show that if f is an isometry of $\mathbf{R}^n$ then $f(x) = Ax + a$ where $A \in O(n)$ and $a \in R^n$ so that every isometry is the composition of an orthogonal transformation and a translation. Hence every isometry is affine.

Exercise If $X \subset \mathbf{R}^n$ is any subset and g: $X \to \mathbf{R}^n$ is an isometric map, show that there is an isometry f: $\mathbf{R}^n \to \mathbf{R}^n$ such that $f|X$ is g. If the affine subspace defined by X has dimension $n - r$, prove that the set of such isometries forms a coset of $O(r)$.

We will now show how every isometry can be written as a product of reflections. This gives an alternative approach to understanding isometries and yields independent proofs of some of our previous results.

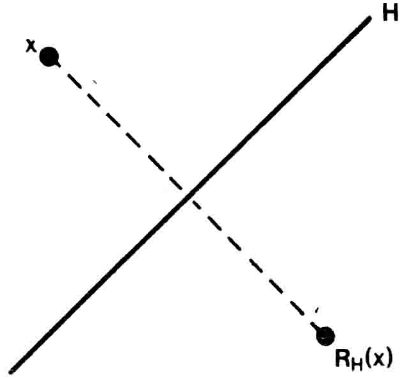
Definition If H is a hyperplane in $\mathbf{R}^n$, **reflection in H** is the isometry R_H of $\mathbf{R}^n$ defined by

$$R_H(x) = y - z$$

where $x = y + z$ with $y \in H$ and $z \perp (H-y)$ using the decomposition given on page 8.

Note that $R_H{}^2$ is the identity and that R_H leaves every point of H fixed. If H is the perpendicular bisector of ab, R_H interchanges a and b.

Example In $\mathbf{R}^3$, regard H as a two-sided mirror, then $R_H(x)$ is the mirror image of x.



Exercise If $0 \in H$, that is if H is a linear hyperplane, show that R_H is orthogonal. If a is a unit vector perpendicular to H, show that $R_H(x) = x - 2(x.a)a$.

Theorem 4 Any isometry f: $\mathbf{R}^n \to \mathbf{R}^n$ that is the identity on an affine (n–r)-dimensional subspace A (that is, fa = a for each $a \in A$) can be expressed as the product of at most r reflections in hyperplanes that contain A. Any isometry can be expressed as the product of at most (n+1) reflections.

Note The last sentence can be regarded as a special case of the first if one makes the (usual) convention that the empty set has dimension −1.

Proof Choose (n–r+1) independent points $a_0, a_1, \ldots, a_{n-r}$ in A and extend them to a set $a_0, a_1, \ldots, a_n$ of (n+1) independent points in $\mathbf{R}^n$. Let $b_i = fa_i$, so $b_i = a_i$ for $0 \leq i \leq n-r$. As f is an isometry, $d(a_i, a_j) = d(b_i, b_j)$ so if H is the perpendicular bisector of $a_{n-r+1}b_{n-r+1}$ it is clear that $a_i \in H$ for $0 \leq i \leq n-r$. The idea now is to consider R_Hf, this is the identity on an (n–r+1) dimensional affine subspace and so one can use induction on r to give the required result. In detail: $R_Hf = R_{H_1} \ldots R_{H_s}$ where $H_1 \ldots H_s$ are s hyperplanes ($s \leq r-1$) containing A and a_{n-r+1}, then $f = R_H R_{H_1} \ldots R_{H_s}$ is a product of at most r reflections in hyperplanes containing A. To prove the last sentence of the theorem, let H be the hyperplane bisecting af(a) for some $a \in \mathbf{R}^n$. Then R_Hf fixes a 0-dimensional affine subspace, and so is the product of at most n reflections. So f is the product of at most n + 1 reflections.