4 Geometrical imagery

4.1 The real axis

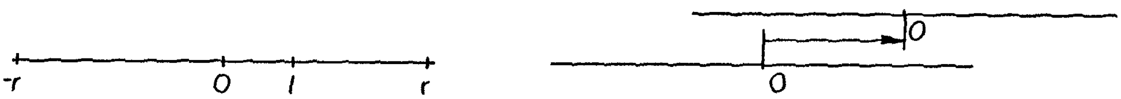
A point is punctated by a pick and a pang of the pencil. This is how it is grossly made visible to the eye. To distinguish between points when referring to them they are distinctly named and labeled, usually by a capital Roman letter. Two points determine one line, and two intersecting lines have but one point in common.

Let O be one point of a line. The line is densely packed with points endlessly extending to the left and to the right of this origin point O. Similarly, our system of real numbers is also densely and endlessly packed downward and upward of zero.

It is a simple, yet one of the most fruitful ideas of mathematics, in that it convenes analysis and geometry, relating numbers to graphs: to locate any point P of a line by its distance to the left or to the right of origin point O.

The line with its myriad points is drawn by a deft uninterrupted draft of the pen. Then origin point 0 is marked on it. Another point at a unit distance to the right of 0 is marked 1. With two labeled points on the line, symmetry is broken, we know to tell on it left from right and the line becomes an *axis*. Positive real number r scores a point at distance r to the right of 0, and negative -r to the left of 0, as in Fig.1a. By an assumption on the nature of these things there corresponds to every real number one point on the line, and conversely, every point on the line is reached by some real number.

The preeminence of this modest plot is that it permits us to *visualize* relationships between real numbers. Magnitude turns thereby into distance—big and small, many and few become long and short— and sign turns thereby into sense—being to the right or to the left of the origin.



(a) Fig. 4.1 (b)

The choice of a right positive direction is an arbitrary but universally accepted convention for the real axis. Restricting ourselves to one dimension, a left axis can be made right through *reflection* in the origin only.

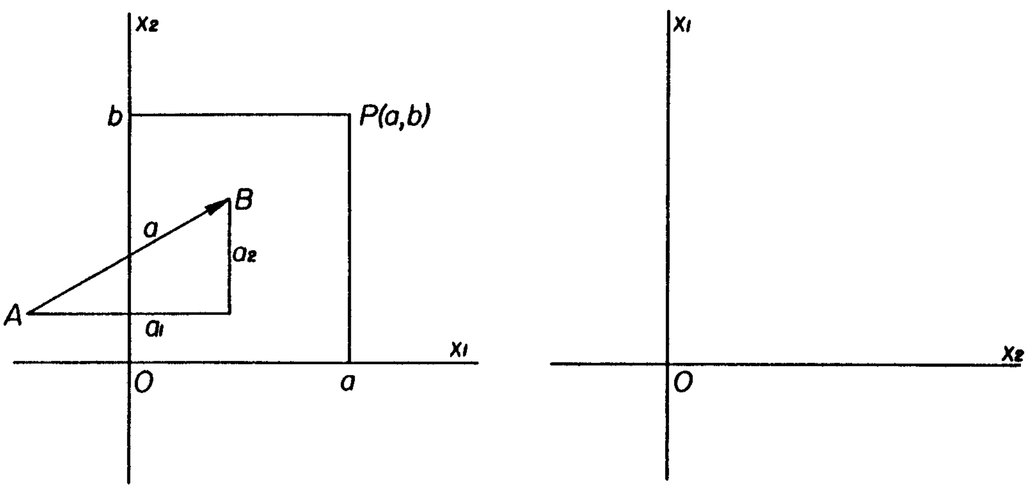
Consider now a *translation* of the real axis relative to itself (see Fig. 1b) in which all points of the axis move in unison as though on a sliding rigid ruler. To graphically represent such a translation, every point of the axis is associated with an arrow or *directed segment* to show the *magnitude* and *direction* (*sense*) of the movement. Since the directed segment is not tied to any specific point it is termed *free*.

Obviously, a right axis cannot be made left by such sliding alone, but *rotation* in the plane of the page will do it.

4.2 Translation of space

More interesting things happen in the plane than on a line as greater freedom fosters greater geometrical possibilities. Analytic geometry is born with the seminal idea of Descartes to unfold the plane with a coordinate system consisting of two perpendicular axes intersecting at their common origin. For distinction one axis is labeled x_1 and the other x_2. An ordered pair of real numbers a and b is geometrically represented as point P shown in Fig. 2a. Different number pairs appear as distinct points, and the string of close points $P(x_1, x_2)$ generated from an analytic rule that relates coordinates x_1 and x_2 graphs a curve in the plane for us to contemplate.

There are other interesting and occasionally even more appropriate ways to numerically locate a point in the plane than by Cartesian coordinates, but we shall not dwell on these here.



(a) Fig. 4.2 (b)

We do not care to distinguish between a right and a left axis in the plane as any two axes can be made to coincide by rotation and translation, but with the system of two axes there are two distinctions. Cartesian coordinate system of Fig. 2b is different from that in Fig. 2a in that the two cannot be made to coincide by planar rotation and translation alone. A reflection is also needed.

Directed segment $\vec{AB}$ drawn in Fig. 2a shows a translation of the whole x_1x_2 plane relative to itself displacing a distance equal to the length of segment AB, parallel to it, and in the arrowed sense. All points of the plane are identical in this respect, directed segment $\vec{AB}$ is free, and can be put anywhere in the plane. We refer to the translation $\vec{AB}$ by a single letter, say a, write $a = \vec{AB}$, and see it as a result of two consecutive sliding a_1 and a_2 performed in any order, parallel to the x_1 axis and then parallel to the x_2 axis. A positive a_1 means movement to the right and a negative a_1 movement to the left, and the same for a_2. The two perpendicular translations that affect $\vec{AB}$ are its *components*, and are ordered columnwise in $a = [a_1\ a_2]^T$. To distinguish between the components of a translation and the coordinates of point P we write the latter as $P(a_1,\ a_2)$.

Some prefer to call translations $[a_1\ 0]^T, [0\ a_2]^T$ the components of translation a, rather than a_1, a_2.

The three-dimensional Cartesian coordinate system consists of three perpendicular axes labeled x_1, x_2, x_3 intersecting at a common origin 0. An ordered triplet $P(a, b, c)$ is marked as point P in space in the manner shown in Fig. 3a.

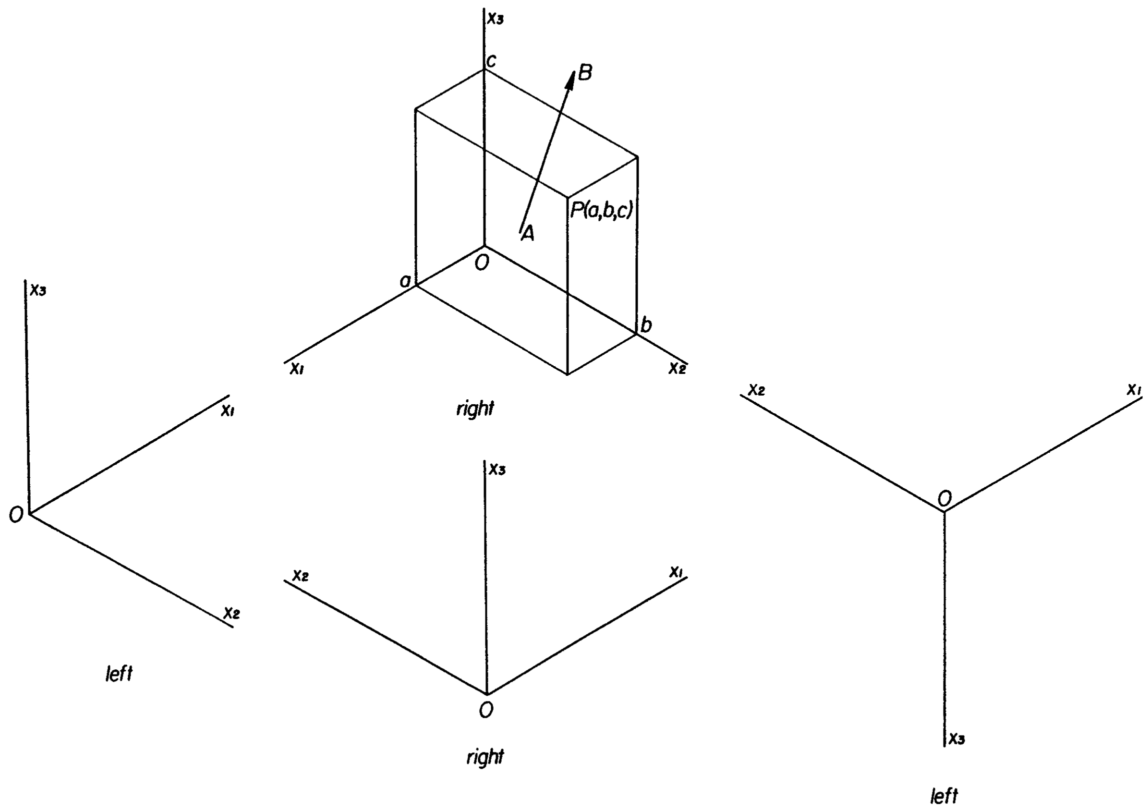


Fig. 4.3: top (a)

(b) (c) (d)

The coordinate system of Fig. 3a is a right-hand system. Reversal of one axis turns it into a left-hand system as in Fig. 3b. A left-hand system cannot be made to coincide with a right-hand system by space rotation and translation alone; a reflection is also needed. Reversal of two axes as in Fig. 3c leaves the system in its kind, while reversal of all axes produces a left-hand system as in Fig. 3d. We shall use the right-hand coordinate system exclusively.

Free directed segment $\vec{AB}$ graphically represents the translation of the whole space relative to itself, a translation that we consider the result of three consecutive sliding a_1, a_2, a_3, performed in any order along x_1, x_2, x_3, respectively. Again we designate translation $\vec{AB}$ by a single letter and write $a = [a_1\ a_2\ a_3]^T = \vec{AB}$ to analytically describe the translation in terms of its three perpendicular movements.

Translation is the most plausible geometrical explanation as to why a directed segment is free. But once we accept this interpretation, we need not be bound by the representation and may consider the directed segment as a geometrical (even ideal physical) entity in its own right. Directed segments that have the same *length*, the same *inclination* relative to the coordinate axes, and the same *sense* represent the same translations and are by themselves *the same* or *equal*. In this sense the ordered list $a = [a_1\ a_2\ a_3]^T$ uniquely determines directed segment $\vec{AB}$, wherever point A is. Conversely, given (geometrically) directed segment $\vec{AB}$ and Cartesian coordinate system $0x_1x_2x_3$, $a = [a_1\ a_2\ a_3]^T$ is unequivocally measured and written for it. An equivalence is thus established between the ordered list, or vector, $a = [a_1\ a_2\ a_3]^T$ and the free directed segment $\vec{AB}$.

Exercises

4.2.1. A Cartesian coordinate system with two orthogonal reference axes is impractical in large scale earth measurements and navigation. Explain how point P of the plane can be located relative to the two fixed points A and B. Also, how point P of space can be located relative to the three fixed points A, B and C.

4.3 Vector geometry

In the previous section we considered the equivalence of an ordered triplet of real numbers and a free directed segment. We formalize the conclusion in

Theorem 4.1. *Vectors $a = [a_1\ a_2\ a_3]^T$ and $b = [b_1\ b_2\ b_3]^T$ represent equal directed segments if and only if $a = b$, if and only if $a_1 = b_1,\ a_2 = b_2,\ a_3 = b_3$.*

In the previous chapters we upheld the convention of writing vectors columnwise, attesting to their origin in systems of linear equations where it is natural to write them in this manner. Geometers prefer to write vectors in the typographically more sensible row form

$a = (a_1, a_2, a_3)$ and are indifferent to transposition. We shall, however, remain faithful to our accepted matrix usage and consistently add the superscript T.

Vector geometry is created in discovering the geometrical meaning of the basic vector operations of:

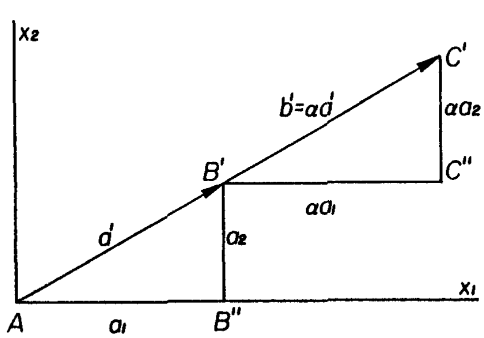
1. Multiplication of vector a by scalar α, $b = \alpha a$.
2. Addition of vectors a and b into vector c, $c = a + b$.
3. Multiplication of vector a by matrix A, $b = Aa$.
4. The scalar product of a and b, $\alpha = a^T b = b^T a$.

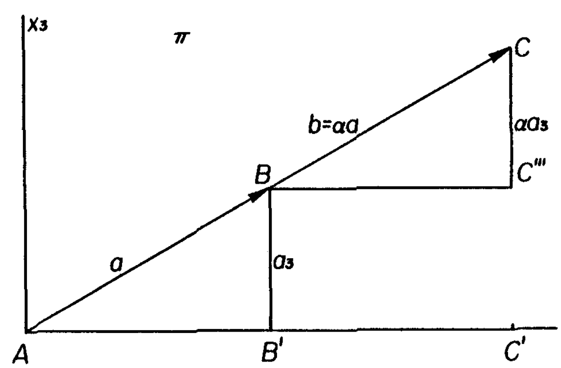
The zero vector $o = [0\ 0\ 0]^T$ represents immobility; it is directed segment $\vec{AA}$ of zero length but no direction to speak of.

For a geometrical interpretation of

$$b = \alpha a = [\alpha a_1\ \alpha a_2\ \alpha a_3]^T \tag{4.1}$$

refer to Fig. 4. Translation $a = \vec{AB}$ of space consists of translation $a' = [a_1\ a_2\ 0]^T$ in the x_1x_2 plane followed by sliding a_3 along x_3. Translation $b = \vec{BC}$ consists of translation $b' = [\alpha a_1\ \alpha a_2\ 0]^T$ in the x_1x_2 plane followed by sliding αa_3 along x_3. Triangles $AB^{''}B'$ and $B'C^{''}C'$ of Fig. 4a are similar, and directed segments $a' = \vec{AB'}$ and $b' = \vec{B'C'}$ are parallel, or colinear. Points A, B', C' are on one line, or colinear, and they are so only if $b' = \alpha a'$.





(a) Fig. 4.4 (b)

Plane π of Fig. 4b contains points A, B', C' and the x_3 axis. Directed segments a and b are on π. Triangles $AB'B$ and $BC^{'''}C$ are similar, and $a = \vec{AB}$ and $b = \vec{BC}$ are parallel, or colinear. Points A, B, C are on one line, or colinear, and they are so only if $b = \alpha a$. Directed segments a and αa have the same inclination relative to the coordinate axes but are not necessarily of the same length and sense.

When $\alpha = -1, (-1)\ a$ is shortened to $-a$. If $a = \vec{AB}$ then $-a = \vec{BA}$. Multiplication by a negative α *reverses* the sense of the corresponding directed segment as shown schematically in Fig. 4.5.

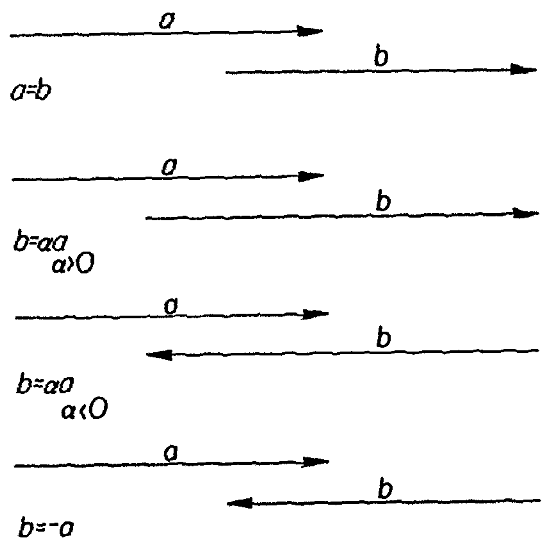


Fig. 4.5

Equivalence or *isomorphism* between different branches of mathematics is always a happy discovery. It allows us to go back and forth between the fields whenever circumstances warrant it. We started with vectors as ordered lists of numbers and found a geometrical image for the triplet. Now we carry a geometrical physical concept back to numbers. Length is a purely geometrical concept, but we introduce the

Definition. *Magnitude* $\|a\|$ *of vector* $a = [a_1\ a_2\ a_3]^T$ *is the length of the directed segment it represents. Vector* u *is said to be a unit vector if* $\|u\| = 1$.

Inasmuch as the components of a are perpendicular translations

$$\|a\| = \sqrt{a_1^2 + a_2^2 + a_3^2} \tag{4.2}$$

and

$$\|\alpha a\| = \sqrt{\alpha^2 a_1^2 + \alpha^2 a_2^2 + \alpha^2 a_3^2} = \sqrt{\alpha^2 (a_1^2 + a_2^2 + a_3^2)} = |\alpha| \|a\|. \tag{4.3}$$

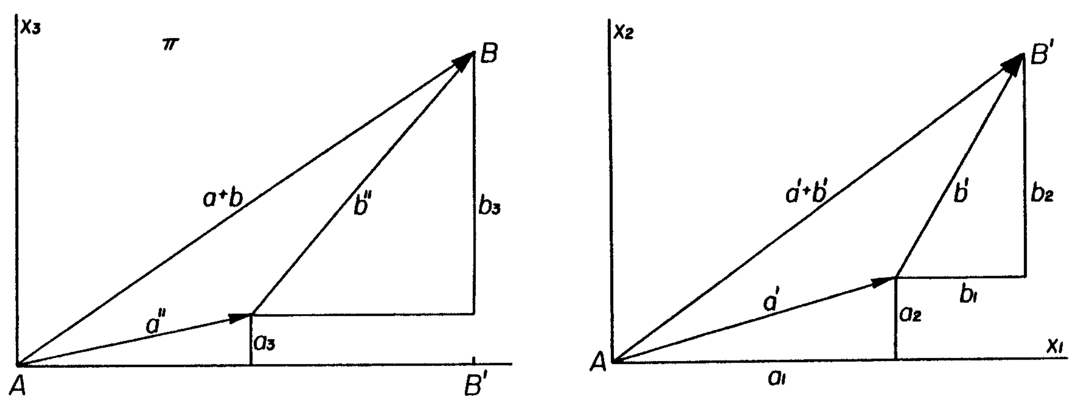
Two distinct points A and B fix a line in space, but vector $a = \vec{AB}$ determines only its inclination: it fixes the line up to an arbitrary translation of all its points. We shall say that vector a generates the line. For any choice of α, vector $\alpha a = \alpha \vec{AB}$ is parallel to the line through A and B. The totality of vectors formed by αa with a given a and an arbitrary α, namely all vectors parallel to the line through A and B, constitutes a *one-dimensional vector space spanned* by a. Vector $b = \alpha a$, that is parallel to the line through A and B, is said to be an *element* of this vector space.

Vector summation

$$a + b = [a_1 + b_1 \ a_2 + b_2 \ a_3 + b_3]^T \tag{4.4}$$

is geometrically interpreted as the compound translation of a followed by b, or b followed by a. Figure 6a shows the projection of a and b on the x_1x_2 plane and the resultant translation

$$a' + b' = [a_1 \ a_2 \ 0]^T + [b_1 \ b_2 \ 0]^T = [a_1 + b_1 \ a_2 + b_2 \ 0]^T. \tag{4.5}$$



(a) Fig. 4.6 (b)

Plane π of Fig. 6b contains points A, B', B and the x_3 axis. Vectors a'' and b'' are the projections of a and b, respectively, on π. Translation $a + b$ is completed with the sliding $a_3 + b_3$ along x_3, and $a, b, \ a + b$ close a triangle in space. Writing $a + b = -c$ we are ready to announce

Theorem 4.2. *The vector sum of a, b, c is zero if and only if the corresponding directed segments close a triangle.*

Recall that vectors a, b, c are free and they close a triangle after translation. See Fig. 7.

For four vectors, $a + b + c + d = o$ if and only if the directed segments close a skew (not necessarily planar) quadrilateral in space. It is easily shown by adding first a and b, and then forming $(a + b) + c + d = o$.

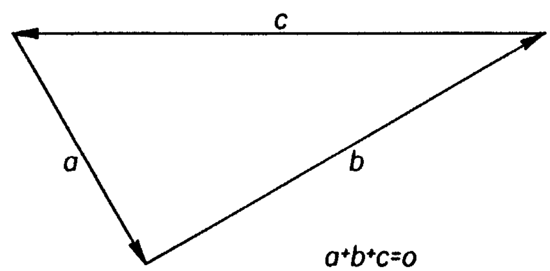


Fig. 4.7

We collapse the triangle and return to parallel vectors.

Theorem 4.3. *Two vectors $a = [a_1\ a_2\ a_3]^T$ and $b = [b_1\ b_2\ b_3]^T$ are colinear; they represent parallel directed segments, if and only if two scalars α, β, not both zero, exist so that*

$$\alpha a + \beta b = o. \tag{4.6}$$

Proof. It is a symmetric statement of the geometrical interpretation of vector by scalar multiplication to include the border case of zero vectors. If in eq. (4.6) $b = o$ and $a \neq o$, then $\alpha = 0\ \beta \neq 0$ satisfy the equation and a and b are colinear. The zero vector is taken thereby to have an arbitrary direction and is colinear with any vector. Writing $b = \alpha a$ would have excluded the colinearity of $a = o,\ b \neq o$; while $a = \beta b$ would have excluded the colinearity of $b = o,\ a \neq o$. End of proof.

Two vectors in the same one-dimensional vector space are colinear. Theorem 4.3 states that the two vectors are colinear and stand for parallel directed segments if and only if homogeneous system (4.6) of three equations in two unknowns has a nontrivial solution. If the only solution to $\alpha a + \beta b = o$ is the trivial $\alpha = \beta = 0$, then the directed segments represented by a and b are not parallel. It is a useful formality to consider immovability as being parallel to any translation.

Three noncolinear points, A, B, C, that is, three points not on one line, uniquely fix a plane in space. But the two noncolinear vectors $a = \vec{AB}$, $b = \vec{AC}$, are free and fix the plane only up to an arbitrary space translation of all its points. We shall say that the two noncolinear vectors a and b generate the plane. Conversely, any two vectors are *coplanar.* To be additionally colinear the two vectors must satisfy the restriction of Theorem 4.3. Notice that the noncolinearity of a and b implies that they are both nonzero.

Four points A, B, C, D are coplanar only conditionally and so are the three directed segments $a = \vec{AB},\ b = \vec{AC},\ c = \vec{AD}$.

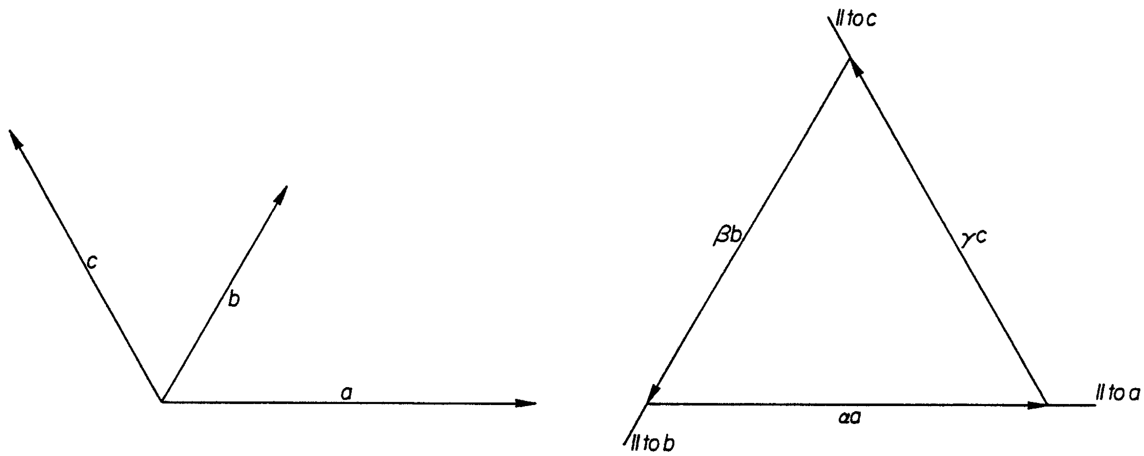


Fig. 4.8

Theorem 4.4. *Three vectors $a = [a_1\ a_2\ a_3]^T, b = [b_1\ b_2\ b_3]^T, c = [c_1\ c_2\ c_3]^T$ are coplanar if and only if three scalars α, β, γ not all zero, exist so that*

$$\alpha a + \beta b + \gamma c = o. \tag{4.7}$$

Proof. Refer to Fig. 8 and assume that no two of the vectors are colinear. If a, b, c are coplanar, then three coplanar nonconcurrent lines drawn parallel to them enclose a triangle and, non-unique, α, β, γ are found so that $\alpha a + \beta b + \gamma c = o$.

If a, b, c are not coplanar, then $\alpha a, \beta b, \gamma c$ cannot be made to close a space triangle with any α, β, γ.

If two vectors, say a and b, are colinear we set in eq. (4.7) $\gamma = 0$ and are assured by Theorem 4.3 that α and β are not both zero. End of proof.

Vectors a, b, o are coplanar, but three *noncoplanar* vectors are all nonzero.

The condition of coplanarity for a, b, c, namely that the homogeneous vector equation $\alpha a + \beta b + \gamma c = o$ have a nontrivial solution is elegantly expressed in determinant notation as

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = 0. \tag{4.8}$$

Corollary 4.5. *If vectors a, b, c are coplanar, and a, b noncolinear, then unique scalars α' and β' exist so that*

$$c = \alpha' a + \beta' b. \tag{4.9}$$

In other words vector c can be written as a unique linear combination of a and b.

Proof. Since a, b, c are coplanar there exists a nontrivial solution to $\alpha a + \beta b + \gamma c = o$. Since a, b are noncolinear γ is nonzero, for otherwise $\alpha a + \beta b = o$ with nonzero α or β contradicts the noncolinearity assumption on the vector pair a, b. Division by γ and some rewriting leads to

$$c = -\frac{\alpha}{\gamma} a - \frac{\beta}{\gamma} b \tag{4.10}$$

which with $\alpha' = -\alpha/\gamma,\ \beta' = -\beta/\gamma$ produces eq. (4.9).

Linear combination (4.9) is unique. Suppose that there are two such combinations $c = \alpha' a + \beta' b$ and $c = \alpha'' a + \beta'' b$. Subtraction yields

$$(\alpha' - \alpha'')a + (\beta' - \beta'')b = o \tag{4.11}$$

and $\alpha' = \alpha''$ and $\beta' = \beta''$ by virtue of a and b being noncolinear. End of proof.

Two noncolinear vectors generate a plane in space. The totality of vectors in this plane constitutes a *two-dimensional* vector space spanned by a and b. Vector c that has the property of being in (or rather parallel to) the plane is said to be an element of the vector space and can be uniquely expressed as $c = \alpha a + \beta b$. A vector space is designated by one single capital Roman letter, say V, and to state concisely the fact that c is in V we write $c \in V$. Any nonzero vector in the plane, i.e. in V, spans a one-dimensional vector space, say

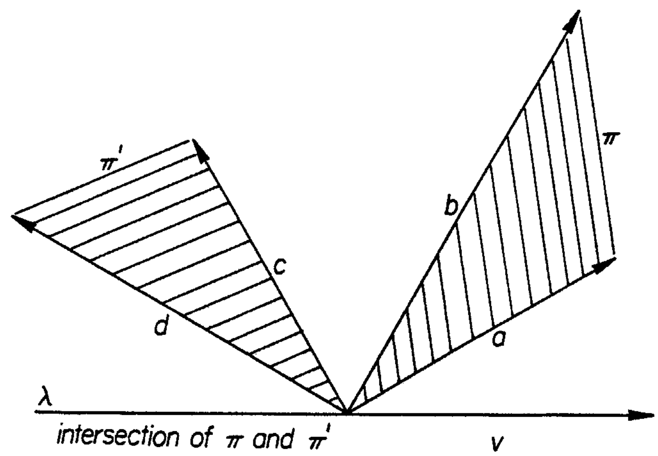


Fig. 4.9

W. Vector space W is a *subspace* of V since if $a \in W$ then also $a \in V$. The converse is of course not true.

Corollary 4.6. *If a, b, c are three noncoplanar vectors, and d a given vector, then unique scalars α, β, γ exist so that*

$$d = \alpha a + \beta b + \gamma c. \tag{4.12}$$

Proof. Vectors a, b, c are all nonzero. If d is colinear with one of the three given vectors, or if it is coplanar with two of them, then we revert to colinearity and coplanarity discussed earlier. We assume therefore that d is noncoplanar with any pair of the three vectors a, b, c. With reference to Fig. 9 let π be the plane generated by a and b, and π' the plane generated by c and d. One must use one's imagination when looking at Fig. 9 since it depicts a three dimensional situation *projected* on a plane —this page. By the assumption that the four vectors a, b, c, d are not coplanar, planes π and π' are not parallel and intersect at line λ. Line λ is at once in π and π'. Let v be a vector parallel to λ. According to Corollary 4.5 there are unique scalars α' and β' such that $v = \alpha' a + \beta' b$, and also unique α'' and β'' so that $d = \alpha'' c + \beta'' v$. Hence $d = \alpha'' c + \beta''(\alpha' a + \beta' b) = \alpha'' c + \beta'' \alpha' a + \beta'' \beta' b$ and unique expression (4.12) is recovered. End of proof.

Three noncoplanar vectors span the three-dimensional space we stand in. This vector space — our whole wide, wonderful and mysterious universe — is designated by R^3. All vectors with three *real* components are in R^3. All vectors parallel to a plane are in a two-dimensional subspace of R^3, while all vectors parallel to a line are in a one-dimensional subspace of R^3.

All vectors on two intersecting planes π and π' constitute two different two-dimensional vector subspaces V and V' of R^3. All vectors parallel to intersection line λ constitute a one-dimensional subspace L of R^3 which is the *intersection* of spaces V and V', $L = V \cap V'$.

On the other hand, the *union* of all vectors in V and V' is not vector space R^3. In fact the union is not a vector space at all.

Speaking of translations, eq. (4.12) implies that arbitrary translation d of R^3 can be uniquely decomposed into a sequence of three slides parallel to the three noncoplanar vectors a, b, c. The magnitudes of these three partial translations are $|\alpha|\,\|a\|, |\beta|\,\|b\|, |\gamma|\,\|c\|$. Factors α, β, γ are the components of the translation along span vectors a, b, c, respectively. Some prefer to call α, β, γ the components of the translation only if $\|a\| = \|b\| = \|c\| = 1$. Others prefer to call $\alpha a, \beta b, \gamma c$ the components of d.

We are now in the position to produce a geometrical interpretation of Ax, and more interestingly the solution of $Ax = f$,

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}. \tag{4.13}$$

Let a_1, a_2, a_3 denote the three columns of A in eq. (4.11). Then

$$Ax = x_1 \begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \end{bmatrix} + x_2 \begin{bmatrix} A_{12} \\ A_{22} \\ A_{32} \end{bmatrix} + x_3 \begin{bmatrix} A_{13} \\ A_{23} \\ A_{33} \end{bmatrix} = f \tag{4.14}$$

is abbreviated as

$$Ax = x_1 a_1 + x_2 a_2 + x_3 a_3 = f. \tag{4.15}$$

Left-hand side vector $x_1 a_1 + x_2 a_2 + x_3 a_3$ of eq.(4.15) means translation of space by three consecutive sliding parallel to $a_1, a_2, , a_3$ with components x_1, x_2, x_3. Solution of the vector equation amounts then to finding the three factors x_1, x_2, x_3 of the translation so that $x_1 a_1 + x_2 a_2 + x_3 a_3$ adds up to given translation f.

If a_1, a_2, a_3 in eq. (4.15) are *noncoplanar*, then there is a unique solution to translation problem (4.15) for any f. If a_1, a_2, a_3 are coplanar then there are solutions to the translation problem if and only if f is coplanar with them. The non-uniqueness of the solution, when it exists, stems from the *redundancy* in the components. If a_1, a_2, a_3 are coplanar or colinear, and f is not coplanar with them or, respectively, not colinear with them, then translation problem (4.15) has no solution.

In the language of vector spaces we describe the solution of $Ax = f$ this way: Vectors a_1, a_2, a_3 —the three columns of A — span subspace V of R^3, possibly even R^3 itself. If $f \in V$ then $Ax = f$ is soluble but if f is not an element of V, then $Ax = f$ is insoluble.

The *rank* of matrix A is the *dimension* of the space spanned by its columns. Matrix $A = A(3 \times 3)$ is nonsingular if and only if its columns span R^3 —if and only if they are noncoplanar, for if the columns of A are coplanar, then vector $x = [x_1\ x_2\ x_3]^T \neq o$ is found so that $Ax = x_1a_1 + x_2a_2 + x_3a_3 = o$. On the other hand, if they are not coplanar, then $Ax \neq o$ whenever $x \neq o$.

It remains for us to do a geometrical interpretation of the scalar product, which we accomplish in the next section.

Exercises

4.3.1. Find α so that the length of $a = \alpha[1\ \ -2\ 1]^T + [1\ 0\ \ -1]^T$ is $\sqrt{8}$.

4.3.2. If points A, B, C are colinear, and point M is the midpoint of segment AB, show that $\vec{CM} = \frac{1}{2}(\vec{CA} + \vec{CB})$.

4.3.3. Are the vectors $a = [1\ \ -2\ 1]^T$, $b = [1\ 0\ \ -1]^T$ and $c = [1\ 1\ 1]^T$ coplanar? If yes produce α, β, γ, not all zero, so that $\alpha a + \beta b + \gamma c = o$.

4.3.4. Are the vectors $a = [1\ \ -1\ 2]^T$, $b = [-2\ 1\ \ -3]^T$ and $c = [4\ \ -1\ 5]^T$ coplanar? If yes produce α, β, γ, not all zero, so that $\alpha a + \beta b + \gamma c = o$.

4.3.5. Find α so that $a = [1\ \ -1\ \ -1]^T, b = [1\ 1\ 2]^T, c = [\alpha\ 1\ \alpha]^T$ are coplanar.

4.5.6. Is vector $c = [-1\ -4\ 7]^T$ an element of the vector space spanned by the two noncolinear vectors $a = [-2\ 1\ \ -1]^T$ and $b = [1\ \ -2\ 3]^T$? If yes find components α and β of c so that $c = \alpha a + \beta b$.

4.4 The scalar dot product

What is written in the coherent matrix notation as $a^Tb = b^Ta$ is also often more simply written in geometrical circumstances as $a \cdot b = b \cdot a$, and is accordingly termed the *dot product* of a and b. We know it to be a very basic binary operation in linear algebra and it has an interesting and important geometrical interpretation.

In terms of their components,

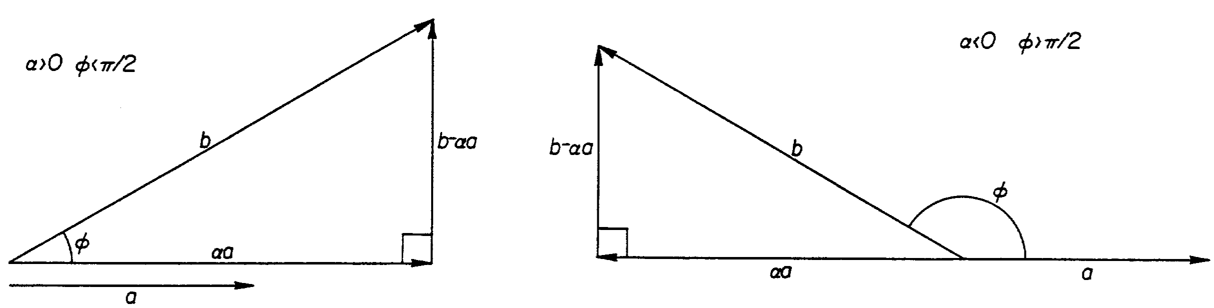
$$a^Tb = b^Ta = a_1b_1 + a_2b_2 + a_3b_3 \tag{4.16}$$

so that

$$\|a\|^2 = a^Ta \tag{4.17}$$

and $\|a\| = 0$ only if $a = o$.

Let a and b be nonzero vectors, drawn in Fig. 10 with angle (measure) ϕ between their positive directions.Factor α is chosen so that directed segment αa is *perpendicular* to $b - \alpha a$. In Fig. 10a $\alpha \geq 0$, while in Fig. 10b $\alpha \leq 0$. Evidently $\alpha = 0$ only if directed segments a and b are mutually perpendicular.



(a) Fig. 4.10 (b)

Because $\cos(180 - \phi) = -\cos\phi$ and by the definition of the cosine

$$\alpha\|a\| = \|b\|\cos\phi. \tag{4.18}$$

Pythagoras assures us that

$$\alpha^2a^Ta + (b - \alpha a)^T(b - \alpha a) = b^Tb \tag{4.19}$$

and

$$\alpha(\alpha a^T a - a^T b) = 0. \tag{4.20}$$

If $\alpha \neq 0$, then

$$a^T b = \alpha a^T a \tag{4.21}$$

becoming, with α from eq. (4.18)

$$a^T b = \|a\| \|b\| \cos\phi \tag{4.22}$$

which is the geometrical interpretation of $a^T b$. Equations (4.21) and (4.22) imply that when $\alpha \neq 0$, when the directed segments are not perpendicular, $a^T b \neq 0$.

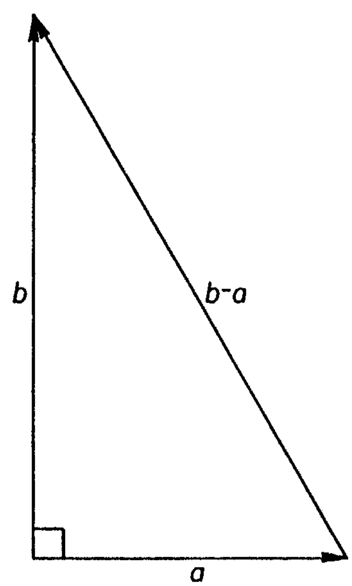


Fig. 4.11

The case $\alpha = 0$, or a and b orthogonal, is resolved with reference to Fig. 11. From Pythagoras we have that

$$a^T a + b^T b = (b-a)^T (b-a) = b^T b + a^T a - 2a^T b \tag{4.23}$$

and if $\alpha = 0$, then $a^T b = 0$. The outcome of all this is

Theorem 4.7. *Directed segments a and b are perpendicular or orthogonal if and only if $a^T b = 0$.*

Being done with this, we write

$$(a-b)^T(a-b) = a^Ta + b^Tb - 2a^Tb \tag{4.24}$$

and with eq. (4.22) elicit from it the expression

$$\|a-b\|^2 = \|a\|^2 + \|b\|^2 - 2\|a\|\,\|b\|\cos\phi \tag{4.25}$$

which is the generalization of Pythagoras' theorem known in trigonometry as the law of cosines.

The Cauchy-Schwarz inequality

$$|a^Tb| \le \|a\|\,\|b\| \tag{4.26}$$

also ensues immediately from eq. (4.20). Since the inequality is obtained by setting 1 for $|\cos\phi|$, it follows that *equality* holds if and only if $\phi = 0^o$ or $\phi = 180^o$; that is when a and b are colinear.

Let V and W be two vector subspaces of R^3. If for every $v \in V$ and every $w \in W$, $v^Tw = 0$, then the two subspaces are *orthogonal*. The vectors parallel to two orthogonal lines constitute such subspaces, as do the vectors parallel to a plane and parallel to a line that is orthogonal to the plane. Vectors parallel to two orthogonal planes do not constitute two orthogonal subspaces of R^3.

Exercises

4.4.1. Write all vectors that are orthogonal to $a = [1\ \ -1]^T$.

4.4.2. Write all *unit* vectors orthogonal to $a = [1\ \ -1]^T$.

4.4.3. Write all vectors orthogonal to both $a = [1\ 1]^T$ and $b = [1\ 2]$.

4.4.4. Write all vectors orthogonal to $a = [1\ \ -1\ 2]^T$.

4.4.5. Write all vectors orthogonal to both $a = [1\ 0\ 1]^T$ and $b = [1\ \ -2\ 1]^T$.

4.4.6. Compute the angle between vectors $a = [1\ 0\ 1]^T$ and $b = [-1\ 1\ \ -1]^T$.

4.4.7. Write all vectors x in the plane of $a = [1\ 1\ 1]^T,\ b = [1\ 2\ -1]^T$ that are orthogonal to $c = [-1\ 3\ 1]^T$.

4.4.8. Write vectors x in the plane of $a = [1\ -1\ 2]^T$ and $b = [-2\ 0\ 1]^T$ that make 60^o with $c = [1\ -2\ 2]^T$. What are the conditions for the existence of such vectors?

4.4.9. Let noncoplanar vectors a, b, c be such that $a^T b = 0$. Find in the plane of a, b vector x that makes minimum angle with c. Show that the plane generated by the optimal x and c is orthogonal to the plane of a, b. What vector x makes widest angle with c?

4.5 Application to plane and solid geometry

One cannot fully appreciate the ingenuity of the calculus of directed segments until one has witnessed the astonishing spectacle of algebra solving geometrical problems. Vectors are used in this section to prove an assortment of plane and solid geometry theorems on incidence of points and lines, colinearity, coplanarity, proportions, and orthogonality.

Let A, M, B be three points fixed in a Cartesian coordinate system with origin O, as in Fig. 12. Directed segment $\vec{OA}$ is called the *position vector* of point A, and we shall denote it

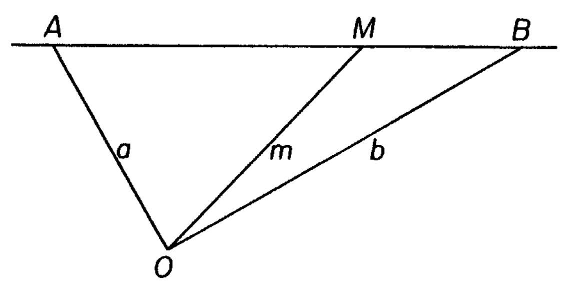


Fig. 4.12

by a. Correspondingly, $m = \vec{OM}$ and $b = \vec{OB}$. The convenience of this notation lies in the fact that the *components* of *vector* a, and the *coordinates* of *point* A are the same numbers. In this notation $\vec{AB} = b - a$.

Theorem 4.8. *If points A, B are distinct, then a necessary and sufficient condition that point M be colinear with them is the existence of scalar λ so that*

$$m = (1 - \lambda)a + \lambda b. \tag{4.27}$$

Proof. Points A, M, B are colinear if and only if λ exists so that $\vec{AM} = \lambda \vec{AB}$. In terms of position vectors a, m, b

$$m - a = \lambda(b - a) \tag{4.28}$$

and eq. (4.27) results. End of proof.

Equation (4.27) is the *parametric equation* of the line fixed by A, B in space.

The variation of parameter λ along the line through points A and B is shown in Fig. 13.

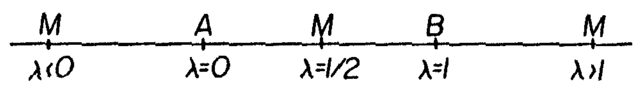


Fig. 4.13

Equation (4.27) may be written symmetrically as

$$m = \alpha a + \beta b \quad \alpha + \beta = 1 \tag{4.29}$$

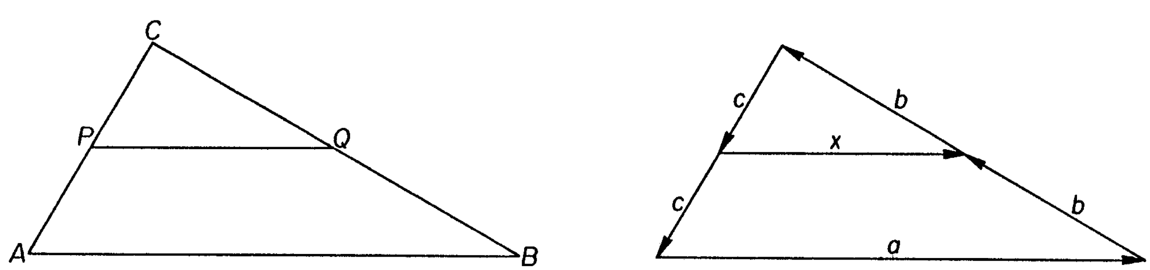
in which case α, β are called the *barycentric coordinates* of M. Or even

$$\alpha a + \beta b + \mu m = o \quad \alpha + \beta + \mu = 0 \tag{4.30}$$

with at least one nonzero among α, β, μ. For equation (4.30) the restriction that points are distinct may be removed. For example, if $a = b$, then $\mu = 0$, $\alpha + \beta = 0$, satisfy the equation for any α.

Theorem 4.9. *If in triangle ABC, P and Q are midpoints, then $PQ \| AB$ and $PQ = \frac{1}{2}AB$.*

Proof. Refer to Fig. 14. The triangle can be considered given by vertices or by edges.



(a) Fig. 4.14 (b)

We shall first take it given by edges and put directed segments along the sides as in Fig. 14b. Vector equations can now be written for closed polygons. Around the larger triangle $a + 2b + 2c = o$, and around the smaller $b + c + x = o$. Elimination of $b + c$ between the two leaves us with $a = 2x$, proving that vectors a and x are parallel and that $\|x\| = \frac{1}{2}\|a\|$. End of proof.

Components did not enter in the above proof for an *intrinsic* property of the triangle, hence the remarkable compactness of the algebra.

Consideration of the triangle given by vertices suggests a position vector proof to the theorem that is short and elegant. Recall that if O is an arbitrary origin in the plane, then $p = \vec{OP}$ and $q = \vec{OQ}$. Since P and Q are midpoints we have from eq. (4.27) that

$$p = \frac{1}{2}(a + c), \quad q = \frac{1}{2}(b + c) \tag{4.31}$$

and upon elimination of c obtain

$$q - p = \frac{1}{2}(b - a) \tag{4.32}$$

meaning that $\vec{PQ} = \frac{1}{2}\vec{AB}$, and the proof is done.

Once the *algebraic* statement of the *geometric* problem is written down the proof is mechanically produced, an advantage of algebraic methods over purely geometric.

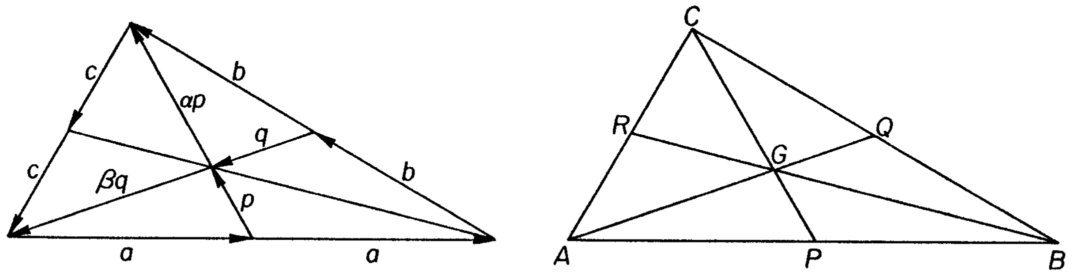
Theorem 4.10. *If in triangle ABC, P, Q, R are midpoints then the three medians PC, QA, RB are concurrent and intersect at a ratio* $1/2$.

Proof. Refer to Fig. 15. First we consider the triangle given by edges and write for:

$$\begin{aligned}
&\text{triangle } ABC \quad 2a + 2b + 2c = o, \\
&\text{triangle } QCG \quad b - \alpha p - q = o, \\
&\text{triangle } APG \quad a + p + \beta q = o, \\
&\text{triangle } CAG \quad 2c - \beta q + \alpha p = o.
\end{aligned} \tag{4.33}$$

Elimination of a, b, c from among the above four vector equations results in

$$(-2 + \alpha)p + (2 - \beta)q = o \tag{4.34}$$



(a) Fig. 4.15 (b)

and since p and q are noncolinear the above equation holds only if $\alpha = 2$ and $\beta = 2$. This is a typical intersection ratio for medians and the third must pass through the intersection point of any two. End of proof.

A simple proof to Theorem 4.10 is given using position vectors by showing that all medians pass through G positioned by

$$g = \frac{1}{3}(a + b + c) \tag{4.35}$$

where a, b, c are the position vectors of vertices A, B, C, respectively.

Indeed, since P is the midpoint between A and B

$$p = \frac{1}{2}(a + b) \tag{4.36}$$

and

$$g = \frac{2}{3}p + \frac{1}{3}c. \tag{4.37}$$

implying, according to Theorem 4.8, that PGC are colinear, and that G divides segment PC at a $1/2$ ratio. It is true for any median, and the proof is done.

Point G given by eq. (4.35) is called the center of area of triangle ABC. If the vertices of the triangle are considered as equal point masses, then point G is also the center of mass of the system.

A generalization of Theorem 4.10 (see Fig. 16, where all lines are straight) is given next, *ab initio*, in its algebraic position vector formulation.

Theorem (Ceva) 4.11. *If in the plane of triangle ABC point G is given by*

$$g = \zeta_1 a + \zeta_2 b + \zeta_3 c \qquad \zeta_1 + \zeta_2 + \zeta_3 = 1 \tag{4.38}$$

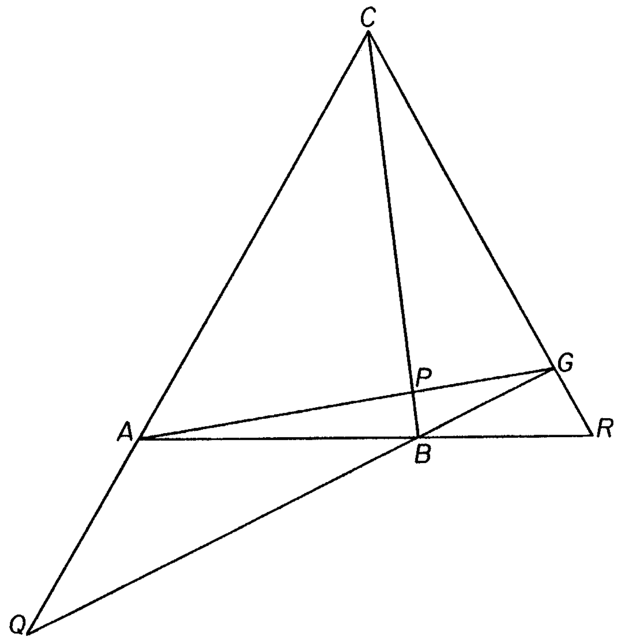


Fig. 4.16

$\zeta_i \neq 0,\ \zeta_i \neq 1,$ and points P, Q, R are such that

$$\begin{aligned} p &= \lambda_1 b + \lambda_1' c \qquad \lambda_1 + \lambda_1' = 1 \\ q &= \lambda_2 c + \lambda_2' a \qquad \lambda_2 + \lambda_2' = 1 \\ r &= \lambda_3 a + \lambda_3' b \qquad \lambda_3 + \lambda_3' = 1, \end{aligned} \tag{4.39}$$

then

$$\begin{aligned} \lambda_1 = \frac{\zeta_2}{\zeta_2+\zeta_3},\ \lambda_2 = \frac{\zeta_3}{\zeta_3+\zeta_1},\ \lambda_3 = \frac{\zeta_1}{\zeta_1+\zeta_2} \\ \lambda_1' = \frac{\zeta_3}{\zeta_2+\zeta_3},\ \lambda_2' = \frac{\zeta_1}{\zeta_3+\zeta_1},\ \lambda_3' = \frac{\zeta_2}{\zeta_1+\zeta_2} \end{aligned} \tag{4.40}$$

and

$$\frac{\lambda_1}{\lambda_1'}\ \frac{\lambda_2}{\lambda_2'}\ \frac{\lambda_3}{\lambda_3'} = 1. \tag{4.41}$$

Proof. To eq. (4.39) we add

$$\begin{aligned} p &= \alpha_1 a + \alpha_1' g \qquad \alpha_1 + \alpha_1' = 1 \\ q &= \alpha_2 b + \alpha_2' g \qquad \alpha_2 + \alpha_2' = 1 \\ r &= \alpha_3 c + \alpha_3' g \qquad \alpha_3 + \alpha_3' = 1 \end{aligned} \tag{4.42}$$

and eliminate g from them with the aid of eq. (3.38). Equating first p of eq. (4.39) with that of eq. (4.42) results in

$$(\alpha_1 + \alpha_1' \zeta_1)a + (\alpha_1' \zeta_2 - \lambda_1)b + (\alpha_1' \zeta_3 - \lambda_1')c = o \tag{4.43}$$

and since A, B, C are distinct and noncolinear eq. (4.43) dictates that

$$\alpha_1 + \alpha_1' \zeta_1 = 0, \ \alpha_1' \zeta_2 - \lambda_1 = 0, \ \alpha_1' \zeta_3 - \lambda_1' = 0. \tag{4.44}$$

From the first of eqs. (4.44) we have, with $\alpha_1 + \alpha_1' = 1$, that $\alpha_1' = 1/(1 - \zeta_1)$ or $\alpha_1' = 1/(\zeta_2 + \zeta_3)$. The two other equations yield the expressions for λ_1 and λ_1'. In exactly the same way we obtain the expressions for λ_2, λ_2' and λ_3, λ_3'. End of proof.

Theorem 4.12. *If points A, B, C are noncolinear, then point M is coplanar with them if and only if scalars α, β, γ exist so that*

$$m = \alpha a + \beta b + \gamma c \quad \alpha + \beta + \gamma = 1. \tag{4.45}$$

Proof. A necessary and sufficient condition that $\vec{AM}$ be coplanar with $\vec{AB}$ and $\vec{AC}$ is the existence of scalars β and γ so that $\vec{AM} = \beta \vec{AB} + \gamma \vec{AC}$. Or in terms of position vectors a, b, c, m,

$$m - a = \beta(b - a) + \gamma(c - a) \tag{4.46}$$

or

$$m = (1 - \beta - \gamma)a + \beta b + \gamma c \tag{4.47}$$

which with $\alpha = 1 - \beta - \gamma$ becomes (4.45). End of proof.

Equation (4.45) is the parametric equation of the plane in space.

Theorem 4.12 may be restated with no preference given to point M in saying that A, B, C, M are coplanar if and only if $\alpha, \beta, \gamma, \mu$ exist, not all zero, so that

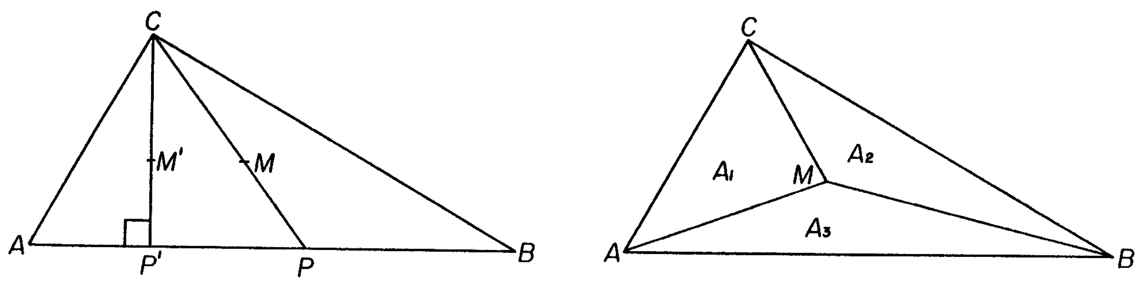
$$\alpha a + \beta b + \gamma c + \mu m = o \quad \alpha + \beta + \gamma + \mu = 0. \tag{4.48}$$

If A, B, C are colinear then eq. (4.48) holds with $\mu = 0$.

Triplet α, β, γ in eq. (4.45) subject to $\alpha+\beta+\gamma=1$, are the barycentric coordinates of point M in the plane through A, B, C. We verify with reference to Fig. 17a that $\vec{PM} = \gamma \vec{PC}$ and also $\vec{P'M'} = \gamma \vec{P'C}$, if MM' is parallel to PP', and conclude that

$$\alpha = \frac{A_1}{A},\ \beta = \frac{A_2}{A},\ \gamma = \frac{A_3}{A} \qquad A_1 + A_2 + A_3 = A \tag{4.49}$$

A_1, A_2, A_3 being the areas of the three sub-triangles in Fig. 17b. If M is not outside the triangle, then $0 \le \alpha \le 1,\ 0 \le \beta \le 1,\ 0 \le \gamma \le 1$, but if M is below the line through A, B, then $\gamma < 0$.



(a) Fig. 4.17 (b)

Vector calculus is even more helpful in three-dimensional, solid, geometry where perspective drawing distorts angle and distance, and where non-intersecting lines in space intersect on paper.

Theorem 4.13. *Let points A, B, C, D be the vertices of a tetrahedron. If point M is the center of area of face ABC, then DM is said to be a median of the tetrahedron. All four medians in the tetrahedron are concurrent and intersect at the ratio* 1/3.

Proof. We shall prove that all medians of the tetrahedron pass through point G (Fig. 18) fixed by

$$g = \frac{1}{4}(a+b+c+d) \tag{4.50}$$

wherever origin O is. Indeed,

$$m = \frac{1}{3}(a+b+c) \tag{4.51}$$

and

$$g = \frac{3}{4}m + \frac{1}{4}d. \tag{4.52}$$

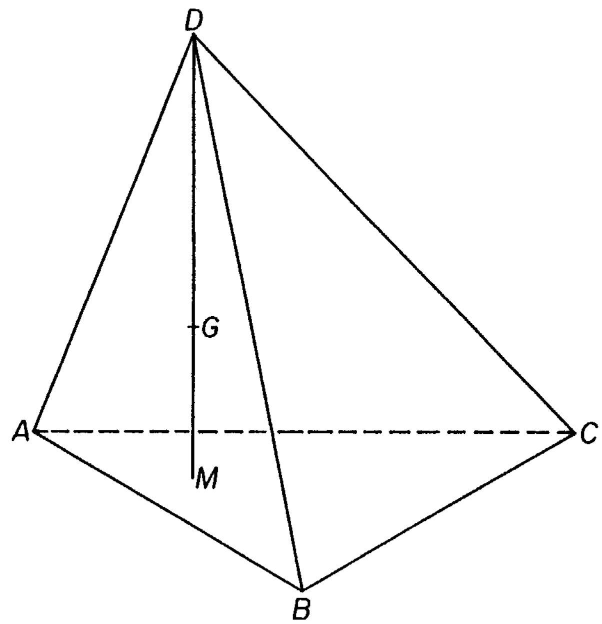


Fig. 4.18

Points MGD are colinear, and G divides MD at the typical ratio of 1/3. It is the same for all medians and the proof is done.

Point G is the center of volume of the tetrahedron or the center of mass of the vertices.

No vector calculus text is crowned without the triumphant algebraic proof to the celebrated theorem of Desargue. But first we prove useful

Lemma 4.14. *Let A, B, C, D be coplanar points with no three being colinear. Denote $a = \vec{OA}, b = \vec{OB}, c = \vec{OC}, d = \vec{OD}$ for point O not on the plane of points A, B, C, D. If both*

$$\alpha a + \beta b + \gamma c + \delta d = o \quad \alpha + \beta + \gamma + \delta = 0$$

and (4.53)

$$\alpha' a + \beta' b + \gamma' c + \delta' d = o \quad \alpha' + \beta' + \gamma' + \delta' = 0$$

hold nontrivially, then scalar λ exists so that

$$\alpha' = \lambda\alpha, \ \beta' = \lambda\beta, \ \gamma' = \lambda\gamma, \ \delta' = \lambda\delta. \tag{4.54}$$

Proof. Multiplication by λ and subtraction produces

$$(\alpha' - \lambda\alpha)a + (\beta' - \lambda\beta)b + (\gamma' - \lambda\gamma)c + (\delta' - \lambda\delta)d = o. \tag{4.55}$$

One of the coefficients, say that of a, may be made zero with λ so that $\alpha' - \lambda\alpha = 0$. By our assumption that B, C, D are noncolinear, and O out of their plane, the remaining coefficients of b, c, d must vanish and eq. (4.54) is obtained. End of proof.

Two triangles are said to be *copolar* if the lines joining their corresponding vertices happen to be *concurrent*. The triangles are said to be *coaxial* if the extensions to their corresponding sides intersect at three *colinear* points. Triangles ABC and $A'B'C'$ of Fig. 19 are copolar and coaxial at once.

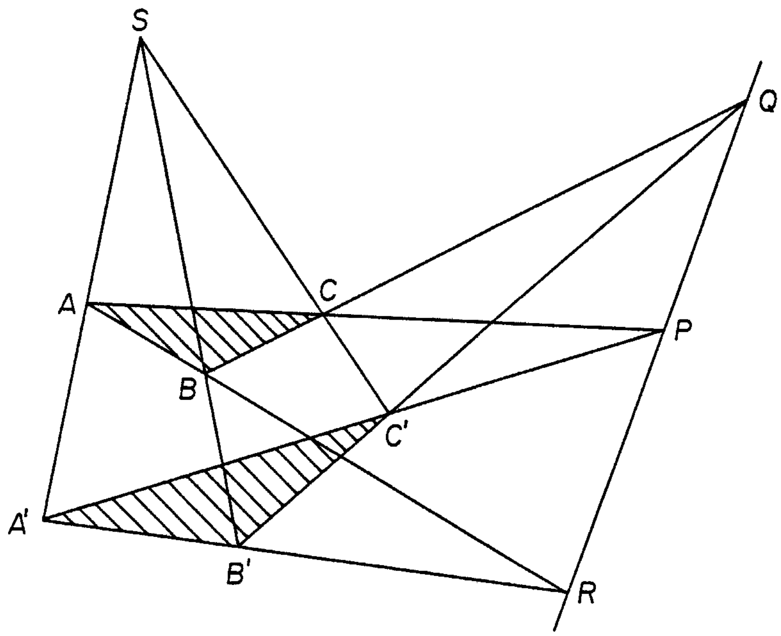


Fig. 4.19

Theorem (Desargue) 4.15. *Copolar triangles are coaxial, and conversely.*

Proof. We shall prove the first part of the theorem only, namely that copolar triangles are coaxial. The converse is left as an exercise. We prove, with reference to Fig. 19, that if lines AA', BB', CC' intersect at common point S, then points P, Q, R are colinear.

The linear algebraic statement of the theorem is expressed in the following three sets of position vector equations,

$$\begin{array}{ll} s = \alpha a + \alpha' a' & \alpha + \alpha' = 1 \\ s = \beta b + \beta' b' & \beta + \beta' = 1 \\ s = \gamma c + \gamma' c' & \gamma + \gamma' = 1 \end{array} \tag{4.56}$$

$$\begin{array}{llll} p = \pi a + \omega c & \pi + \omega = 1 & p = \pi' a' + \omega' c' & \pi' + \omega' = 1 \\ q = \sigma b + \tau c & \sigma + \tau = 1 & q = \sigma' b' + \tau' c' & \sigma' + \tau' = 1 \\ r = \rho b + \mu a & \rho + \mu = 1 & r = \rho' b' + \mu' a' & \rho' + \mu' = 1. \end{array} \tag{4.57}$$

Considering first equations that involve a, a', b, b' we deduce from eqs. (4.56) and (4.57) that

$$\begin{aligned} \alpha a + \alpha' a' - \beta b - \beta' b' &= o \\ \rho b + \mu a - \rho' b' - \mu' a' &= o. \end{aligned} \tag{4.58}$$

By Lemma 4.14 scalar λ exists so that

$$\begin{matrix} \lambda\alpha = \mu & \lambda\alpha' = -\mu' \\ -\lambda\beta = \rho & -\lambda\beta' = -\rho' \end{matrix} \tag{4.59}$$

and hence

$$\lambda(\alpha - \beta) = \mu + \rho = 1 \tag{4.60}$$

so that $\lambda = 1/(\alpha - \beta)$.

The equality $\alpha = \beta$ imports that $\alpha' = \beta' = 1 - \alpha$, implying that point R is at infinity, or that segment AB is parallel to $A'B'$. We wish to reject this possibility for now.

In a likewise manner we obtain

$$\begin{matrix} (\alpha - \beta)r = \alpha a - \beta b & \alpha \neq \beta \\ (\beta - \gamma)q = \beta b - \gamma c & \beta \neq \gamma \\ (\gamma - \alpha)p = \gamma c - \alpha a & \gamma \neq \alpha \end{matrix} \tag{4.61}$$

and then

$$(\alpha - \beta)r + (\beta - \gamma)q + (\gamma - \alpha)p = o \tag{4.62}$$

proving, by way of Theorem 4.8, that R, Q, P are colinear. End of proof.

Figure 19 acquires sudden sense and clarity if we imagine it depicting a three-dimensional situation whereby triangle ABC floats above triangle $A'B'C'$ on an inclined plane. Point S can be imagined as being a source of light with rays that pierce exactly through the corresponding vertices of the two triangles. Triangle $A'B'C'$ is the shadow, or the perspective projection, that triangle ABC casts upon the lower plane.

Algebraic proofs to geometrical theorems are methodical but the writing can become lengthy. Synthetic proofs require insights but can be short and elegant. Look again at Fig. 19 and consider it showing triangles ABC and $A'B'C'$ being on inclined planes in space. Triangles SAC and $SA'C'$ are coplanar, as are the lines through A, C and A', C', and they intersect at P. Triangles SAB and $SA'B'$ are coplanar and hence intersection point R. Triangles SBC and $SB'C'$ are coplanar and hence intersection point Q. Points P, Q, R are on the intersection of the plane of triangle ABC and the plane of triangle $A'B'C'$, and they are therefore colinear.

Desargue's theorem says in effect that every pair of arbitrary triangles can be considered the perspective projection, or the shadow, of each other. It results simply from the fact that two triangles that share a vertex are at once copolar and coaxial, as is seen in Fig. 20.

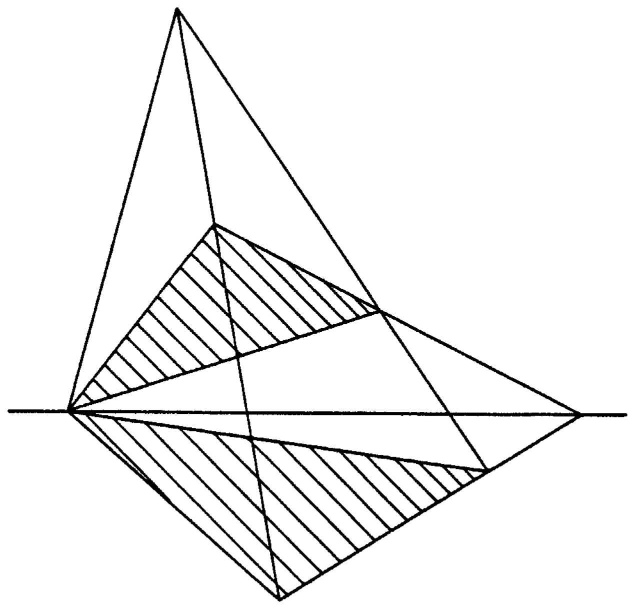


Fig. 4.20

None of the previous theorems of this section involved angle or distance. Vectors provide short proofs for the forthcoming theorems on orthogonalities in space. While reading the next theorem look at Fig. 21 but be wary of the fact that it is an angle-distorting projection of a three-dimensional situation. Symbol $\perp$ stands for orthogonal.

Theorem 4.16. *Let points A and B on the two nonintersecting orthogonal lines λ_1 and λ_2 in space be such that $AB \perp \lambda_2$. Then also $CB \perp \lambda_2$ for any point C on λ_1.*

Proof. Around triangle ABC

$$\alpha p + r + s = o. \tag{4.63}$$

Premultiplication of eq. (4.63) by q^T collapses it into the scalar equation

$$\alpha q^T p + q^T r + q^T s = 0 \tag{4.64}$$

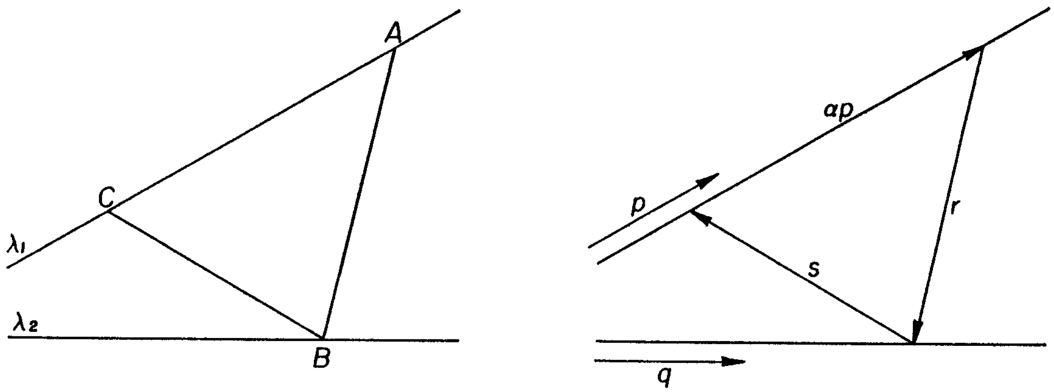


Fig. 4.21

but since $p^Tq = 0$ by virtue of λ_1 and λ_2 being orthogonal, and because $q^Tr = 0$ by construction, all that remains of eq. (4.64) is $q^Ts = 0$, and the proof is done.

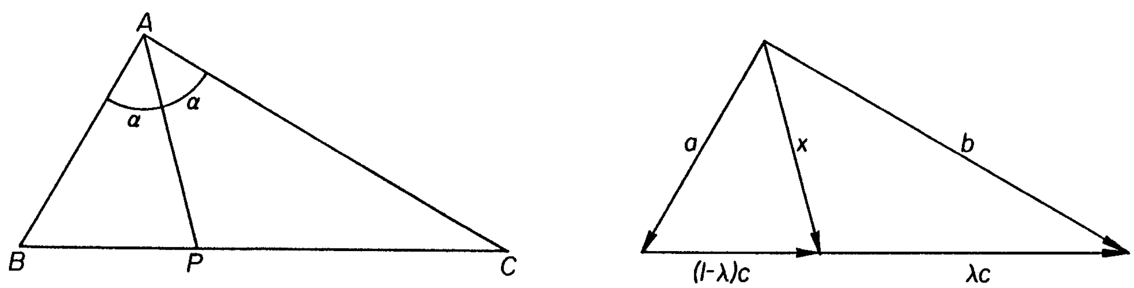


Fig. 4.22

Theorem 4.17. *If in triangle ABC the angle bisector of A meets BC at P, then $BP/PC = AB/AC$.*

Proof. With reference to Fig. 22 we write the two vector equations

$$\begin{aligned} x + \lambda c - b &= o \\ a + c - b &= o \end{aligned} \tag{4.65}$$

and seek to prove that $(1-\lambda)/\lambda = \|a\|/\|b\|$. Elimination of c from system (4.65) results in

$$x = (1-\lambda)b + \lambda a \tag{4.66}$$

and by virtue of x being a bisector

$$\frac{a^Tx}{\|a\|} = \frac{b^Tx}{\|b\|}. \tag{4.67}$$

With eq. (4.66) and $a^Tb = \|a\|\ \|b\|\cos 2\alpha$, eq. (4.67) becomes

$$\|b\|(1-\lambda)(\cos 2\alpha - 1) = \lambda\|a\|(\cos 2\alpha - 1) \tag{4.68}$$

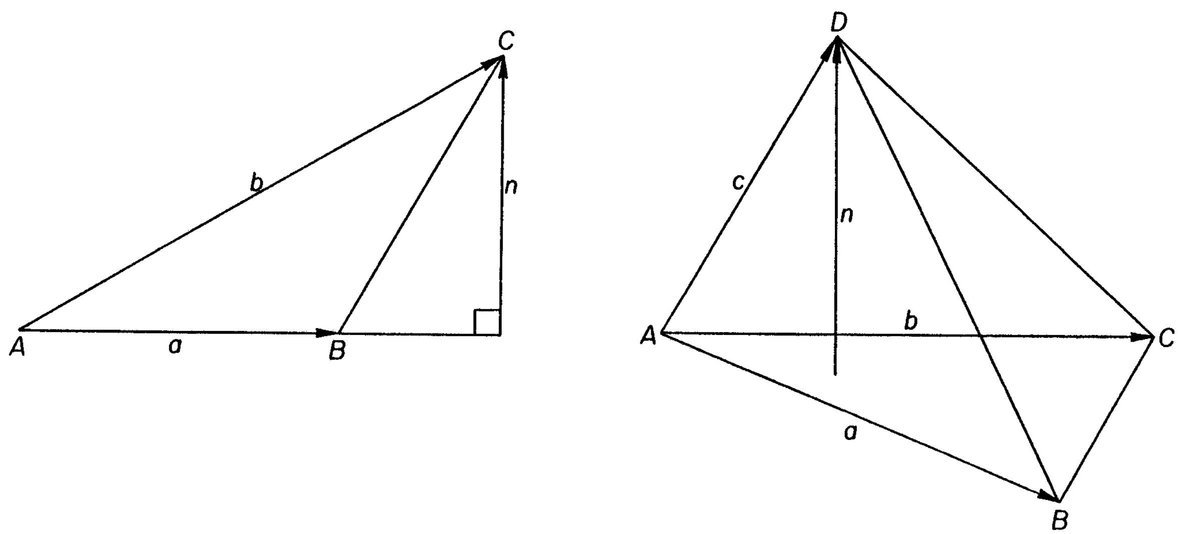
but since $\cos 2\alpha \neq 1$, we are left with $\|b\|(1-\lambda) = \lambda\|a\|$, and the proof is done.

Theorem 4.18. *Let the vertices of triangle ABC be at $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$, and write $a = \vec{AB}$, $b = \vec{AC}$. Let the vertices of tetrahedron $ABCD$ be at $A(x_1, y_1, z_1)$ $B(x_2, y_2, z_2)$, $C(x_3, y_3, z_3)$, $D(x_4, y_4, z_4)$, and write $a = \vec{AB}, b = \vec{AC}, c = \vec{AD}$. Then*

$$\text{Area } ABC = \frac{1}{2}\begin{vmatrix} a^Ta & a^Tb \\ b^Ta & b^Tb \end{vmatrix}^{1/2} = \frac{1}{2}\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \tag{4.69}$$

$$\text{Volume } ABCD = \frac{1}{6}\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \frac{1}{6}\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{vmatrix}. \tag{4.70}$$

Proof. According to Fig. 23a



(a) Fig. 4.23 (b)

$$\text{Area } ABC = \frac{1}{2}\sqrt{(n^Tn)(a^Ta)}. \tag{4.71}$$

Inasmuch as n is orthogonal to a and is coplanar with a and b

$$n = -\frac{a^Tb}{a^Ta}a + b \tag{4.72}$$

and

$$\text{Area } ABC = \frac{1}{2}\Big((a^Ta)(b^Tb) - (a^Tb)^2\Big)^{1/2} = \frac{1}{2}\begin{vmatrix} a^Ta & a^Tb \\ b^Ta & b^Tb \end{vmatrix}^{1/2} \tag{4.73}$$

which is true, we notice, for any triangle in space. In terms of the components of spatial a and b

$$\text{Area } ABC = \frac{1}{2}(\Delta_1^2 + \Delta_2^2 + \Delta_3^2)^{1/2} \tag{4.74}$$

where

$$\Delta_1 = \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}, \ \Delta_2 = \begin{vmatrix} a_3 & b_3 \\ a_1 & b_1 \end{vmatrix}, \ \Delta_3 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}. \tag{4.75}$$

In case triangle ABC lies in a plane parallel to the xy plane $a_3 = b_3 = 0$, and

$$\text{Area } ABC = \frac{1}{2}\sqrt{(a_1b_2 - a_2b_1)^2} = \frac{1}{2}\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \tag{4.76}$$

or by vertex coordinates

$$\text{Area } ABC = \frac{1}{2}\Big((x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)\Big) = \frac{1}{2}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}. \tag{4.77}$$

The sign ambiguity raised by the square root will be resolved after the proof is completed.

For tetrahedron $ABCD$ of Fig. 23b we have that

$$\text{Volume } ABCD = \frac{1}{3}\|n\| \text{ Area } ABC. \tag{4.78}$$

Vector n is solved, using Cramer's rule, from the three equations $a^Tn = 0,\ b^Tn = 0,\ c^Tn = n^Tn$,

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = n^Tn\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.79}$$

as

$$n = \frac{n^Tn}{\Delta}[\Delta_1\ \Delta_2\ \Delta_3]^T \tag{4.80}$$

where $\Delta_1, \Delta_2, \Delta_3$ are given in eq. (4.75), and where

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}. \tag{4.81}$$

Writing $n^T n$ for n in eq. (4.80) we obtain

$$\Delta = (n^T n)^{1/2} (\Delta_1^2 + \Delta_2^2 + \Delta_3^2)^{1/2}.$$

We recognize that $(\Delta_1^2 + \Delta_2^2 + \Delta_3^2)^{1/2} = 2$ Area ABC, and conclude that volume $ABCD = \Delta/6$. Putting the components of a, b, c as differences of vertex coordinates we obtain the second determinant for the volume. End of proof.

To resolve the sign ambiguity of the triangle area formula we write

$$a = \|a\| \, [\cos\alpha \ \ \sin\alpha]^T, \ \ b = \|b\| \, [\cos\beta \ \ \sin\beta]^T \tag{4.83}$$

and have that

$$\text{Area } ABC = \frac{1}{2} \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = \frac{1}{2} \|a\| \, \|b\| \sin(\beta - \alpha) \tag{4.84}$$

If the vertices of triangle ABC are listed sequentially *counterclockwise*, then $\beta - \alpha > 0$ and the area of the triangle is positive. The triangle, we say, has positive *orientation*. Permutation of two vertices reverses the sign of Area ABC, and hence its orientation. Planar rotation and translation does not affect the orientation of a triangle; only a flip-over changes it.

A triangle with vertices at $A(0,0), B(1,0), C(0,1)$ has positive orientation, and so does a tetrahedron with vertices at $A(0,0,0), B(1,0,0), C(0,1,0), D(0,0,1)$.

Let's muse more and delve deeper into the nature of the affinity between algebra and geometry. Once we have decided that an ordered pair of numbers makes a point, and once we have formulated the distance between two points—that is, once we have fixed the *metric* of the space, algebra becomes fully equipped to travel its own way through geometry. Analytic geometry is then completely determined by the nature of real numbers with their inherent properties of order and continuity. Nothing more is needed, except the analytic formulation of the geometric problem. We need construct noting say nothing, nor imagine anything. A point is a pair of numbers, that's all, not a dent not a peck not a peg, and not a speck of light or matter. the mysteriously primitive object called line in geometry, supposedly being "determined" by two points, becomes a simple functional relationship between two variables with two parameters that are exactly fixed by two distinct points. The geometer who is constantly torn between the existential and the philosophical, between the real and the ideal, must be forever on guard to separate the obvious from the assumed. For the geometer, a coherent geometrical proof to the simplest and most obvious of theorems can turn into a feat of reasoning. Algebra relieves all that. Yet algebra is not geometry. One cannot divide plots nor build houses with algebra. To return to the physical world we shall need to know where and how to pick a point, how to construct a line, and how to compare distances.

Exercises

4.5.1. In trapezoid $ABCD$, P and Q are midpoints (Fig. 24.)

Prove that $PQ \| AB$, and $PQ = \frac{1}{2}(AB + CD)$.

4.5.2. In quadrilateral $ABCD$, P, Q, R, S are midpoints, and so are K and L. Prove that K, O, L are colinear (Fig. 25.)

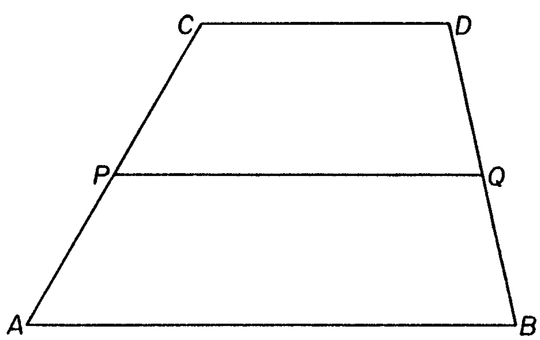


Fig.4.24

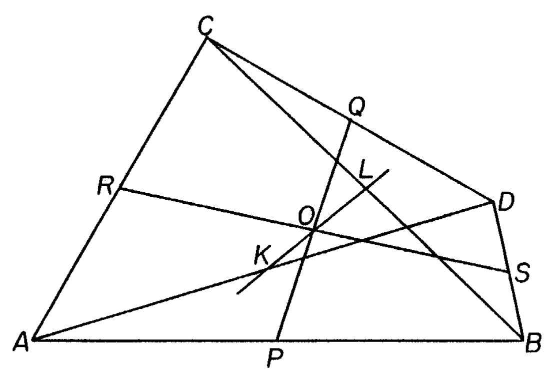


Fig.4.25

4.5.3. Prove that the diagonals of a parallelogram bisect each other.

4.5.4. Prove that if P is the midpoint of AA', Q the midpoint of BB', and R the midpoint of CC', then P, Q, R are colinear.

Also, that

$$(1 + \mu)m = b' + \mu b, \ (1 - \mu)m' = b' - \mu b$$

or

$$\vec{B'M} = \mu \vec{MB}, \quad \vec{B'M'} = -\mu \vec{M'B}.$$

Such division is called *harmonic*. The two diagonals CC' and AA' divide the third diagonal BB' harmonically (Fig. 26.)

5.5.5. Prove that if

$$\vec{CA} = -\lambda \vec{CB},\ \vec{C'A'} = -\lambda' \vec{C'B'}$$

and

$$\vec{DA} = -\mu \vec{DB},\ \vec{D'A'} = -\mu' \vec{D'B'}$$

then $\lambda/\lambda' = \mu/\mu'$. This is the celebrated cross-ratio theorem (Fig. 27.)

4.5.6. Prove that a parallelogram is a rhomb if and only if its diagonals are orthogonal.

4.5.7. Prove that a parallelogram is a rectangle if and only if its diagonals are equal.

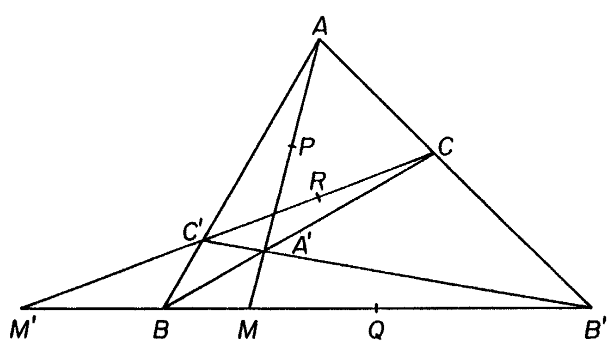


Fig.4.26

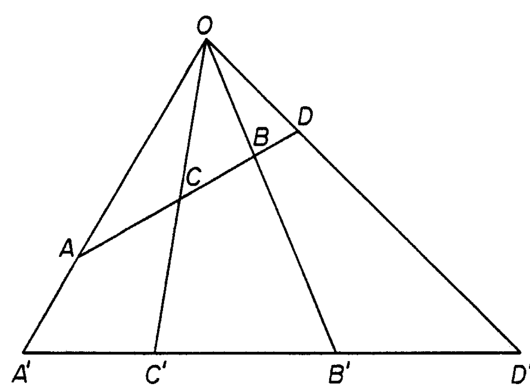


Fig.4.27

4.5.8. Let $a = [1\ 1]^T, b = [1\ -2]^T$. Find $x = [x_1\ x_2]^T$ so that $a^T x = 2,\ b^T x = -7$. Numbers 2 and -7 are called by some the *covariant* components of x. Also, for a, b, x find α and β so that $x = \alpha a + \beta b$. Numbers α, β are called by some the *contravariant* components of x.

With reference to Fig.28 prove that if a and b are unit vectors and $a^T b = 0$, then the covariant and contravariant components of x are equal.

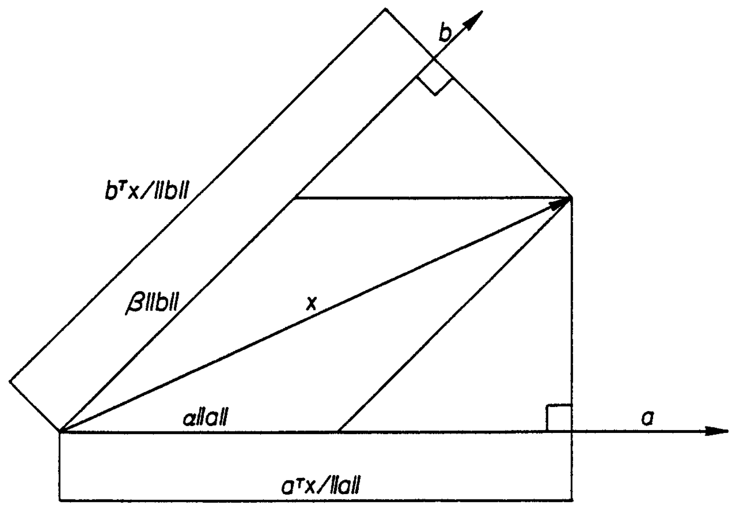


Fig.4.28

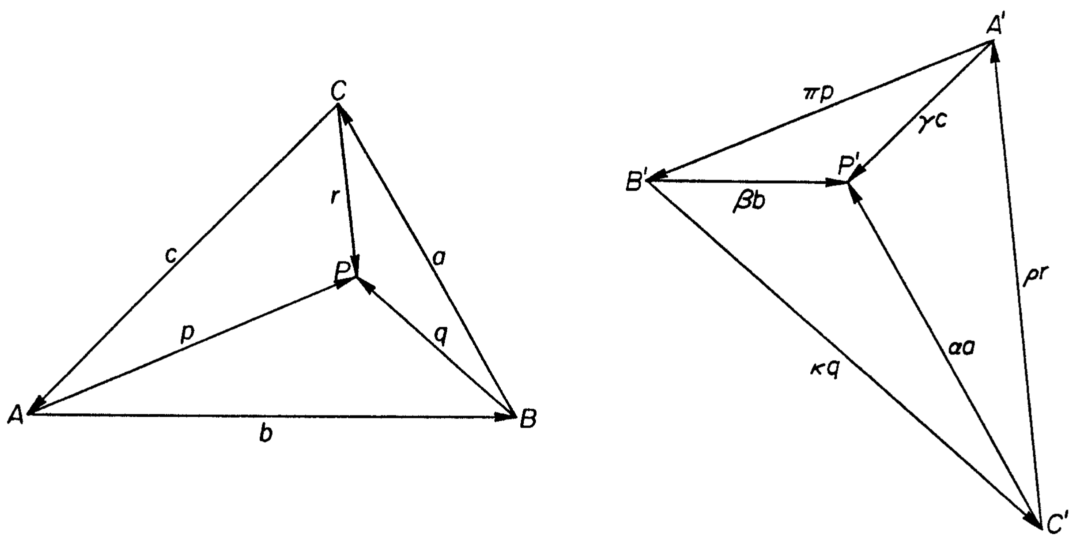


Fig.4.29

4.5.9. Point P is in the plane of ABC. If $A'B' \| AP, B'C' \| BP, C'A' \| CP$, show that the lines through A', B', C' parallel to AC, BA, CB respectively, are concurrent (Fig. 29.)

4.5.10. Show that if in quadrilateral $ABCD$, $AB \perp CD$, then $(AC)^2 + (BD)^2 = (AD)^2 + (BC)^2$, and vice versa (Fig. 30.)

4.5.11. Show that if coplanar a, b, x are such that $\|a\| = \|b\| = \|x\|$ and $a^T b = 0$, then $x = a \cos\alpha + b \cos\beta$. Also, that if $\|x\|$ is arbitrary, then $x = (\|x\|/\|a\|)(a \cos\alpha + b \cos\beta)$ (Fig. 31.)

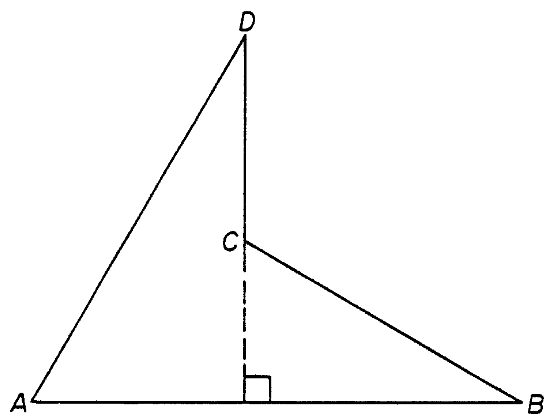


Fig.4.30

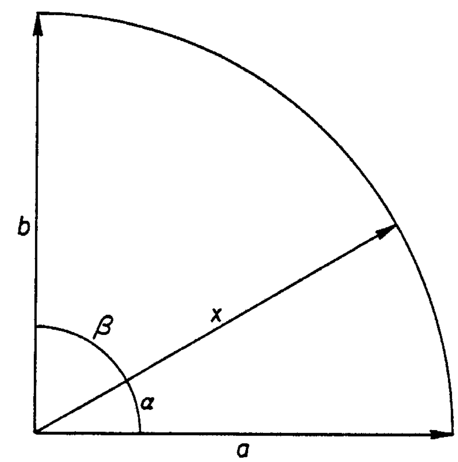


Fig.4.31

4.5.12. Show that if coplanar a, b, x are such that $\|a\| = \|b\| = \|x\|$, $a^T x = \|a\|^2 \cos\alpha$, $b^T x = \|a\|^2 \cos\beta$ and $a^T b = \|a\|^2 \cos(\alpha + \beta)$, then

$$x = \frac{\sin\beta}{\sin(\alpha+\beta)} a + \frac{\sin\alpha}{\sin(\alpha+\beta)} b.$$

4.5.13. Show that if x in the plane of a and b bisects the angle between them, then $x = \alpha a + \beta b$ with

$$\alpha = \frac{\|x\|}{\|a\|} \frac{1}{2\cos\theta} \ , \ \beta = \frac{\|x\|}{\|b\|} \frac{1}{2\cos\theta}$$

(Fig. 32.)

4.5.14. Prove that if $OA = OB, OA \perp OB$, and $OC = OD$, $OC \perp OD$, then $AD \perp BC$ (Fig. 33.)

4.5.15. Prove that if A, M, B are collinear points so that

$$m = \frac{\|q\|}{\|r\|} a + \frac{\|p\|}{\|r\|} b \ , \ r = p + q$$

then

$$\|r\| \ \|m\|^2 = \|q\| \ \|a\|^2 + \|p\| \ \|b\|^2 - \|p\| \ \|q\| \ \|r\|.$$

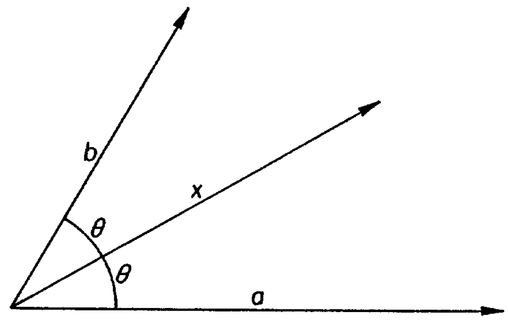


Fig.4.32

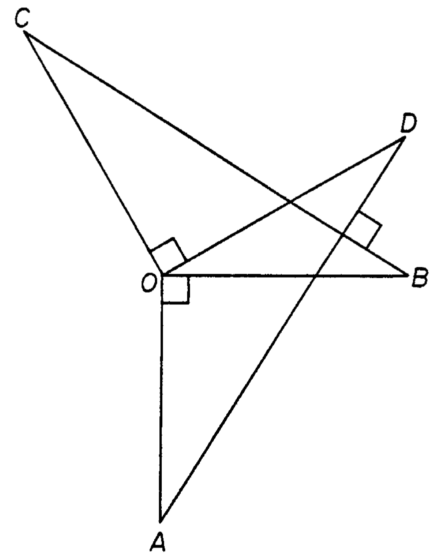


Fig.4.33

This is Stewart's formula (Fig. 34.)

4.5.16. Show that

$$\|x\|^2 = \|c\|^2(1-\lambda) + \|b\|^2\lambda - \|a\|^2\lambda(1-\lambda)$$

(Fig. 35.)

4.5.17. Let

$$a = \alpha_1 \vec{OA_1} + \alpha_2 \vec{OA_2} + \cdot + \alpha_n \vec{OA_n}.$$

Show that if $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 0$, then vector a is independent of point O.

4.5.18. Points $A_1, A_2, \ldots, A_n$ fix point G according to

$$g = \alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_n a_n,\ \ \alpha_1 + \alpha_2 + \cdots + \alpha_n = 1$$

where $g = \vec{OG}$ and $a_i = \vec{OA_i}$. Prove that G is independent of O.

4.5.19. Show that if $a^T b = o$, then

$$\|p\|^2 + \|q\|^2 = 5\|c\|^2$$

(Fig. 36.)

4.5.20. If point I is the center of the inscribed circle in triangle ABC, show that in terms of the position vectors,

$$(\alpha + \beta + \gamma)i = \alpha a + \beta b + \gamma c$$

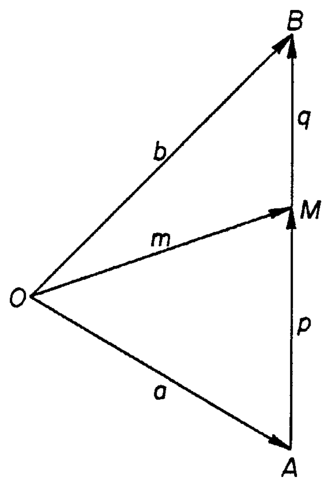


Fig.4.34

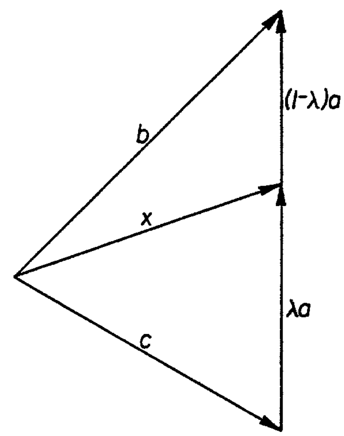


Fig.4.35

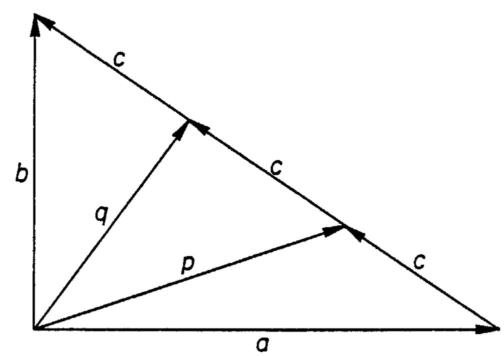


Fig.4.36

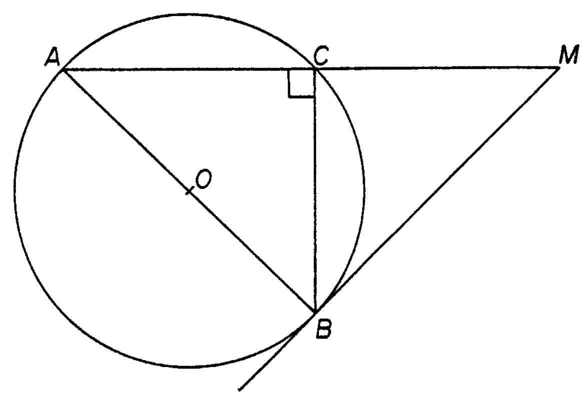


Fig.4.37

where $\alpha = BC,\ \beta = CA,\ \gamma = AB$.

4.5.21 Let $AM \perp CB,\ MB \perp BA$, and write $\vec{MC} = c,\ \vec{MA} = (1+\alpha)c,\ \vec{MB} = t,\ \vec{OM} = m,\ \vec{OA} = a$. Prove that

$$(1+\alpha)c^T c = t^T t = m^T m - a^T a$$

(Fig. 37.)

4.5.22. If in triangle ABC, P, Q, R are midpoints, show that

$$\text{Area}PQR/\text{Area } ABC = 1/4.$$

Generally, if $g = (\alpha a + \beta b + \gamma c)/(\alpha + \beta + \gamma)$, then

$$\text{Area } PQR/\text{Area } ABC = 2\alpha\beta\gamma/((\alpha+\beta)(\beta+\gamma)(\gamma+\alpha))$$

(Fig. 38.)

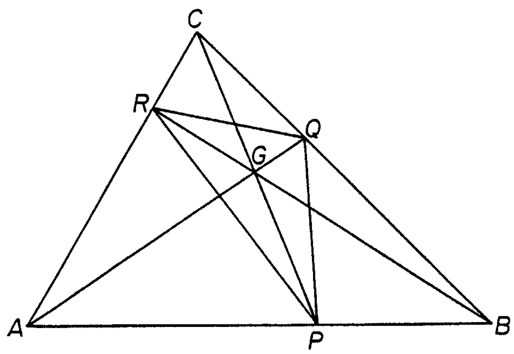


Fig.4.38

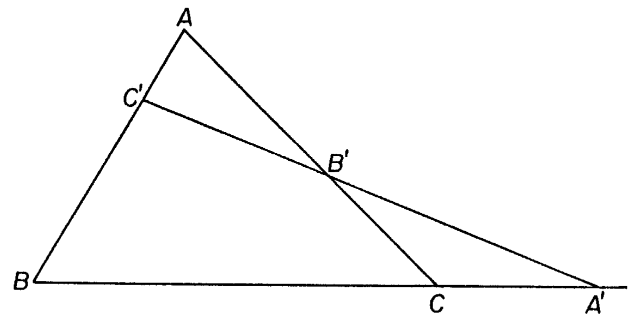


Fig.4.39

4.5.23. Following Menelaus prove that a necessary and sufficient condition for points A', B', C' to be colinear is that

$$\alpha\beta\gamma = 1,\ \alpha = \frac{A'B}{A'C},\ \beta = \frac{B'C}{B'A},\ \gamma = \frac{C'A}{C'B}$$

(Fig. 39.)

4.5.24. Prove that if ρ is the radius of the circumscribing circle to triangle ABC, then

$$\text{Area ABC } = \frac{1}{4}\frac{\alpha\beta\gamma}{\rho},\ \alpha = AB,\ \beta = BC,\ \gamma = CA.$$

4.5.25. Let the barycenter (center of mass) of points $A_1, A_2, \ldots, A_n$ and points B_1,B_2,. . .,B_m be G_1 and G_2 respectively,

$$\begin{array}{l} g_1 = \alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_n a_n,\ \alpha_1 + \alpha_2 + \cdots + \alpha_n = 1 \\ g_2 = \beta_1 b_1 + \beta_2 b_2 + \cdots + \beta_n b_n,\ \beta_1 + \beta_2 + \cdots + \beta_n = 1 \end{array}.$$

Show that barycenter G of all points is the barycenter of the barycenters,

$$g = \gamma_1 g_1 + \gamma_2 g_2,\ \gamma_1 + \gamma_2 = 1.$$

4.5.26. Triangle ABC has vertices at $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$. Show that Cartesian coordinates x, y are obtained from the barycentric coordinates $\zeta_1, \zeta_2, \zeta_3, \zeta_1 + \zeta_1 + \zeta_3 = 1$ as

$$x = x_1\zeta_1 + x_2\zeta_2 + x_3\zeta_3$$

$$y = y_1\zeta_1 + y_2\zeta_2 + y_3\zeta_3.$$

4.5.27. Show that in terms of the barycentric coordinates $\zeta_1, \zeta_2, \zeta_3$ a complete quadratic is written as $\zeta_1^k\zeta_2^l\zeta_3^m \;\; k + l + m = 2$ with k, l, m being positive integers.

4.5.28. Using barycentric coordinates write the equation of the line parallel to AB through C. Also the line orthogonal to AB through A.

4.5.29. Write the area of triangle $P'Q'R'$ in terms of the area of triangle ABC, given that

$$r = (\alpha_1 a + \beta_1 b)/(\alpha_1 + \beta_1)$$

$$p = (\beta_2 b + \gamma_1 c)/(\beta_2 + \gamma_1)$$

$$q = (\gamma_2 c + \alpha_2 a)/(\gamma_2 + \alpha_2).$$

(Fig. 40.)

4.5.30. Equation 4.84 states that if a_1 and a_2 are in R^2, then $det[a_1 \; a_2]$ is the area of the parallelogram formed by a_1, a_2 on its sides. Discuss using area to define $det[a_1 \; a_2]$ for rectangular matrix A with columns of more than two components. If A is rectangular and B square, is it still true that $det(AB) = det(A)det(B)$? Generalize the argument to $det[a_1 \; a_2 \; a_3]$, etc.

4.5.31. Write the lengths of the angle bisectors in a triangle in terms of the three sides.

4.5.32. Prove that if the bisectors of two internal angles of a triangle are equal, then the triangle is isosceles. This, not so easy problem, is the Steiner-Lehmus theorem.

4.5.33. Discuss Desargue's theorem (4.15) for the case that points A, A' and B, B' coincide. Also for the case that points C, C' coincide and $AB \| A'B'$.

4.5.34. Prove the converse to Theorem 4.15.

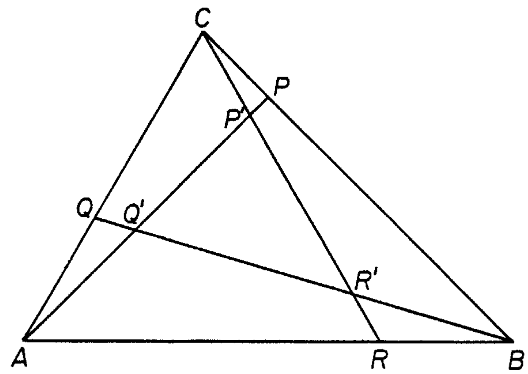


Fig.4.40

4.6 Linear transformations

In Chapter One we looked with deliberate care at linear system $Ax = f$ with f given. Here we consider the linear relationship $Ax = x'$ between independent *variable* vector x and dependent *variable* vector x'. When x varies over a certain vector space, over what vector space does x' range? This is essentially the question before us now.

Recall that a one-dimensional vector space holds all vectors $x = \alpha a$ for given $a \neq o$ and arbitrary α, that is, all vectors colinear with the line generated by a. A two-dimensional vector space includes the collection of all vectors $x = \alpha a + \beta b$ for noncolinear a, b and arbitrary α, β; that is, the set of all vectors parallel to the plane generated by a, b. The three-dimensional vector space R^3 includes all vectors with three real components, that is, all vectors $x = \alpha_1[1\ 0\ 0]^T + \alpha_2[0\ 1\ 0]^T + \alpha_3[0\ 0\ 1]^T$ for arbitrary $\alpha_1, \alpha_2, \alpha_3$, or all vectors $x = \alpha a + \beta b + \gamma c$ for noncoplanar a, b, c and arbitrary α, β, γ.

Suggestive terminology calls $x' = Ax$ the *linear transformation* of the corresponding vector spaces of x and x', or the *linear mapping* of x into its *image* x'. One may also want to think of A in $x' = Ax$ as being a *linear operator* that acts upon x to produce x'.

When A is nonsingular the transformation is said to be nonsingular. Nonsingular transformations are special to a degree that they alone are sometimes found worthy of the name transformation.

Suppose that

$$x = [x_1 \; x_2]^T = x_1[1 \; 0]^T + x_2[0 \; 1]^T = x_1e_1 + x_2e_2 \tag{4.85}$$

for arbitrary x_1, x_2. This vector space, R^2, spanned by e_1, e_2 is the *domain* of the mapping. Vector space

$$x' = Ax = x_1Ae_1 + x_2Ae_2 = x_1a_1 + x_2a_2 \tag{4.86}$$

spanned by columns a_1, a_2 of A is the *range* of the transformation, and if $A = A(2 \times 2)$ is nonsingular, then the range of the mapping is also R^2. Every vector $x \in R^2$ has a unique image $x' \in R^2$, and vice versa. The nonsingular mapping is one-to-one, or in the emphatic terminology of linear transformations the mapping is *onto*; R^2 is mapped onto R^2 and we write $R^2 \rightarrow R^2$.

Similarly R^3 is mapped by nonsingular $A = A(3 \times 3)$ onto R^3; if $x' = Ax$, then also $x = A^{-1}x'$.

Since the linear transformation $x' = Ax$ is given entirely by matrix A, we may interchangeably speak about the vector spaces associated with the transformation or the vector spaces associated with A itself.

Consider, for instance, the linear operator (alias matrix)

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{4.87}$$

in the linear mapping $x' = Ax$ and see Fig. 41. With no restrictions imposed on x the domain of this mapping is R^2; x is any

vector with two real components, but since the columns of A are colinear it is singular and the range of the mapping is only a one-dimensional vector space R spanned by $r = [-1 \; 1]^T$. Every image vector x', whatever x, is colinear with vector r. All vectors x such that $Ax = o$ constitute the *nullspace* of matrix A. Nullspace N of A in eq. (4.87) is a one-dimensional vector space spanned by $n = [1 \; 1]^T$. Every vector of the form $x = \alpha[1 \; 1]^T$, and only vectors of this form, have as image the zero vector $o = [0 \; 0]^T$.

Vectors that are already in the range of this mapping, i.e. $x = \alpha[-1 \; 1]^T$ return image vector x' to the range. Here, for the matrix in eq.(4.87),

$$x' = Ax = 2x \tag{4.88}$$

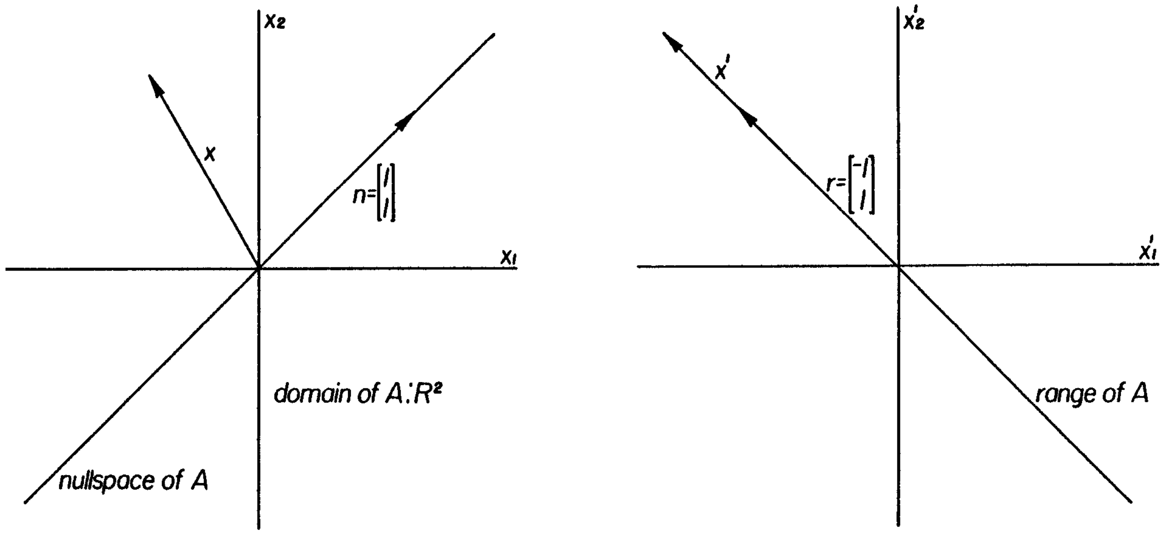


Fig. 4.41

and x and Ax are colinear. A nonzero vector x that is colinear with $Ax,\ Ax = \lambda x$, is said to be an *eigenvector* of square matrix A with *eigenvalue* λ. Vector $x = [1\ 1]^T$ that spans the nullspace of A is also an eigenvector of A since for this vector $Ax = 0x$, and the eigenvalue is zero.

Matrix A in linear mapping $x' = Ax$ need not be square. For example

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad , \ x' = Ax \tag{4.89}$$

sends $x \in R^2$ into the two-dimensional subspace, a plane, of R^3 spanned by the noncolinear columns $a_1 = [1\ -1\ 1]^T,\ a_2 = [1\ 1\ 1]^T$ of A.

The nullspace of mapping (4.89) includes only one element, the zero vector, since a_1 and a_2 are not colinear. A mapping has a nontrivial, or nonempty, nullspace only if matrix A that affects it has colinear or coplanar columns.

Generally, let the domain of $x' = Ax,\ A = [a_1\ a_2\ a_3]$, be R^3 so that $x' = x_1a_1 + x_2a_2 + x_3a_3$ for arbitrary x_1, x_2, x_3. If the columns of A are colinear and $a_1 \neq o$, then we may write $a_2 = \alpha_2 a_1,\ a_3 = \alpha_3 a_1$ and the image of x becomes $x' = (x_1 + \alpha_2 x_2 + \alpha_3 x_3)a_1$, signifying that the range of A is a one-dimensional subspace of R^3 spanned by $a_1 \neq o$. If the columns of A are coplanar such that $a_1 = \alpha_2 a_2 + \alpha_3 a_3$, with a_2, a_3 being noncolinear, then

$$x' = (x_1\alpha_2 + x_2)a_2 + (x_1\alpha_3 + x_3)a_3 = \alpha'_2 a_2 + \alpha'_3 a_3 \tag{4.90}$$

signifying, since α_2' and α_3' are arbitrary, that the range of A is a two-dimensional subspace of R^3 spanned by the noncolinear a_1, a_2. The range of A is the vector space spanned by its columns; it is the *column space* of A. The dimension of the range is the *rank* of A.

If in the mapping $x' = Ax$ of $x \in R^3$ the columns of A are coplanar, then the range of A is a subspace R of R^3 of dimension less than three. Say x' is on a plane. Vector $x \in R$ already on the plane creates an image x' confined to a subspace R' of R, $R' \subset R \subset R^3$. For example, if $x \in R^3$, then the range of

$$A = [a_1\ a_2\ a_3] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -3 \\ 1 & 1 & 1 \end{bmatrix} \tag{4.91}$$

constitutes the totality of vectors $x' = \alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3$ for arbitrary $\alpha_1, \alpha_2, \alpha_3$. It turns out that a_1, a_2, a_3 are coplanar so that $a_3 = -a_1 + 2a_2$ and the range of A is two-dimensional, $x' = \alpha_1' a_1 + \alpha_2' a_2$, where $\alpha_1' = \alpha_1 - \alpha_3,\ \alpha_2' = \alpha_2 + 2\alpha_3$ are arbitrary. Taking $x = \alpha_1 a_1 + \alpha_2 a_2$ in R we obtain the image $x' = \alpha_1 A a_1 + \alpha_2 A a_2 = (3\alpha_1 + \alpha_2)[1\ \ -1\ 1]^T$, and R' is a one-dimensional vector space spanned by $r' = [1\ \ -1\ 1]^T$. Now, if $x \in R'$, that is, if $x = \alpha[1\ \ -1\ 1]^T$, then $Ax = x$, and x is an eigenvector of A with eigenvalue equal to 1. Solving $Ax = o$ we discover that the nullspace of A is the one-dimensional vector space that includes the set of all vectors $x = \alpha[1\ \ -2\ 1]^T$. The nullspace of A happens here to be in its range,$-a_1 + 3a_2 = 2[1\ \ -2\ 1]^T$.

One more example. Matrix

$$A = [a_1\ a_2\ a_3] = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \tag{4.92}$$

of linear mapping $x' = Ax,\ x \in R^3$, has coplanar columns so that $a_1 = a_2 + a_3$. We verify that the nullspace of A is the one-dimensional vector space spanned by $n = [-1\ 1\ 1]^T$. The range $x' = \alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3$ of the mapping reduces by virtue of the coplanarity of a_1, a_2, a_3 to $x' = \alpha_2' a_2 + \alpha_3' a_3$ where $\alpha_2' = \alpha_1 + \alpha_2,\ \alpha_3' = \alpha_1 + \alpha_3$ are arbitrary. The totality of image vectors x' constitutes a two-dimensional subspace R, spanned by a_2, a_3 of R^3. Restricting x to R, $x = \alpha_2 a_2 + \alpha_3 a_3$, we obtain

$$x' = Ax = \alpha_2 \begin{bmatrix} -1 \\ 1 \\ -3 \end{bmatrix} + \alpha_3 \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} = \alpha_2 a_2' + \alpha_3 a_3'. \tag{4.93}$$

Vectors a_2' and a_3' are not colinear and are in R; they span R. In fact, $a_2' = a_2 - a_3$, and $a_3' = a_2 + 3a_3$ so that

$$x' = (\alpha_2 + \alpha_3)a_2 + (-\alpha_2 + 3\alpha_3)a_3 = \alpha_2' a_2 + \alpha_3' a_3. \tag{4.94}$$

Since

$$\begin{array}{r} \alpha_2' = \alpha_2 + \alpha_3 \\ \alpha_3' = -\alpha_2 + 3\alpha_3 \end{array} \quad \text{and conversely} \quad \begin{array}{l} \alpha_2 = \frac{3}{4}\alpha_2' - \frac{1}{4}\alpha_3' \\ \alpha_3 = \frac{1}{4}\alpha_2' + \frac{1}{4}\alpha_3' \end{array} \tag{4.95}$$

every vector $x \in R$ has a unique image $x' \in R$. Vector space R is an *invariant subspace* of A. It contains an interesting subspace R' spanned by $a = [1\ 1\ 1]^T$, given by $\alpha_2 = \alpha_3 = 1/2$, such that if $x = \alpha a$, then $Ax = 2x$. Vector x is then an eigenvector of A with eigenvalue 2.

Theorem 4.19. *If range R of mapping $x' = Ax$, $x \in R^3$, is a subspace of R^3 of dimension $dim(R)$, and nullspace N of the mapping is a subspace of R^3 of dimension $dim(N)$, then $dim(R) + dim(N) = 3$.*

Proof. Let the columns of A be a_1, a_2, a_3, and assume the mapping nontrivial so that at least one of the columns, say a_1 is nonzero.

If the range is one-dimensional, ($dim(R) = 1$), then a_2 and a_3 are parallel to a_1, and $a_2 = \alpha_2 a_1$ and $a_3 = \alpha_3 a_1$. The nullspace totals the vectors $n = [n_1\ n_2\ n_3]^T$ such that $n_1 a_1 + n_2 a_2 + n_3 a_3 = o$. Because a_1, a_2, a_3 are colinear, two components of n, here n_2 and n_3, are independent and n_1 depends on them. Vectors n are on one plane. Indeed, under the condition that $n_1 a_1 + n_2 a_2 + n_3 a_3 = o$ and with $a_2 = \alpha_2 a_1$ and $a_3 = \alpha_3 a_1$ it results that

$$(n_1 + n_2\alpha_2 + n_3\alpha_3)a_1 = o. \tag{4.96}$$

Since $a_1 \neq o$, $n_1 = -n_2\alpha_2 - n_3\alpha_3$ and

$$n = n_2 \begin{bmatrix} -\alpha_2 \\ 1 \\ 0 \end{bmatrix} + n_3 \begin{bmatrix} -\alpha_3 \\ 0 \\ 1 \end{bmatrix} \tag{4.97}$$

showing n to be in the plane generated by the noncoplanar $[-\alpha_2\ 1\ 0]^T$ and $[-\alpha_3\ 0\ 1]^T$.

In case the range is two-dimensional, the columns of A are coplanar with at least two of them, say a_1 and a_2, being noncolinear so that the third column may be expressed as

$$a_3 = \alpha_1 a_1 + \alpha_2 a_2. \tag{4.98}$$

Substituting this into $n_1a_1 + n_2a_2 + n_3a_3 = o$ yields

$$(n_1 + \alpha_1 n_3)a_1 + (n_2 + \alpha_2 n_3)a_2 = o. \tag{4.99}$$

Because a_1, a_2 are noncolinear $n_1 = -\alpha_1 n_3$, $n_2 = -\alpha_2 n_3$, and $n = n_3[-\alpha_1 \ \ -\alpha_2 \ \ 1]^T$ is in a one-dimensional vector space. End of proof.

Range R and nullspace N of A in eq. (4.87) are orthogonal subspaces complementing R^2. This is always the case if A is *symmetric.*

Theorem 4.20. *If matrix $A = A(3 \times 3)$ is symmetric, $A = A^T$, then the range and nullspace of the linear transformation $x' = Ax$ are orthogonal complements of R^3. In case the range is one-dimensional, orthogonality happens only if $A = A^T$.*

Proof. Suppose first a one-dimensional range so that the three columns of A are colinear. If A is symmetric, $A = A^T$, then $An = o$ means that $a_1^T n = 0,\ a_2^T n = 0,\ a_3^T n = 0$, and n is in the plane orthogonal to the range.

To prove that if the range of A is one-dimensional then the nullspace of A is orthogonal to it only if $A = A^T$, assume $A - A^T \neq O$ and consider the range of A spanned by vector r and the nullspace of A spanned by vectors n_1, n_2 taken so that $n_1^T n_2 = 0$ and $n_1^T r = n_2^T r = 0$. Vectors n_1, n_2, r span R^3. Writing

$$A^T n_1 = \nu_1 n_1 + \nu_2 n_2 + \rho r \tag{4.100}$$

and using the fact that ${n_1}^T A^T n_1 = {n_1}^T A n_1 = n_2^T A n_1 = n_1^T A n_2 = 0$ we readily conclude that $A^T n_1 = o$, and in the same way that $A^T n_2 = o$, and also that $A^T r = Ar$, or

$$(A - A^T)n_1 = (A - A^T)n_2 = (A - A^T)r = o. \tag{4.101}$$

But for a 3×3 matrix this means that $A - A^T = O$, since it implies that for any $x \in R^3$, $\ (A - A^T)x = o$. Our assumption that $A - A^T \neq O$ is contradicted and the plane orthogonal to r is not the nullspace of A.

If the range of A is two-dimensional, then a_1, a_2, a_3 generate a plane in R^3, and $a_1^T n = a_2^T n = a_3^T n = 0$ means that n is orthogonal to that plane. End of proof.

The dimension of the range is the *rank* of A and the dimension of the nullspace is the *nullity* of A.

As an example consider

$$A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 4 & -5 \\ 2 & -1 & 3 \end{bmatrix} \tag{4.102}$$

that has a two-dimensional range spanned by r_1, r_2, and a one-dimensional nullspace spanned by n,

$$r_1 = [1\ 1\ 0]^T,\ r_2 = [1\ 0\ 1]^T,\ n = [-1\ 1\ 1]^T. \tag{4.103}$$

Matrix A of eq. (4.102) is not symmetric but its nullspace is orthogonal to its range, $r_1^T n = r_2^T n = 0$. Matrix A is of rank 2 and nullity 1. According to Theorem 4.17 *rank* $A+$ *nullity* $A = 3$.

Theorem 4.21. *The range of A and the nullspace of A^T are orthogonal.*

Proof. The range of A is spanned by a_1, a_2, a_3, while nullspace N of A^T consists of all vectors n so that $a_1^T n = a_2^T n = a_3^T n = 0$. Hence every $n \in N$ is orthogonal to the range or the column space of A. End of proof.

Exercises

4.6.1. Find the range and nullspace of

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & -1 \\ -1 & 1 & 2 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ -\beta & -\gamma & 0. \end{bmatrix}$$

What are the rank and nullity of A and T?

4.6.2. The range and nullspace of matrix

$$A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

are the same vector space spanned by $a = [1\ \ -1]^T$. Write all such (2×2) matrices. Can A ever be symmetric? Do the same for (3×3) matrices.

4.6.3. In the transformation $x' = Ax$,

$$A = \begin{bmatrix} -4 & 1 \\ -1 & 1 \end{bmatrix},$$

find all x so that $x^T x' = 0$.

4.6.4. Does a linear mapping $x' = Ax$ of R^2 onto itself exist by which the orthonormal x_1, x_2 are mapped such that $Ax_1 = \lambda_1 x_2,\ Ax_2 = \lambda_2 x_1,\ \lambda_1 \neq \lambda_2$? Hint: $x_1^T A x_1 = x_2^T A x_2 = 0$.

4.7 Projection into subspaces

A notable conclusion of the previous section has been that if A in linear mapping $x' = Ax$, is singular, then $x \in R^3$ is relegated by the mapping to a lower-dimensional vector space. In this section we shall consider a geometrically significant and linear algebraically important singular linear mapping having the property that *every* vector x in the range R of A is an eigenvector of A with *unit eigenvalue*

$$Ax = x \quad x \in R. \tag{4.104}$$

Singular matrix A having this property is said to be a *projection matrix*, and the mapping a projection of x into the range.

Matrix

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{4.105}$$

is not a projection matrix since

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{4.106}$$

and every vector $x = \alpha[1 \ \ -1]^T$ in the range of A is extended by the mapping to *twice* its length.

The essential property of projection embodied in eq. (4.104) imports that for any $x \in R^3$, $A^2 x = Ax$ and the projection matrix is characterized by $A^2 = A$, $A \neq I$. One and zero are the only eigenvalues of a projection matrix.

A common notation for the projection matrix is P, and if in addition to $P^2 = P$ also $P = P^T$, then the projection is orthogonal —the nullspace of P is orthogonal to its range.

For a geometric interpretation of the projection of R^2 into (not onto!) a line see Fig. 42. Projection Px of $x \in R^2$ on line λ with unit vector v along it is geometrically imagined as the shadow that x casts upon λ by rays of light parallel to unit vector u. When x is parallel

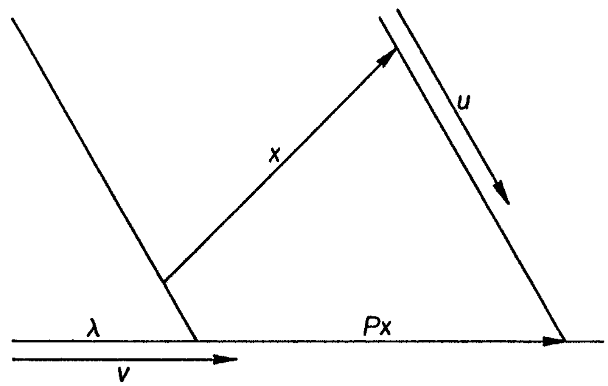


Fig. 4.42

to the line, $Px = x$, but when x is parallel to u, $Px = o$; vector u spans the nullspace of the projection. We write with reference to Fig. 42

$$P = \frac{1}{1-\alpha^2}(vv^T - \alpha vu^T),\ \alpha = u^T v \tag{4.107}$$

and verify that $Pv = v$ and that $Pu = o$. When $u^T v = 0$

$$P = vv^T,\ v^T v = 1 \tag{4.108}$$

and the projection is *orthogonal*. It is an easy matter to show that in both eqs. (4.107) and (4.108) $P = P^2$.

If in projection matrix P of eq. (4.108) $v \in R^3$, then the nullspace of this projection consists of the plane orthogonal to v, and every $x \in R^3$ is orthogonally projected on line λ.

Projection matrix P in eq. (4.108) is written in terms of vector v that spans the range, but it may also be written in terms of the vectors that span its nullspace. Let u_1 and u_2 be two orthogonal unit vectors on the plane orthogonal to v. Then instead of eq. (4.108) we have

$$P = I - u_1u_1^T - u_2u_2^T \tag{4.109}$$

and verify that $Pv = v$, and that $Pu_1 = Pu_2 = o$.

Orthogonal projection (see Fig. 43) of R^3 on plane π generated by the orthonormal (i.e. unit and orthogonal) u_1, u_2 is accomplished by projection matrix

$$P = u_1u_1^T + u_2u_2^T. \tag{4.110}$$

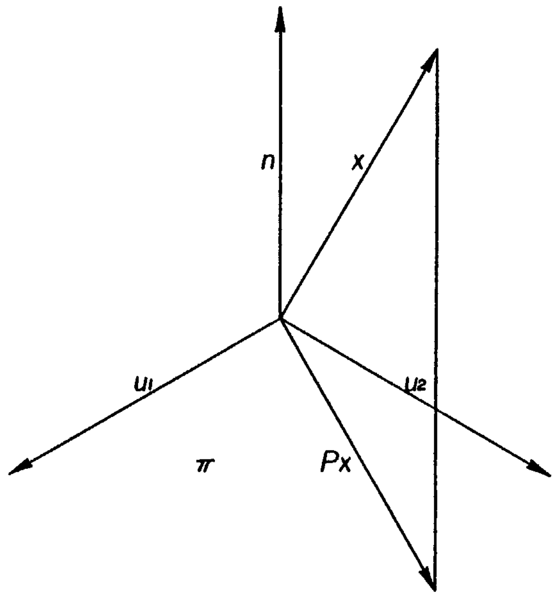


Fig. 4.43

We verify that $Pu_1 = u_1,\ Pu_2 = u_2$, and that $Pn = o$, where n is a unit vector perpendicular to π. The same projection matrix is written in terms of n as

$$P = I - nn^T,\ n^T n = 1 \tag{4.111}$$

by the formal use of $u_1u_1^T + u_2u_2^T + nn^T = I$.

Orthogonal projection of $x \in R^3$ on a plane is imagined as the shadow x casts under an overhead sun.

Orthogonal projection is of great interest because of

Theorem 4.22. *Orthogonally projected vector Px is the best approximation to x in the range R of P;*

$$\|x - Px\| \leq \|x - x'\| \quad x' \in R,\ x \in R^3 \tag{4.112}$$

with equality holding if and only if $x' = Px$.

Proof. We shall produce an analytic proof to this geometrically obvious theorem. Let u_1, u_2, n be an orthonormal system of three vectors of which u_1, u_2 span the range of the projection (a plane), and of which n spans the nullspace (orthogonal line) of the projection, so that $Pu_1 = u_1,\ Pu_2 = u_2,\ Pn = o$. If $x = \alpha_1u_1 + \alpha_2u_2 + \nu n$, then $Px = \alpha_1u_1 + \alpha_2u_2$,

$x - Px = \nu n,\ \|x - Px\|^2 = \nu^2$. Every vector x' in the range can be expressed as $x' = \alpha'_1 u_1 + \alpha'_2 u_2$, and

$$\|x - x'\|^2 = (\alpha_1 - \alpha'_1)^2 + (\alpha_2 - \alpha'_2)^2 + \|x - Px\|^2. \tag{4.113}$$

Hence the inequality of the theorem. Equality happens in the above equation for $\alpha_1 = \alpha'_1,\ \alpha_2 = \alpha'_2$ only, or $x' = \alpha_1 u_1 + \alpha_2 u_2 = Px$. End of proof.

Exercises

4.7.1. Which of the following matrices

$$P_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ P_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\ P_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},\ P_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},\ P_5 = \begin{bmatrix} 1 & 0 \\ \alpha & 0 \end{bmatrix}$$

is a projection matrix? Find the range and nullspace of P_5.

4.7.2. Show that if $A^2 = A,\ A \neq I$, then $P = I - A$ is a projection matrix. Also, that if $A^2 = I,\ A \neq I$, then $P = \frac{1}{2}(I + A)$ is also a projection matrix.

4.7.3. Prove that $P = A(A^TA)^{-1}A^T$ is a projection matrix.

4.7.4. For what value of α is

$$P = \alpha \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

a projection matrix.

4.7.5. Is the product of two projection matrices a projection matrix? Let $P_1 = I - u_1u_1^T$ and $P_2 = I - u_2u_2^T$ be two orthogonal projections of R^3. Prove that their product is an orthogonal projection if and only if $u_1^Tu_2 = 0$ or $P_1P_2 = P_2P_1$.

4.8 Nonsingular mappings

Nonsingular mappings have remarkable geometrical properties that essentially emanate from

Theorem 4.23. *Nonsingular matrix A maps noncolinear vectors into noncolinear vectors and noncoplanar vectors into noncoplanar vectors.*

Proof. If A is nonsingular, then in $x' = Ax$, x and x' are both zero or both nonzero. Suppose v_1, v_2, v_3 noncoplanar. Then for any $\alpha_1, \alpha_2, \alpha_3$, not all zero

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = v \neq o \tag{4.114}$$

and hence

$$\alpha_1 A v_1 + \alpha_2 A v_2 + \alpha_3 A v_3 = Av \neq o. \tag{4.115}$$

End of proof.

Linear transformation $x' = Ax$ can be looked upon as the mapping of vector spaces, or what to geometry is more pertinent, as the mapping of points. In the latter, position vector x of point $X(x_1, x_2, x_3)$ is mapped by $x' = Ax$ into position vector x' of point $X'(x'_1, x'_2, x'_3)$ with reference to the same Cartesian coordinate system. The origin is mapped into itself, and if the mapping is nonsingular, then to one point X there corresponds one point X' and vice versa.

We conclude from Theorem 4.23 that nonsingular linear transformations preserve lines (the mapping is a *collineation*). Indeed, if the equation of the line passing through A, B is given in terms of position vectors as

$$x = \alpha a + \beta b \quad \alpha + \beta = 1 \tag{4.116}$$

with a and b being noncolinear then

$$Ax = \alpha A a + \beta A b \tag{4.117}$$

or

$$x' = \alpha a' + \beta b' \quad \alpha + \beta = 1 \tag{4.118}$$

in which a', b' remain noncolinear.

In the same way one shows that planes remain plane under nonsingular linear mappings.

A nonsingular linear mapping is essentially a reversible, one-to-one, transformation of a vector space onto itself. Geometrically speaking a nonsingular linear transformation maps a line onto a line, a plane onto a plane, and space onto itself. Such transformations realistically occur when a rigid body tilts in space or when a pliable solid undergoes an elastic squeezing deformation. If the elastic deformation is not too severe and no overlapping or tearing occurs in it, then it is reasonable to assume that the correspondence between the points on the undeformed solid and the deformed are one-to-one. Even simpler, we may think of the transformation of a plane as an elastic sheet being stretched differently in different directions. Figures drawn on the sheet are distorted in the stretching, but for obvious geometrical reasons, and admittedly also for analytic tractability, we are most interested in transformations that preserve straight lines.

Before returning to linear algebraic formalities we want to consider the geometrically significant, one-to-one mapping, taking place in *perspective projection*.

We look first at the simpler case of projection of lines. Refer to Fig. 44 and think of point S as being a source of light. A ray of light emitted from S traverses point P on inclined line λ, and casts it as point P' on line λ'. Let x be the coordinate of point P on λ, and x' the coordinate of point P' on λ', for this choice of origins.

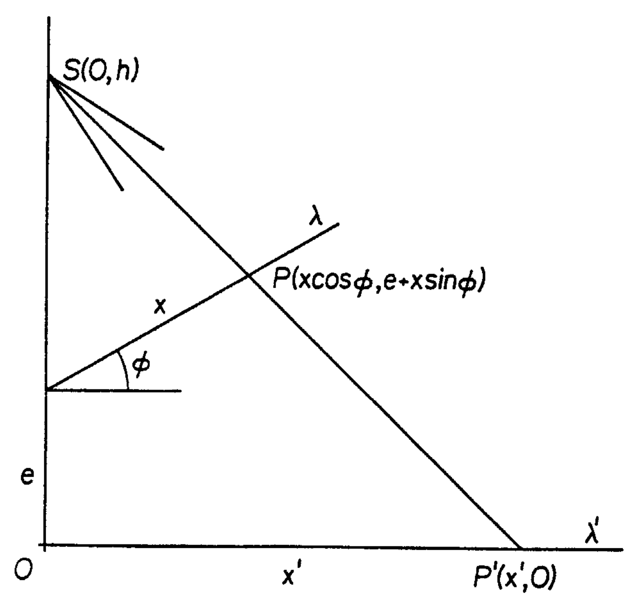


Fig. 4.44

Some simple algebra leads to

$$x' = \frac{x\cos\phi}{1-(x/h)\sin\phi}$$
$$x = \frac{x'}{\cos\phi + (x'/h)\sin\phi} \tag{4.119}$$

implying analytically what is geometrically obvious, that the mapping is reversible with a one-to-one correspondence between x and x', provided that the ray of light is not parallel to line λ', or that $1-(x/h)\sin\phi \neq 0$. Projective mapping by a point source of light extends the line, but not uniformly. Point A, B, C, D on line λ in Fig. 45 are

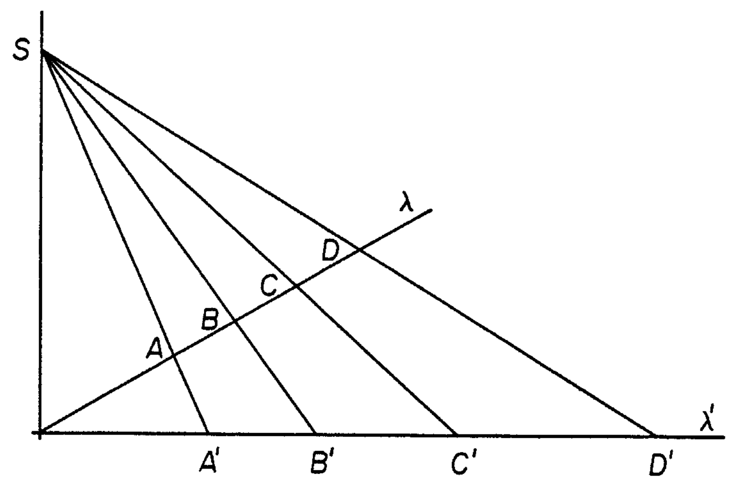


Fig. 4.45

equidistant but not so their images A', B', C', D'. Betweenness is preserved by the mapping of Fig.45—point B is between points A and C—and so is point B' that is between points A' and C', but *ratio* is not preserved by it. Point B is midway between points A and C, but point B' is not midway between points A', C'. What is preserved, or what is *invariant*, under the mapping is the ratio of ratios, or *cross ratio* $(AC/BC)/(AD/BD)$, the proof of which is left as an exercise. But if point S recedes into infinity, then the rays of light come down in parallel sheafs, and transformation (4.119) simplifies into $x' = x\cos\phi$, which amounts to a uniform extension or contraction of any segment on line λ by the ratio of $1/\cos\phi$.

Perspective projection of planes is similar. In Fig. 46 a point light source at $S(s_1, s_2, s_3)$, given in

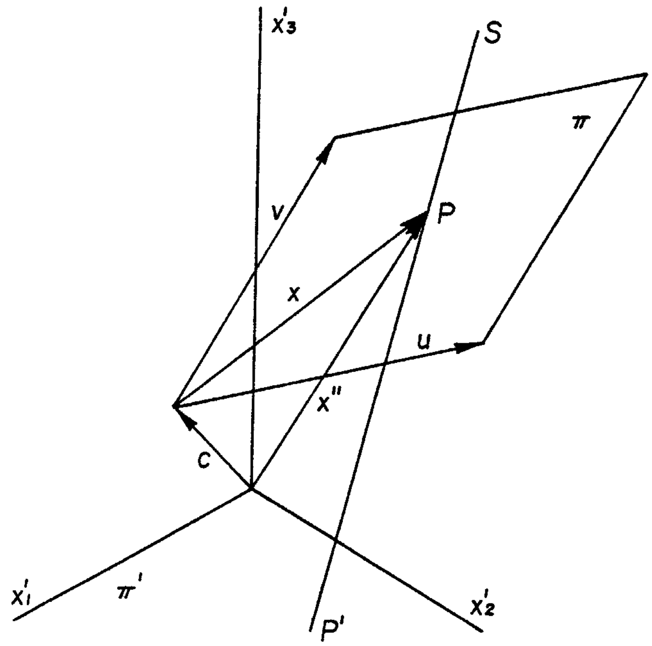


Fig. 4.46

reference to Cartesian coordinate system $0x_1{}'x_2{}'x_3{}'$ sheds point P of plane π upon point P' of plane $\pi', x_3{}' = o$.

Plane π is fixed by point $C(c_1, c_2, c_3)$ and the orthonormal vector pair u, v. With x_1, x_2 taken as parameters, the parametric equation of plane π assumes the form

$$\begin{bmatrix} x_1^{''} \\ x_2^{''} \\ x_3^{''} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + x_1 \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + x_2 \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} . \tag{4.120}$$

The parametric equation of the line through P and S is

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \lambda \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} + (1-\lambda) \begin{bmatrix} x_1^{''} \\ x_2^{''} \\ x_3^{''} \end{bmatrix} \tag{4.121}$$

and at $x_3' = 0$

$$\lambda = -\frac{x_3^{''}}{s_3 - x_3^{''}} \quad , \ 1 - \lambda = \frac{s_3}{s_3 - x_3^{''}} \quad . \tag{4.122}$$

When $s_3 = x_3^{''}$, when the ray of light is parallel to plane π', point P' falls at infinity.

Substitution of $x^{''}$ from eq. (4.120) into eqs. (4.121) and (4.122) results in the *rational* mapping

$$
\begin{aligned}
x_1' &= \frac{(u_1s_3 - u_3s_1)x_1 + (v_1s_3 - v_3s_1)x_2 + (c_1s_3 - c_3s_1)}{-u_3x_1 - v_3x_2 + s_3 - c_3} \\
x_2' &= \frac{(u_2s_3 - u_3s_2)x_1 + (v_2s_3 - v_3s_2)x_2 + (c_2s_3 - c_3s_2)}{-u_3x_1 - v_3x_2 + s_3 - c_3}
\end{aligned}
\tag{4.123}
$$

or in short

$$
x_1' = \frac{\alpha_1x_1 + \alpha_2x_2 + \alpha_3}{\beta_1x_1 + \beta_2x_2 + \beta_3},\ x_2' = \frac{\alpha_1'x_1 + \alpha_2'x_2 + \alpha_3'}{\beta_1x_1 + \beta_2x_2 + \beta_3}
\tag{4.124}
$$

and it *preserves lines*. If

$$
\alpha'x_1' + \beta'x_2' + \gamma' = 0,
\tag{4.125}
$$

then, so long as $\beta_1x_1 + \beta_2x_2 + \beta_3 \neq 0$,

$$
\alpha x_1 + \beta x_2 + \gamma = 0,
\tag{4.126}
$$

where

$$
\alpha = \alpha'\alpha_1 + \beta'\alpha_1' + \gamma^{\gamma}\beta_1,\ \beta = \alpha'\alpha_2 + \beta'\alpha_2' + \gamma'\beta_2,\ \gamma = \alpha'\alpha_3 + \beta'\alpha_3' + \gamma'\beta_3.
\tag{4.127}
$$

The case of parallel planes is simplified with the choice of $u = [1\ 0\ 0]^T,\ v = [0\ 1\ 0]^T$, that reduces the mapping to

$$
\begin{aligned}
x_1' &= \frac{s_3}{s_3 - c_3}x_1 + \frac{c_1s_3 - c_3s_1}{s_3 - c_3} \\
x_2' &= \frac{s_3}{s_3 - c_3}x_2 + \frac{c_2s_3 - c_3s_2}{s_3 - c_3}
\end{aligned}
\tag{4.128}
$$

which apart from the shift is pure enlargement.

To write the mapping of inclined planes by parallel rays we take $u = [1\ 0\ 0]^T,\ v = [0\ \cos\phi\ \sin\phi]^T$, set $\sigma_1 = s_1/s_3$, $\sigma_2 = s_2/s_3$, and let $s_3 \to \infty$, resulting in

$$
\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} 1 & -\sigma_1\sin\phi \\ 0 & \cos\phi - \sigma_2\sin\phi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_1 - \sigma_1c_3 \\ c_2 - \sigma_2c_3 \end{bmatrix}
\tag{4.129}
$$

which is linear.

In the case of an overhead sun, $\sigma_1 = \sigma_2 = 0$, and if we also take $c_1 = c_2 = 0$, projection (4.129) becomes

$$
x_1' = x_1,\ x_2' = x_2\cos\phi
\tag{4.130}
$$

and π' is stretched along x_2' only. Orthogonal projection (4.130) is not general enough to map a triangle of any shape into another triangle of any shape. We shall see in Section 10 that for such general mapping two independent stretches, or magnifications, are needed along x'_1 and x'_2. To parallel ray projection (4.130) of inclined planes we must add an *enlargement* κ so as to have

$$x'_1 = \kappa x_1,\ x'_2 = \kappa \cos\phi x_2, \tag{4.131}$$

or we must project twice to produce noncolinear extensions.

We return to linear mappings.

Theorem 4.24. *Linear mapping of R^3 onto itself is uniquely determined by three noncoplanar (two noncolinear in R^2) vectors and their three respective noncoplanar images.*

Proof. Let the three vectors be x_1, x_2, x_3 and their three images under the supposed mapping $x' = Ax$ be x'_1, x'_2, x'_3 so that

$$[x'_1\ x'_2\ x'_3]\ = A[x_1\ x_2\ x_3] \tag{4.132}$$

or in short $X' = AX$. The three columns of X are noncoplanar and matrix X is invertible. So is X', $X'X^{-1} = A$, and A is nonsingular. End of proof.

Considering linear mapping as that of points, Theorems 4.23 and 4.24 may be restated as

Corollary 4.25. *Under nonsingular linear mapping:*

1. Three noncolinear points are mapped into three noncolinear points, and three colinear points are mapped into three colinear points.

2. Four noncoplanar points are mapped into four noncoplanar points, and four coplanar points are mapped into four coplanar points.

3. Parallel lines are mapped into parallel lines and intersecting lines into intersecting lines, with the intersections being the image of one another. Parallel planes are mapped into parallel planes and intersecting planes into intersecting planes, with the intersections being the image of one another.

4. Three arbitrary points in the plane are mapped into three arbitrary points in the plane, and four arbitrary points in space are mapped into four arbitrary points in space.

Statement 4 asserts that an arbitrary triangle is mapped onto an arbitrary triangle. All geometrical properties that are preserved by the nonsingular linear mapping hold true for both triangles. A theorem involving these properties proved for *one particular* triangle is proved for an *arbitrary* triangle. The study of invariants under nonsingular linear mappings belongs to *affine geometry*.

Length, angle, area, and volume are not preserved by nonsingular linear transformations, and are therefore outside the scope of affine geometry, but betweenness and length ratios on a line are maintained.

Projective geometry studies the invariants left under perspective projection or rational mapping. It excludes mention not only of angle, length, and area, but also of segment ratio, which is not preserved by it. The theorem of Pythagoras belongs not in affine geometry and not in projective geometry. Theorem 4.10 on the intersection of medians in a triangle belongs in affine geometry but not projective. Desargue's theorem on coaxial and copolar triangles in one plane is a theorem of projective geometry as it deals with the incidence of lines and points only.

Theorem 4.26. *Let A, M, B be three colinear points and A', M', B' their image under a nonsingular linear mapping. If point M is between points A and B, then so is M'; and wherever M is, $AM/AB = A'M'/A'B'$.*

Proof. Suppose A and B are noncoincidental. Then M is uniquely fixed by λ in the position vector equation

$$m = (1-\lambda)a + \lambda b \tag{4.133}$$

and when $0 < \lambda < 1$ point M is between points A and B. For any λ,

$$\|m - a\| = |\lambda| \, \|b - a\| \tag{4.134}$$

and $AM/AB = |\lambda|$. Mapping by the nonsingular A produces

$$m' = (1-\lambda)a' + \lambda b' \tag{4.135}$$

where $m' = Am$, and where $a' = Aa$ and $b' = Ab$ are noncolinear. Point M' is between points A' and B', and $A'M'/A'B' = |\lambda| = AM/AB$. End of proof.

An interesting result of the last theorem is

Corollary 4.27. *Nonsingular matrix A maps the interior of a triangle onto the interior of the image triangle.*

The proof of which is left to the reader.

Theorem 4.28. *If triangle ABC is mapped by the nonsingular $A = A(2\times 2)$ into triangle $A'B'C'$, and tetrahedron $ABCD$ is mapped by nonsingular $A = A(3 \times 3)$ into tetrahedron $A'B'C'D'$, then*

$$\begin{aligned} \text{Area } A'B'C' \ / \ \text{Area } ABC &= \det(A) \\ \text{Volume } A'B'C'D' \ / \ \text{Volume } ABCD &= \det(A). \end{aligned} \tag{4.136}$$

Proof. For triangle ABC let $a = \vec{AB}$ and $b = \vec{AC}$. Then $\det[a\ b] = 2$ Area ABC. The mapping is written as $[a'\ b'] = A[a\ b]$, and

$$\det[a'\ b'] = \det(A)\ \det[a\ b] \tag{4.137}$$

proves the first part of the theorem. The proof for the tetrahedron is analogous. End of proof.

If $\det(A) > 0$, then the triangle (tetrahedron) is mapped into a triangle (tetrahedron) of the same orientation, but when $\det\ (A) < 0$ the orientation is reversed.

Example. Triangle ABC be given by $a = \vec{AB} = [1\ \ -2\ 1]^T$ and $b = \vec{AC} = [1\ 1\ 1]^T$. It is mapped by

$$A = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -2 \end{bmatrix}, \ \det(A) = -2 \tag{4.138}$$

into triangle $A'B'C'$ with $a' = [1\ \ -2\ \ -2]^T$, $b' = [1\ 1\ \ -2]^T$, and we compute from eq. (4.73)

$$\text{Area } ABC = 3\sqrt{2}/2,\ \text{Area } A'B'C' = 3\sqrt{5}/2,\ \text{Area } A'B'C'/\text{Area } ABC = \sqrt{5/2}. \tag{4.139}$$

Segment length is not invariant under nonsingular mapping. Contraction and elongation of the image segment takes place as a function of inclination. Suppose that $x' = Ax$, $A = A(2\times 2)$, and that $x = [\cos\theta\ \sin\theta]^T$, $x^Tx = 1$. As θ varies between 0^o and 360^o position vector x traces a unit circle, while image vector x' changes in length as $\rho^2 = x'^T x' = x^T A^T A x$. Because the mapping is nonsingular

$$B = A^T A = \begin{bmatrix} A_{11}^2 + A_{21}^2 & A_{11}A_{12} + A_{22}A_{21} \\ A_{11}A_{12} + A_{22}A_{21} & A_{12}^2 + A_{22}^2 \end{bmatrix} \tag{4.140}$$

is symmetric and positive definite.

We expand $x^T B x$ into

$$\rho^2 = B_{11}\cos^2\theta + B_{22}\sin^2\theta + 2B_{12}\sin\theta\cos\theta \tag{4.141}$$

and use the trigonometrical identities

$$\sin 2\theta = 2\sin\theta\cos\theta,\ \sin^2\theta = \frac{1}{2}(1-\cos 2\theta),\ \cos^2\theta = \frac{1}{2}(1+\cos 2\theta) \tag{4.142}$$

to rewrite quadratic form (4.141) as

$$\rho^2 = \frac{1}{2}(B_{11}+B_{22}) + \frac{1}{2}(B_{11}-B_{22})\cos 2\theta + B_{12}\sin 2\theta. \tag{4.143}$$

To group the last two terms of eq. (4.143) we introduce parameter κ and angle θ_0 through

$$\frac{1}{2}(B_{11}-B_{22}) = \kappa\cos 2\theta_0,\ B_{12} = \kappa\sin 2\theta_0 \tag{4.144}$$

and with $\cos^2 2\theta_0 + \sin^2 2\theta_0 = 1$ determine that

$$\kappa = \frac{1}{2}\sqrt{(B_{11}-B_{22})^2 + 4B_{12}^2},\ \tan 2\theta_0 = \frac{2B_{12}}{B_{11}-B_{22}}. \tag{4.145}$$

Now

$$\rho^2 = \frac{1}{2}(B_{11}+B_{22}) + \kappa\cos 2(\theta-\theta_0) > 0 \tag{4.146}$$

and ρ is seen to reach two extrema, a maximum at $\theta-\theta_0 = 0$ or $\theta-\theta_0 = \pi$, and a minimum at $\theta-\theta_0 = \pi/2$ or $\theta-\theta_0 = 3\pi/2$. The two extreme values of ρ, λ_1 and λ_2, given by

$$\begin{aligned} \min_\theta \rho^2 = \lambda_1^2 &= \frac{1}{2}(B_{11}+B_{22}) - \frac{1}{2}\sqrt{(B_{11}-B_{22})^2 + 4B_{12}^2} \\ \max_\theta \rho^2 = \lambda_2^2 &= \frac{1}{2}(B_{11}+B_{22}) + \frac{1}{2}\sqrt{(B_{11}-B_{22})^2 + 4B_{12}^2} \end{aligned} \tag{4.147}$$

are the principal *stretch ratios* of the mapping, and for any θ the relative segment stretch lies between the two extremes.

Orthogonal lines $\theta - \theta_0 = 0$ and $\theta - \theta_0 = \pi/2$ are the *principal axes* of the mapping, and we verify that

$$x_1 = [-\sin\theta_0 \;\; \cos\theta_0]^T \quad \text{and} \quad x_2 = [\cos\theta_0 \;\; \sin\theta_0]^T \tag{4.148}$$

are two orthogonal eigenvectors of B,

$$Bx_1 = \lambda_1^2 x_1 \;\; \text{and } Bx_2 = \lambda_2^2 x_2. \tag{4.149}$$

Angle θ measures the inclination of x, but ρ measures the length of x'. To identify the curve that x' traces as x rounds the unit circle we write the inverse map $x = A^{-1}x'$, and have that $1 = x^T x = {x'}^T C x'$, where

$$C = A^{-T}A^{-1} = \frac{1}{\Delta^2}\begin{bmatrix} A_{22}^2 + A_{21}^2 & -A_{22}A_{12} - A_{11}A_{21} \\ -A_{22}A_{12} - A_{11}A_{21} & A_{11}^2 + A_{12}^2 \end{bmatrix}, \; \Delta = \det(A), \tag{4.150}$$

is symmetric and positive definite.

With $x' = [\rho\cos\theta' \;\; \rho\sin\theta']^T$, ${x'}^T C x' = 1$ becomes

$$\frac{1}{\rho^2} = \frac{1}{2}(C_{11} + C_{22}) + \kappa' \cos 2(\theta' - \theta_0') \tag{4.151}$$

where, in analogy to eq. (4.144)

$$\frac{1}{2}(C_{11} - C_{22}) = \kappa' \cos 2\theta_0', \;\; C_{12} = \kappa' \sin 2\theta_0'. \tag{4.152}$$

Equation (4.151) is identified as that of an *ellipse* with principal axes at $\theta' - \theta_0' = 0$ and $\theta' - \theta_0' = \pi/2$. See Fig. 47.

Comparing B in eq. (4.140) and C in eq. (4.150) reveals that when $A = A^T$, when A is *symmetric*, $C_{11} = B_{22}/\Delta^2, \;\; C_{22} = B_{11}/\Delta^2, \;\; C_{12} = -B_{12}/\Delta^2$, and

$$\cos 2\theta_0' = -\cos 2\theta_0, \;\; \sin 2\theta_0' = -\sin 2\theta_0 \tag{4.153}$$

so that $\theta_0' = \theta_0 + \pi/2$. This implies that x_1 and x_1' are colinear, and also x_2 and x_2'. Scalars μ_1 and μ_2 exist so that $x_1' = \mu_1 x_1$ and $x_2' = \mu_2 x_2$,

$$Ax_1 = \mu_1 x_1, \;\; Ax_2 = \mu_2 x_2 \tag{4.154}$$

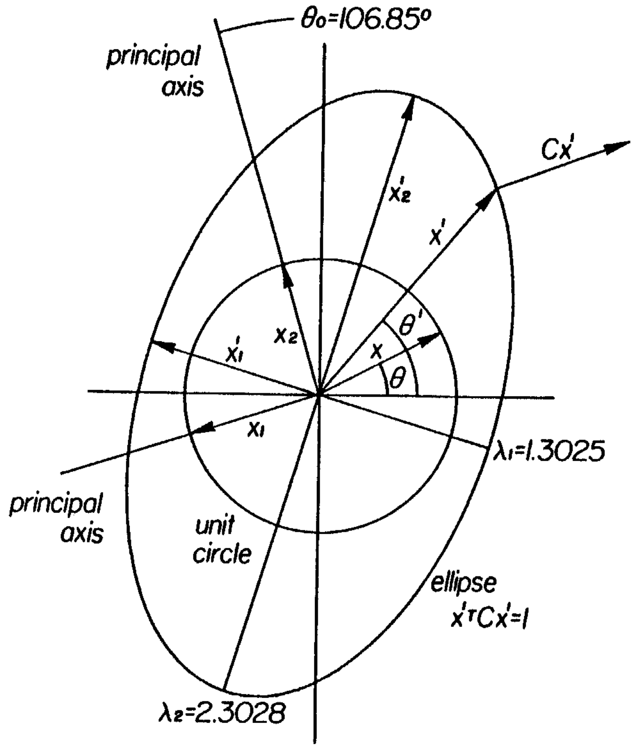


Fig. 4.47

and x_1 and x_2 are the orthogonal eigenvectors of the symmetric A with the corresponding eigenvalues $\mu_1^2 = \lambda_1^2$ and $\mu_2^2 = \lambda_2^2$.

Equation (4.149) is written for C as

$$Cx'_1 = \lambda_1^{-2} x'_1 \ , \ Cx'_2 = \lambda_2^{-2} x'_2 \tag{4.155}$$

geometrically explained with reference to Fig. 46. If we write $\phi(x') = {x'}^T Cx'$, then $\phi(x') = 1$ is one contour line of the quadratic function. Differentiation with respect to x' produces

$$grad\ \phi = 2Cx' \tag{4.156}$$

and a necessary and sufficient condition that $\|x'\|/\|x\|$ be extremal is that the gradient of $\phi(x')$ be colinear with position vector x'.

Figure 47 is drawn for the example of

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \ , \ B = \begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix} \ , \ C = \frac{1}{9} \begin{bmatrix} 5 & -1 \\ -1 & 2 \end{bmatrix} \tag{4.157}$$

with the computed values of θ_0, λ_1 and λ_2 shown on the graph.

An explicit analysis of this kind for the nonsingular mapping of the unit *sphere* is utterly prohibitive. It is left to us to remark that for $x \in R^3$ and $x' = Ax$, $x'^T Cx' = 1$ traces the surface of an *ellipsoid*. Every plane through the origin cuts the surface in an ellipse. Let u_1, u_2 be two orthonormal vectors in space. If x' is restricted to the plane generated by them, $x' = \alpha_1 u_1 + \alpha_2 u_2$, then

$$\alpha_1^2 u_1^T C u_1 + \alpha_2^2 u_2^T C u_2 + 2\alpha_1 \alpha_2 u_1^T C u_2 = 1 \tag{4.158}$$

and

$$C' = \begin{bmatrix} u_1^T C u_1 & u_1^T C u_2 \\ u_2^T C u_1 & u_2^T C u_2 \end{bmatrix} \tag{4.159}$$

is positive definite since $C = A^{-T} A^{-1}$ is positive definite. Equation (4.158) describes an ellipse. Among all ellipses on the different planes there is one with *longest* major axis and *shortest* minor axis.

Let λ_1 be the minimal stretch ratio of the mapping and λ_3 the maximal, occurring for orthogonal $x' = x'_1$ and $x' = x'_3$ so that $Cx'_1 = \lambda_1^{-2} x'_1$ and $Cx'_3 = \lambda_3^{-2} x'_3$. If x'_2 is a (unit) vector orthogonal to both x'_1 and x'_3, then also $Cx'_2 = \lambda_2^{-2} x'_2$. Indeed, if we write

$$Cx'_2 = \alpha_1 x'_1 + \alpha_2 x'_2 + \alpha_3 x'_3 \tag{4.160}$$

then premultiplication by $x_1'^T$ yields

$$\alpha_1 x_1'^T x'_1 = x_1'^T C x'_2 = x_2'^T C x'_1 = \lambda_1^{-2} x_2'^T x'_1 = 0 \tag{4.161}$$

and $\alpha_1 = 0$ since $x_1'^T x'_1 \neq 0$. The same happens to α_3; we are left with $Cx'_2 = \alpha_2 x'_2$, and $\alpha_2 = \lambda_2^{-2}$.

Look at Fig. 48. Mapping $R^2 \to R^2$ carries the circle into an ellipse and the tangent square into a tangent parallelogram with tangent points going onto tangent points (why?) As the square rotates the parallelogram rotates with it until it becomes a rectangle with sides parallel to the principal axes of the ellipse. Then the orthogonal $\vec{OP}, \vec{OQ}$ are mapped into the orthogonal $\vec{OP}', \vec{OQ}'$. In R^3 the situation is essentially the same except that for

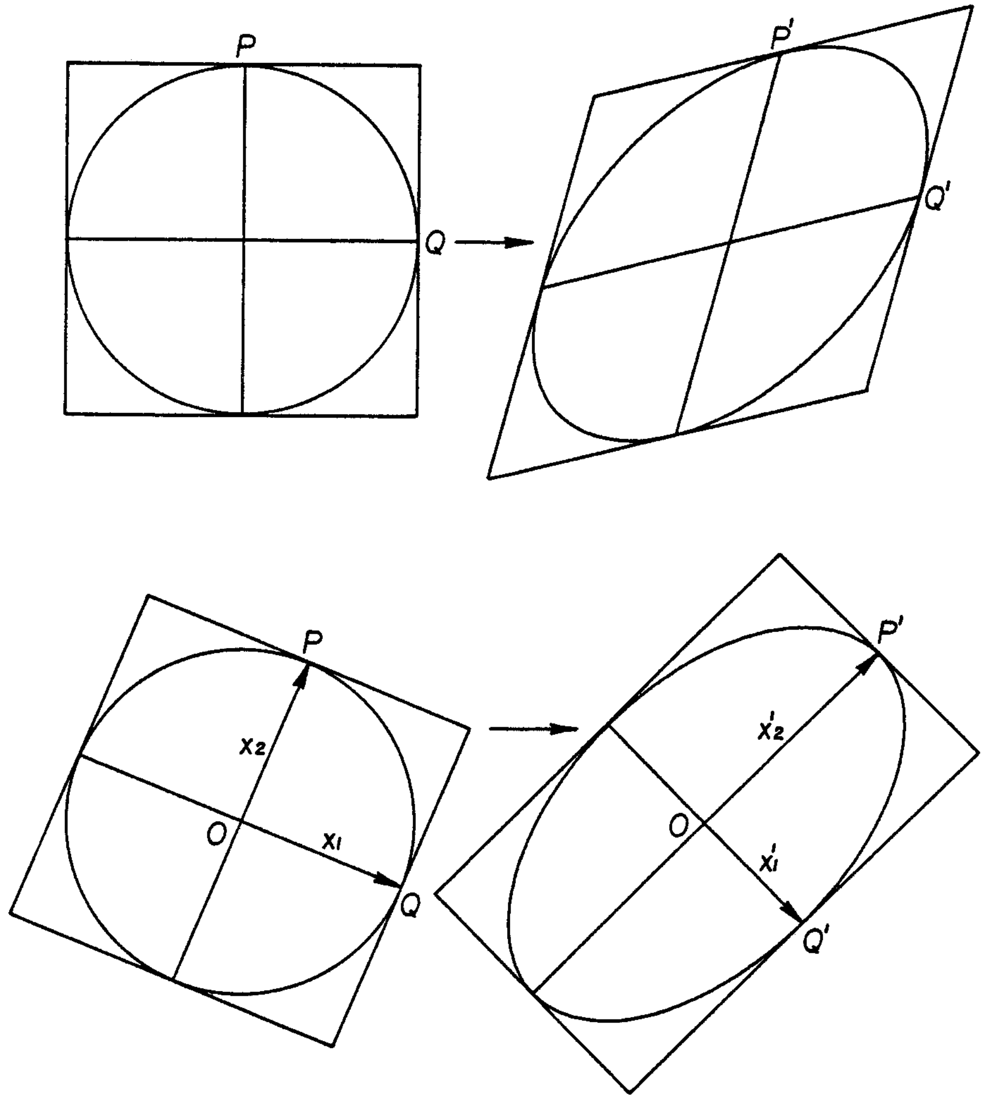


Fig. 4.48

circle and ellipse we have sphere and ellipsoid, and for square and parallelogram we have cube and parallelepiped.

In every nonsingular mapping of R^3 onto itself there exists an orthonormal vector triplet x_1, x_2, x_3 that is mapped into the orthogonal vector triplet x'_1, x'_2, x'_3.

Exercises

4.8.1. In the $R^2 \to R^2$ mapping given by linear transformation $x' = Ax$, find matrix A so that points $A(0,0), B(1,0), C(0,1)$ are mapped to points $A'(0,0), B'(2,1), C'(1,1)$, respectively.

4.8.2. Let distinct points A, B, C be on line λ and distinct points A, B', C' on λ'. Point A

is shared by both. Show a unique perspective projection exists by which points B, C are placed on points B', C', respectively. Discuss the projection that puts point B on point C' and point C on B'. What happens to the metaphor of light sources and point shadows?

4.8.3. Distinct points A, B, C are on line λ, and distinct points A', B', C' are on line λ'. Show that if the lines are coplanar, then two successive perspective projections can place one set of points on the other. What happens if the lines are not coplanar?

4.8.4. Let x be the coordinate of point P on line λ, and x' the coordinate of point P' on line λ'. Perspective projection of lines is algebraically described by the rational mapping

$$x' = \frac{\alpha_1 x + \alpha_2}{\beta_1 x + \beta_2}$$

which is the general form of eq.(4.119), and which is the one-dimensional counterpart to the perspective projection of planes described in eq.(4.124). Prove the fundamental theorem on the projective geometry of lines asserting that the projection is uniquely determined by three distinct points on λ and their three distinct images on λ'.

4.8.5. Figure 49 shows planes π and π' of Fig.46 on top of each other. Show that for this noncoplanar arrangement of triangles on planes that intersect at the line through A, B, equilateral triangle ABC can be projected by either radial or parallel rays into arbitrary triangle $A'B'C'$, and vice-versa. Is inclination $\phi \neq 0$ of π' relative to π unique?

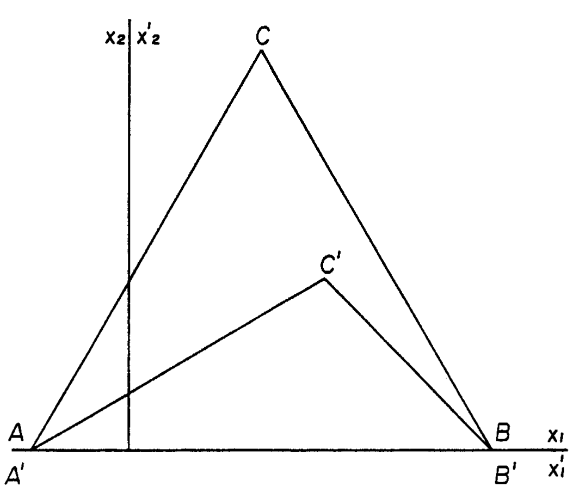


Fig. 4.49

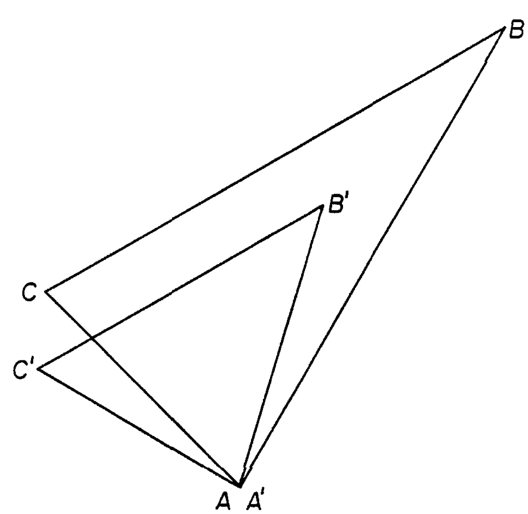


Fig. 4.50

4.8.6. With reference to Fig. 50, show that if arbitrary triangles ABC and $A'B'C'$ are placed in space so that vertex A coincides with vertex A', and sides CB and $C'B'$ are parallel, then the lines through B, B' and C, C' intersect. Hence prove that projection of planes by a point source of light is capable of transforming an arbitrary triangle into an arbitrary triangle. Is the projection unique? How are triangles of opposite orientation projected?

The fundamental theorem on the perspective projection of planes, which claims that four points A, B, C, D, no three of which are colinear, on plane π; and four points A', B', C', D', no three of which are colinear, on plane π' fix the projection uniquely, is more difficult to prove.

4.8.7. Fix the coefficients in transformation (4.124) so that $A(0,0), B(1,0), C(0,1), D(1,1)$ are respectively mapped to $A'(0,0), B'(1,1/2), C'(1/2,1), D'(1/2,1/2)$.

4.8.8. In the mapping

$$x_1' = \frac{2x_1 - x_2}{x_1 + 2x_2} \quad , \quad x_2' = \frac{-x_1 + x_2}{x_1 + 2x_2}$$

find all lines of the form $\alpha' x_1' + \beta' x_2' = 0$ that map into $\lambda\alpha' x_1 + \lambda\beta' x_2 = 0, \ \lambda \neq 0$.

4.8.9. Linear transformation $x' = Ax$ of planes is bounded and invariably maps the unit circle into an ellipse. Cylindrical perspective projection by parallel rays does the same. Conical perspective projection of planes is capable of sending a point to infinity and is therefore able to transform the unit circle into a general conic section. Consider rational transformation

$$x_1{}' = \frac{x_1 + \alpha x_2}{x_1 + x_2 + \gamma} \quad , \quad x_2{}' = \frac{\beta x_1 + x_2}{x_1 + x_2 + \gamma}$$

with parameters α, β, γ. Invert the transformation and write x_1, x_2 in terms of $x_1{}', x_2{}'$. Study the possible second-order curves on the x_1, x_2 plane corresponding to $x_1{}'^2 + x_2{}'^2 = 1$.

4.8.10. Prove that nonsingular linear mapping $x' = Ax$ maps the barycenter of a system of points to the barycenter of the mapped system of points.

4.8.11. Compute the principal axes and principal stretch ratios of $x' = Ax$ for

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$$

4.9 Orthogonal transformations; rotations and reflections

Linear mappings that preserve length are important. Movement of a *rigid body* is an example for such a mapping. It consists of *translations* whereby all material points move along parallel lines, of *rotations* around a fixed axis, or their combination. Mapping $x' = Ax$ of vectors x and x' excludes translation as the origin is always a *fixed point* of the transformation, but rotation of space is possible.

In this section we shall be concerned with the linear mapping $x' = Qx$ that preserves length; for which $\|x'\| = \|x\|$ for any $x \in R^3$.

Definition. *Square matrix Q is orthogonal if $Q^TQ = I$, that is if its columns are orthonormal.*

If Q is orthogonal, then $Q^T = Q^{-1}$ and also $QQ^T = I$. *Not only are the columns of Q orthonormal but also the rows.* A linear transformation that preserves length is said to be an *isometry.*

Theorem 4.29. *A necessary and sufficient condition that linear mapping $x' = Qx$ be an isometry is that Q is orthogonal.*

Proof. We have that ${x'}^T x' = x^TQ^TQx$ and hence if $Q^TQ = I$, then ${x'}^T x' = x^Tx$ and the condition is sufficient. To show necessity we choose $x = e_1, x = e_2,\ x = e_3$ and obtain that $(Q^TQ)_{ii} = 1$. The choice $x = [1\ 1\ 0]^T,\ x = [1\ 0\ 1]^T,\ x = [0\ 1\ 1]^T$ establishes that $(Q^TQ)_{ij} = 0,\ i \neq j$. End of proof.

Theorem 4.30. *The product of orthogonal matrices is an orthogonal matrix.*

Proof. Let Q_1 and Q_2 be orthogonal. If $Q = Q_1Q_2$, then $Q^TQ = Q_2^TQ_1^TQ_1Q_2 = I$. End of proof.

The linear mapping $x' = Qx$ with $Q^TQ = I$ is an *orthogonal transformation.*

Theorem 4.31. *Orthogonal transformation is conformal; angles are preserved by it.*

Proof. If $x_1' = Qx_1$ and $x_2' = Qx_2$, then ${x_1'}^T x_2' = x_1^TQ^TQx_2 = x_1^Tx_2$. Also, by the nature of the transformation, ${x_1}^Tx_1 = {x'_1}^Tx'_1,\ x_2^Tx_2 = {x'_2}^Tx_2'$, and hence according to eq.4.22 the angle between x_1 and x_2 is the same as the angle between x'_1 and x'_2. Notice, however, that $\cos\theta = \cos\theta'$ happens for both $\theta = \theta'$ and $\theta = -\theta'$. End of proof.

From generalities we descend to specifics and look first at plane orthogonal transformations for which the matrix is written generally as

$$Q = \begin{bmatrix} \cos\alpha & \cos\beta \\ \sin\alpha & \sin\beta \end{bmatrix} \ , \ \cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\beta-\alpha) = 0. \tag{4.162}$$

Angle β is related to α in eq. (4.162) either by $\beta = \alpha + \pi/2$, or $\beta = \alpha + 3\pi/2$ and we accordingly have the two basic plane orthogonal matrices

$$Q_1 = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \quad \text{and} \quad Q_2 = \begin{bmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{bmatrix}. \tag{4.163}$$

Matrix Q_1, for which $\det(Q_1) = 1$, preserves triangle orientation, and matrix Q_2, for which $\det(Q_2) = -1$, reverses it. Orthogonal mapping with Q_1 is said to be of the *first kind*, and that with Q_2 is said to be of the *second kind*.

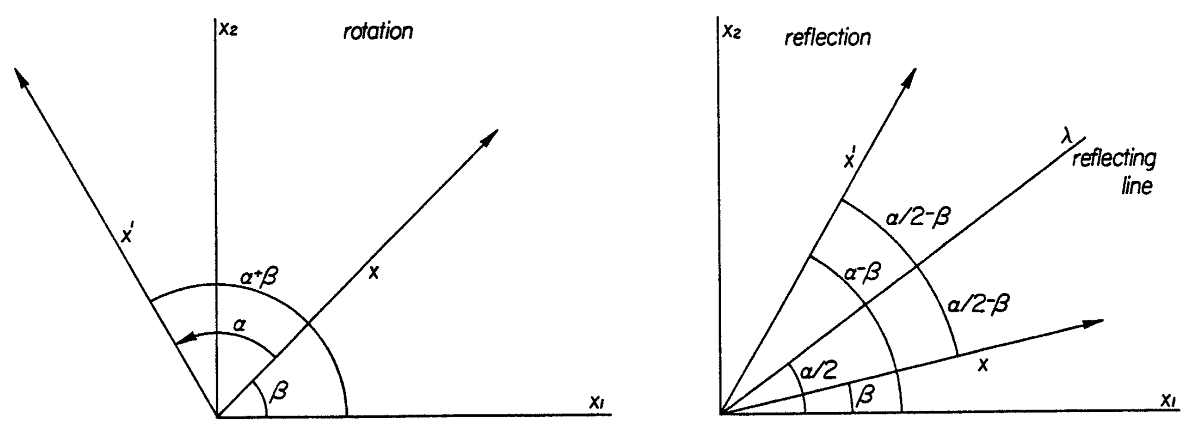
To see what Q_1 and Q_2 do to a unit vector we write $x = [\cos\beta \ \ \sin\beta]^T$ and have

$$x' = Q_1 x = [\cos(\alpha+\beta) \ \ \sin(\alpha+\beta)]^T \tag{4.164}$$

showing that matrix Q_1 *rotates* the plane by α degrees, as in Fig. 51a, whereas

$$x' = Q_2 x = [\cos(\alpha-\beta) \ \ \sin(\alpha-\beta)]^T \tag{4.165}$$

and matrix Q_2 *reflects* the plane in the line generated by $u = [\cos\alpha/2 \ \ \sin\alpha/2]^T$ as in Fig. 51b. Clearly $Q_1(-\alpha) = Q_1^T(\alpha)$, and $Q_1(\alpha)Q_1(\beta) = Q_1(\alpha+\beta)$.



(a) Fig. 4.51 (b)

Reflection of the plane in line λ can also be interpreted as a half-turn of the plane *in space* around the reflecting line.

Matrix Q_1 that performs rotation has generally no real eigenvector and no corresponding real eigenvalue since *all* vectors are turned by it, except for $\alpha = 180^o$ when $Q_1 = -I$. Then every $x \in R^2$ is an eigenvector of Q_1 with eigenvalue -1, $-Ix = -x$. Reflection matrix Q_2 has two orthogonal eigenvectors. If vector x is parallel to the reflecting line (the mirror), then $x = x'$, and if x is orthogonal to the reflecting line then $x = -x'$. The two eigenvalues of reflection matrix Q_2 are 1 and -1.

Let q_1 and q_2 denote the orthonormal columns (rows) of orthogonal matrix $Q = Q(2 \times 2)$. Figure 52 shows them for Q_1 and Q_2.

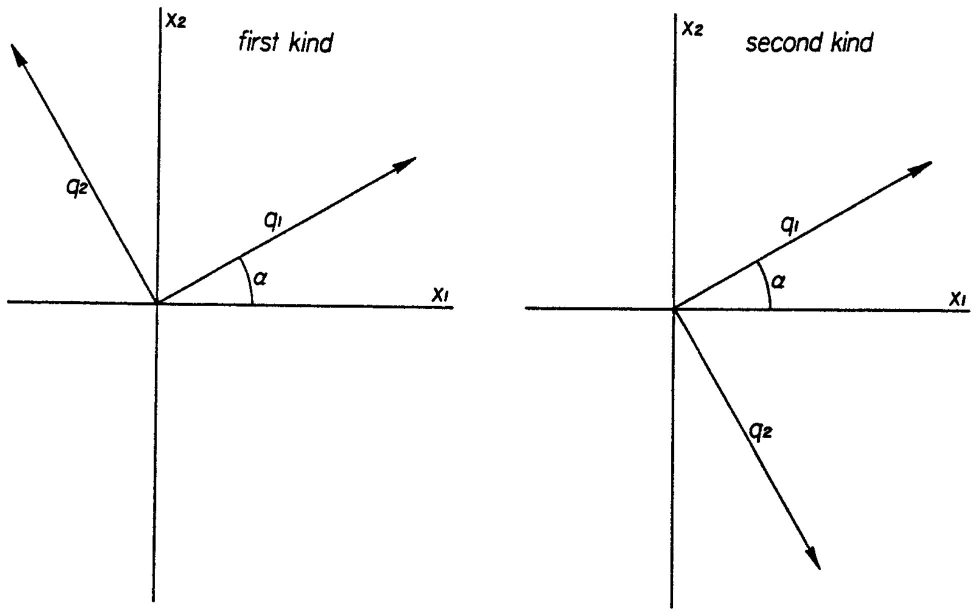


Fig. 4.52

Theorem 4.32. *Every plane rotation is the product of two reflections.*

Proof.

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{bmatrix} \begin{bmatrix} \cos\gamma & \sin\gamma \\ \sin\gamma & -\cos\gamma \end{bmatrix} = \begin{bmatrix} \cos(\beta-\gamma) & -\sin(\beta-\gamma) \\ \sin(\beta-\gamma) & \cos(\beta-\gamma) \end{bmatrix}. \tag{4.166}$$

End of proof.

One readily imagines two half-turns of a plane around two intersecting lines on it causing an in-plane rotation. This is what happens when a cone rolls without slipping on a plane.

The reflection matrix can be written with the aid of the projection matrix. With reference to Fig. 53 we establish that

$$x' = Px - (x - Px) = 2Px - x = (2P - I)x \tag{4.167}$$

and reflection matrix S (for Spiegelung) is

$$S = 2P - I = 2uu^T - I, \ u^T u = 1. \tag{4.168}$$

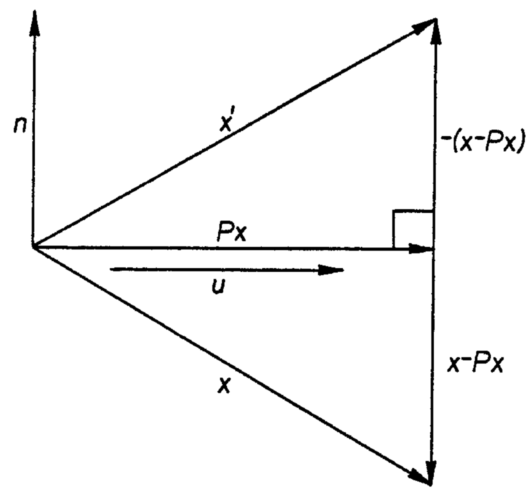


Fig. 4.53

Projection matrix P is such that $Pu = u$, $Pn = o$, whereas reflection matrix S is such that $Su = u$ and $Sn = -n$. Notice that matrix S is symmetric as is Q_2 in eq. (4.163).

We turn now to isometries of space.

Theorem 4.33. *Let $Q = Q(3 \times 3)$ be orthogonal. If $\det(Q) = 1$, then there exists $x \neq o$ so that $Qx = x$. If $\det(Q) = -1$, then there exists $x \neq o$ so that $Qx = -x$.*

Proof. To prove that if $\det Q = 1$, then there exists $x \neq o$ such that $(Q - I)x = o$, it is enough that we show that $Q - I$ is singular or $\det(Q - I) = 0$. Indeed

$$\begin{aligned} \det(Q - I) = \det(Q - QQ^T) &= \det Q(I - Q^T) \\ &= \det(I - Q^T)^T = \det(-I)\det(Q - I) \end{aligned} \tag{4.169}$$

and since for $I = I(n \times n)$, $\det(-I) = (-1)^n$ it results that $2\det(Q - I) = 0$.

To prove that if $\det(Q) = -1$, then there exists $x \neq o$ such that $(Q + I)x = o$, it is sufficient that we show that $Q + I$ is singular, or that $\det(Q + I) = 0$. In fact

$$\begin{aligned}\det(Q+I) &= \det(Q+QQ^T) = \det Q \det(Q+I) \\ &= -\det(Q+I)\end{aligned} \tag{4.170}$$

and $2\det(Q + I) = 0$. End of proof.

Theorem 4.34. *Let $x' = Qx$ be an isometry of R^3, $Q^TQ = I$. If $\det(Q) = 1$, then the isometry is a rotation. If $\det(Q) = -1$, then the isometry is a reflection followed by rotation.*

Proof. First let $\det(Q) = 1$. By Theorem 4.33 there exists a right hand orthonormal system q_1', q_2', q_3' in R^3 so that $Qq_1' = q_1'$. Let the images of q_2' and q_3' be denoted by q_2'' and q_3'', $Qq_2' = q_2''$, $Qq_3' = q_3''$. The image system $q_1'' = q_1', q_2'', q_3''$ is orthonormal and therefore q_2'', q_3'' are in the plane of q_2', q_3'. If Q' is with columns q_1', q_2', q_3', then $\det(Q'^TQQ') = 1$ and therefore

$$Q'^TQQ' = \begin{bmatrix} 1 & & \\ & {q'_2}^T q_2'' & {q_2'}^T q_3'' \\ & {q_3'}^T q_2'' & {q_3'}^T q_3'' \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \cos\alpha & -\sin\alpha \\ & \sin\alpha & \cos\alpha \end{bmatrix} \tag{4.171}$$

and space is rotated. Eigenvector q_1' spans the axis of rotation and all points move on planes orthogonal to q_1' such that if $x = q_2'\cos\theta + q_3'\sin\theta$, then $x' = Qx = q_2''\cos\theta + q_3''\sin\theta$, and $x^Tx' = \cos\alpha$ for any θ.

Now let $\det(Q)$ be -1. In this case $\det Q'^TQQ' = -1$, and if we choose q_1' so that $Qq_1' = -q_1'$, then

$$Q'^TQQ' = \begin{bmatrix} -1 & & \\ & {q_2'}^T q_1'' & {q_2'}^T q_3'' \\ & {q_3'}^T q_2'' & {q_3'}^T q_3'' \end{bmatrix} = \begin{bmatrix} -1 & & \\ & \cos\alpha & -\sin\alpha \\ & \sin\alpha & \cos\alpha \end{bmatrix} \tag{4.172}$$

and we have reflection followed by rotation. The reflecting plane, the mirror, is the plane orthogonal to q_1'. Vector x orthogonal to the mirror is reflected into vector $-x$, and vector x parallel to the mirror is rotated by α degrees in the plane. When $\alpha = 0$

$$Q'^TQQ' = \begin{bmatrix} -1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \tag{4.173}$$

which is pure reflection. End of proof.

To write the matrix for the rotation of R^3 around an axis spanned by the unit vector n, we choose the orthonormal pair u v to span the plane perpendicular to the axis of rotation. Figure 54 shows this plane with n pointing up.

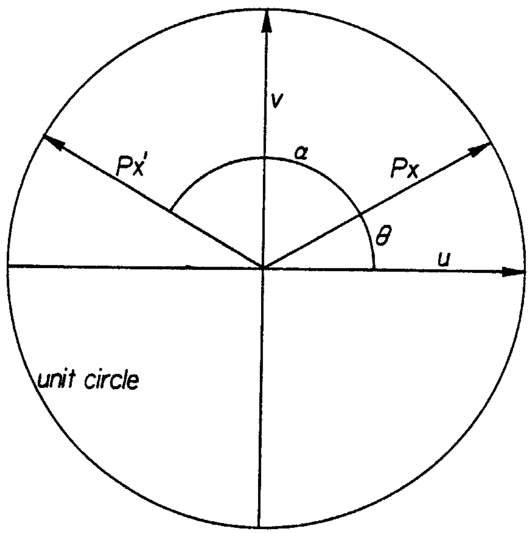


Fig. 4.54

Vector $x \in R^3$ is projected upon the u, v plane by

$$Px = (uu^T + vv^T)x \tag{4.174}$$

or, with reference to Fig. 54

$$Px = \|Px\|(u\cos\theta + v\sin\theta) \tag{4.175}$$

and $x = Px + (n^T x)n$. Rotation of space by α degrees around axis n produces the image

$$x' = \|Px\|(u\cos(\theta+\alpha) + v\sin(\theta+\alpha)) + (n^T x)n \tag{4.176}$$

that with $u^T Px = \|Px\|\cos\theta,\ v^T Px = \|Px\|\sin\theta$, and $uu^T + vv^T + nn^T = I$, is written as

$$x' = Rx,\ R = I + (\cos\alpha - 1)(uu^T + vv^T) + \sin\alpha(vu^T - uv^T) \tag{4.177}$$

and R is the space rotation matrix. We verify that, since $u^T n = v^T n = 0$,

$$Rn = n$$

$$Ru = u\cos\alpha + v\sin\alpha, Rv = -u\sin\alpha + v\cos\alpha = u\cos(\frac{\pi}{2}+\alpha) + v\sin(\frac{\pi}{2}+\alpha) \tag{4.178}$$

$$R(-\alpha) = R^T(\alpha) = R^{-1}(\alpha)$$

and

$$R(180^o) = I - 2(uu^T + uv^T) = I - 2P = -I + 2nn^T. \tag{4.179}$$

To write rotation matrix R in terms of unit vector $n = [n_1\ n_2\ n_3]^T$ that generates the rotation axis we introduce $N = vu^T - uv^T$ for which we have

$$N[u\ v\ n] = [v\ \ -u\ o]. \tag{4.180}$$

Vectors u, v, n are orthonormal, matrix $[u\ v\ n]$ is orthogonal, and

$$N = [v\ \ -u\ o][u\ v\ n]^T \tag{4.181}$$

or explicitly

$$N = \begin{bmatrix} 0 & v_1u_2 - u_1v_2 & v_1u_3 - u_1v_3 \\ v_2u_1 - u_2v_1 & 0 & v_2u_3 - u_2v_3 \\ v_3u_1 - u_3v_1 & v_3u_2 - u_3v_2 & 0 \end{bmatrix}. \tag{4.182}$$

We observe that

$$n = [v_3u_2 - u_3v_2\ \ v_1u_3 - u_1v_3\ \ v_2u_1 - u_2v_1]^T \tag{4.183}$$

is orthogonal to u and v, $v^Tn = u^Tn = 0$, and $n^Tn = 1$. In particular if $u = e_1$ and $v = e_2$, then $n = e_3$, and

$$N = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix} \tag{4.184}$$

with n_1, n_2, n_3 being the components of unit vector n that generates the rotation axis. Making use once more of $uu^T + vv^T + nn^T = I$ we write rotation matrix R in final form

$$R = I\cos\alpha + (1-\cos\alpha)nn^T + \sin\alpha N. \tag{4.185}$$

One is tempted to represent rotation α around axis n as vector αn, but the end effect of two rotations is *not* their vector sum. Finite rotation is not a vector. Only infinitesimal rotation and angular velocity are vectors. Two successive rotations are described by the ordered product of the two corresponding rotation matrices (Prove!)

Reflection of space in a plane perpendicular to unit vector n is accomplished by matrix

$$S = I - 2nn^T. \tag{4.186}$$

If u is any vector in the reflecting plane (mirror), then $Su = u$, but $Sn = -n$.

Reflection in a line generated in space by unit vector u is done with

$$S = -I + 2uu^T \tag{4.187}$$

which we compare with $R(180^o)$ in eq. (4.179). When n and u are $e_3 = [0\ 0\ 1]^T$, the reflection matrices in eqs. (4.186) and (4.187) become

$$S = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix} \text{ and } S = \begin{bmatrix} -1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \tag{4.188}$$

respectively.

Theorem 4.35. *Rotation of R^3 is the product of two reflections in planes containing the axis of rotation.*

Proof. Let u_1 and u_2 be two unit vectors in the plane perpendicular to the axis of rotation. With reference to Fig. 54 let $u_1 = u$, and $u_2 = u\cos\phi + v\sin\phi$. The product of the two reflection matrices $S_1 = I - 2u_1u_1^T$ and $S_2 = I - 2u_2u_2^T$ is

$$S_1S_2 = I + (\cos 2\phi - 1)(uu^T + vv^T) + (uv^T - vu^T)\sin 2\phi \tag{4.189}$$

and if $2\phi = -\alpha$, then $S_1S_2 = R$ as in eq. (4.177). End of proof.

Theorem 4.36. *Every orthogonal transformation is either a rotation or the product of a rotation and a reflection.*

Proof. This theorem is a restatement of Theorem 4.34. Let the columns of Q be q_1, q_2, q_3 and consider the two orthonormal systems of R^3 : $e_1 = [1\ 0\ 0]^T, e_2 = [0\ 1\ 0]^T, e_3 = [0\ 0\ 1]^T$, and q_1, q_2, q_3. There exists a rotation R_1 in the e_1, q_1 plane so that $R_1q_1 = e_1$. Vectors $q'_1 = e_1 = R_1q_1,\ q'_2 = R_1q_2,\ q'_3 = R_1q_3$ are still orthonormal with q'_3 and q'_2 being in the

e_2, e_3 plane. A second rotation around e_1 turns q_2' into $q_2'' = e_2$. Then $q_3'' = \pm e_3$. In matrix form

$$R_1 Q = [e_1 \ R_1 q_2 \ R_1 q_3] \ , \ R_2 R_1 Q = [e_1 \ e_2 \ \pm e_3] \tag{4.190}$$

For the plus sign we have $R_2 R_1 Q = I$, or $Q = R_1^T R_2^T$ which is a rotation matrix, whereas for the minus sign we have $S R_2 R_1 Q = I$ where S is the reflection matrix

$$S = I - 2 e_3 e_3^T = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix}. \tag{4.191}$$

Then $Q = R_1^T R_2^T S$. End of proof.

Corollary 4.37. *Every orthogonal transformation of R^3 is the product of at most three reflections,*

$$Q = (I + (\lambda_1 - 1) u_1 u_1^T)(I + (\lambda_2 - 1) u_2 u_2^T)(I + (\lambda_3 - 1) u_3 u_3^T) \tag{4.192}$$

where $\lambda_1, \lambda_2, \lambda_3$ are ± 1, and where $u_1^T u_1 = u_2^T u_2 = u_3^T u_3 = 1$.

Proof. According to Theorem 4.36 $Q = RI$ or $Q = RS$ where R is rotation and S reflection. Theorem 4.35 states that R itself is the product of two reflections and hence Q is the product of at most three reflections. End of proof.

Exercises

4.9.1. Compute the eigenvalues and eigenvectors of

$$Q_1 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \ Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \ Q_3 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \ Q_4 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and show that Q_1 performs a half-turn, and Q_3 a quarter-turn. Verify that Q_2 and Q_4 perform reflection and find the mirror.

Hint: $Q_4 e_1 = e_2, \ Q_4 e_2 = e_1, \ Q_4(e_1 + e_2) = e_1 + e_2$.

Show that $Q_3^2 = Q_1, \ Q_1^2 = I, \ Q_2^2 = I$ and that $Q_2 Q_3 = Q_4$.

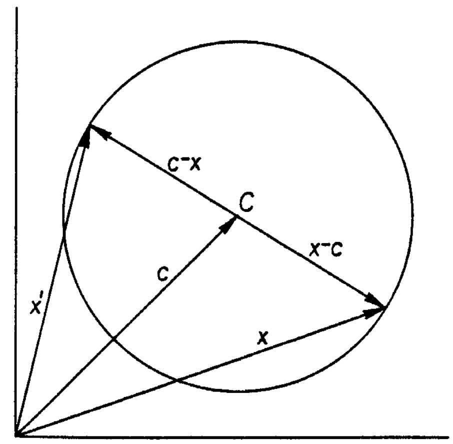


Fig.4.55

4.9.2. Verify that the punctual transformations of $R^2 \to R^2$

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad x' = Ix + c$$

and

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = -\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad x' = -Ix + 2c$$

are a translation and a half-turn around point $C(c_1, c_2)$, respectively.

Show that the product of two half-turns is a translation, the product of three half-turns is a half-turn, and that half-turns generally do not commute. When do half-turns commute? (Fig. 55.)

4.9.3. Verify that the punctual transformation of $R^2 \to R^2$

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x_1 - c_1 \\ x_2 - c_2 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad x' = R(x-c) + c$$

is a rotation around $C(c_1, c_2)$.

Show that the product of two plane rotations is a plane rotation of their angle sum. Find the compound center (Fig. 56.)

4.9.4. Verify that the punctual transformation $R^2 \to R^2$,

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2\gamma \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \quad x' = Sx + 2\gamma v$$

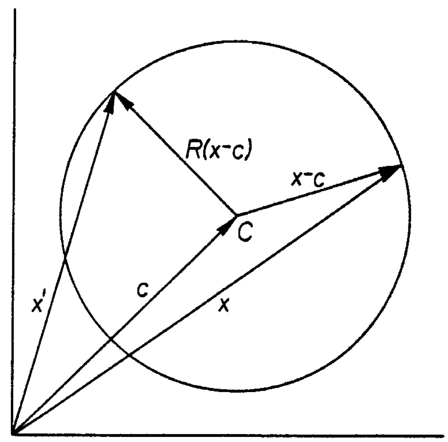


Fig.4.56

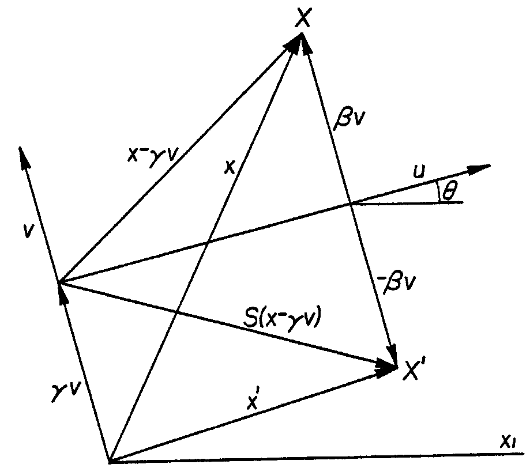


Fig.4.57

where $S = 2uu^T - I$, is a reflection in the line generated by $u = [\cos\theta \;\; \sin\theta]^T$, passing at a distance γ from the origin (Fig. 57.)

Show that the product of two reflections in parallel lines is a translation.

Show that the product of two reflections in lines making angle θ is a rotation of angle 2θ around their intersection point.

In particular, show that the product of two reflections in perpendicular lines is a half-turn.

4.9.5. The mapping $x' = \rho Q x$, where $\rho > 0$ and Q is a rotation matrix is called a *similarity* transformation (when $Q = I$ it is a *dilation*). Show that length is not preserved by it, but angle is.

4.9.6. Show that if Q rotates R^2 by angle θ, then

$$Q + Q^T = 2I\cos\theta, \text{ and } a^T(Q + Q^T)b = 2a^T b\cos\theta.$$

In particular if $a^T b = 0$ then $a^T Q b + b^T Q a = 0$.

4.9.7. Prove that if R is a rotation matrix turning R^3 by angle θ around $n = [n_1 \; n_2 \; n_3]^T$, then

$$\frac{1}{2}(R_{11} + R_{22} + R_{33} - 1) = \cos\theta$$

and

$$\frac{1}{2}(R_{12}-R_{21})=-n_3\sin\theta,\ \frac{1}{2}(R_{13}-R_{31})=n_2\sin\theta,\ \frac{1}{2}(R_{23}-R_{32})=-n_1\sin\theta$$

or

$$(R_{12}-R_{21})^2+(R_{13}-R_{31})^2+(R_{23}-R_{32})^2=4\sin^2\theta.$$

Show further that

$$\det(R+I)=4(1+\cos\theta).$$

Show that

$$Q=\begin{bmatrix} \sqrt{6}/6 & \sqrt{2}/2 & \sqrt{3}/3 \\ -2\sqrt{6}/6 & 0 & \sqrt{3}/3 \\ \sqrt{6}/6 & -\sqrt{2}/2 & \sqrt{3}/3 \end{bmatrix}$$

rotates R^3, find the axis and angle of rotation.

4.9.8. Prove that two rotations of R^3 commute if and only if they share the same axis.

4.9.9. Prove that if R_1 and R_2 are rotations of R^3 around axes passing through the origin, then R_1R_2 is a rotation around some axis through the origin.

4.9.10. Verify that if R rotates R^3 around n by angle θ, then the same rotation, around an axis that does not necessarily pass through the origin, is done by

$$x'=R(x-p)+p$$

where p is the position vector of an arbitrary point on the axis of rotation (Fig. 58.)

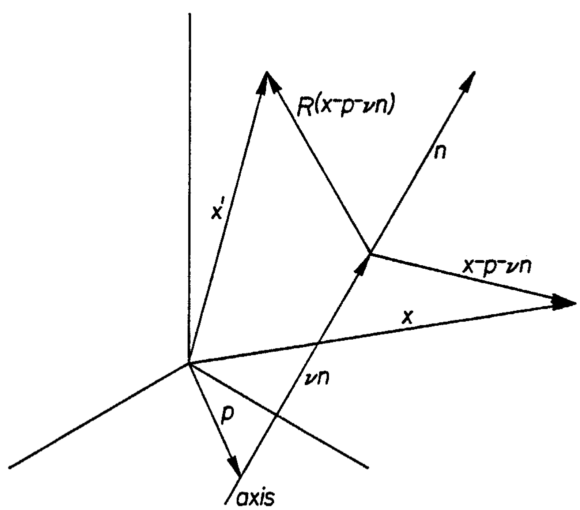


Fig.4.58

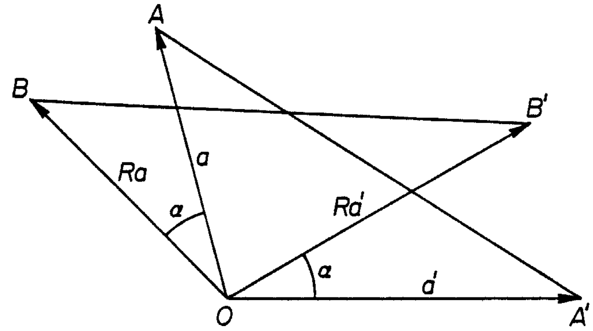


Fig.4.59

4.9.11. Prove that the product of rotations around parallel axes in R^3 is either a rotation or a translation.

4.9.12. Prove that the product of rotations around non-intersecting axes in R^3 is neither a rotation nor a translation, but rather a rotation followed by a translation.

4.9.13. Prove that if $OA = OB$, $OA' = OB'$ and angle AOB = angle $A'OB' = \alpha$, then AA' and BB' intersect at angle α. Make use of rotation matrix R (Fig. 59.)

4.9.14. Prove that the centroids G_1, G_2, G_3 of equilateral triangles erected externally on the sides of an arbitrary triangle form an equilateral triangle. Use the rotation matrix R (Fig. 60.)

4.9.15. Triangle ABC has right angle at C. Show that if G and G' are centers of exterior squares, and M is a midpoint, then GMG' is isosceles with a right angle at M. Use the rotation matrix R (Fig. 61.)

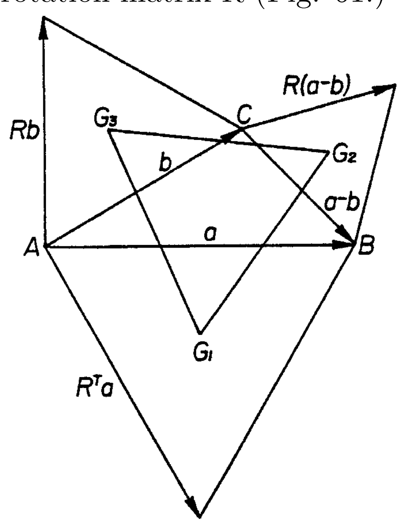


Fig.4.60

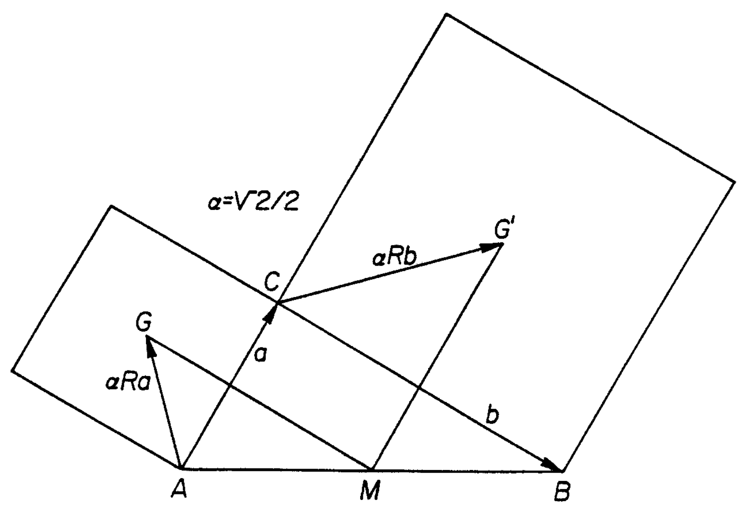


Fig.4.61

4.10 Spectral decomposition

Let u_1, u_2, u_3 be an orthonormal system in R^3. Since $I = u_1u_1^T + u_2u_2^T + u_3u_3^T$, reflection matrix $S = I - 2u_3u_3^T$ can also be written as

$$S = u_1u_1^T + u_2u_2^T - u_3u_3^T \tag{4.193}$$

so that $Su_3 = -u_3,\ Su_1 = u_1,\ Su_2 = u_2$, and $S^{-1} = S$. Reflection, we know, is the basic isometry with which any orthogonal mapping is generated.

Presently we introduce the more general *stretch* matrix

$$T = u_1u_1^T + u_2u_2^T + \lambda u_3u_3^T \quad T^{-1} = u_1u_1^T + u_2u_2^T + \lambda^{-1}u_3u_3^T \tag{4.194}$$

where u_1, u_2, u_3 are orthonormal, that we may also write as

$$T = I + (\lambda - 1)u_3u_3^T. \tag{4.195}$$

We will show it to be a basic element in the general nonsingular mapping of R^3.

Stretch matrix T is such that $Tu_1 = u_1,\ Tu_2 = u_2,\ Tu_3 = \lambda u_3$, and u_1, u_2, u_3 are three orthonormal eigenvectors of T corresponding to eigenvalues $1, 1, \lambda$. Writing $x = \alpha_1u_1 + \alpha_2u_2 + \alpha_3u_3$ and seeking λ' so that $Tx = \lambda'x$ we ascertain that stretch matrix T has no more eigenvalues. Vectors in the plane generated by u_1 and u_3 are unaffected by mapping T, but vectors along u_3 extend by factor λ. The unit sphere is mapped by T into an ellipsoid of revolution.

Three matrices T_1, T_2, T_3

$$T_1 = I + (\lambda_1 - 1)u_1u_1^T,\ T_2 = I + (\lambda_2 - 1)u_2u_2^T,\ T_3 = I + (\lambda_3 - 1)u_3u_3^T \tag{4.196}$$

affect successive orthogonal stretchings of R^3 and the compound $T = T_1T_2T_3$ is

$$T = \lambda_1u_1u_1^T + \lambda_2u_2u_2^T + \lambda_3u_3u_3^T. \tag{4.197}$$

As a prelude to a more general discussion, we raise at this point the following question on plane linear mappings: Given two noncolinear vectors a_1, a_2, does there exist an orthogonal double-stretching of their plane

$$T = \lambda_1u_1u_1^T + \lambda_2u_2u_2^T \tag{4.198}$$

that maps a_1, a_2 into an orthonormal q_1, q_2?

We write

$$[a_1\ a_2] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, [q_1\ q_2] = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, [u_1\ u_2] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{4.199}$$

and require that $Ta_1 = q_1,\ Ta_2 = q_2$. Elimination of λ_1 and λ_2 between the two equations yields

$$(u_1^T q_1)(u_1^T a_2) - (u_1^T q_2)(u_1^T a_1) = 0, \quad (u_2^T q_2)(u_2^T a_1) - (u_2^T q_1)(u_2^T a_2) = 0 \tag{4.200}$$

respectively, or

$$\begin{aligned}(A_{22} - A_{11})\sin(2\theta - \phi) + (A_{12} + A_{21})\cos(2\theta - \phi) + (A_{11} + A_{22})\sin\phi \\ + (A_{12} - A_{21})\cos\phi = 0, \\ (A_{22} - A_{11})\sin(2\theta - \phi) + (A_{12} + A_{21})\cos(2\theta - \phi) - (A_{11} + A_{22})\sin\phi \\ - (A_{12} - A_{21})\cos\phi = 0\end{aligned} \tag{4.201}$$

respectively. Hence

$$\begin{aligned}(A_{11} + A_{22})\sin\phi + (A_{12} - A_{21})\cos\phi = 0, \\ (A_{22} - A_{11})\sin(2\theta - \phi) + (A_{12} + A_{21})\cos(2\theta - \phi) = 0\end{aligned} \tag{4.202}$$

that can be solved for ϕ and θ.

In case $A_{12} = A_{21}$ we may take $\phi = 0$ so that $q_1 = e_1,\ q_2 = e_2$. Expressed as matrix products, if $A = A^T$, then $TA = I$ or $A = T^{-1}$ *and every nonsingular symmetric mapping of* R^2 *is the result of two stretchings along orthogonal axes. The singular case is trivial.* Agreeing that T is generic for stretch we write for $A = A^T,\ A = T$, and have that

$$Au_1 = \lambda_1 u_1 \text{ and } Au_2 = \lambda_2 u_2. \tag{4.203}$$

In every nonsingular symmetric mapping of R^2 there are two orthogonal vectors u_1, u_2 that have as images $u_1' = \lambda_1 u_1$ and $u_2' = \lambda_2 u_2$. This is not new to us. We have already seen it in Section 4.8 when we discussed the mapping of the unit circle into an ellipse.

Writing $A = T,\ x = \alpha_1 u_1 + \alpha_2 u_2$ and restricting x to $x^T x = 1$, we have that $\alpha_1^2 + \alpha_2^2 = 1$. Since $A^{-1} = (1/\lambda_1)u_1 u_1^T + (1/\lambda_2)u_2 u_2^T$ writing the image $x' = Ax$ as $x' = \alpha_1' u_1 + \alpha_2' u_2$ yields

$$\left(\frac{\alpha_1'}{\lambda_1}\right)^2 + \left(\frac{\alpha_2'}{\lambda_2}\right)^2 = 1 \tag{4.204}$$

which is an ellipse.

For nonsymmetric $A = A(2 \times 2)$ the above argument implies the existence of stretch matrix T and rotation matrix Q, $Q^TQ = I$, so that $TA = Q$, differently written as $A = TQ$, and we have

Theorem 4.38. *Every nonsingular mapping of R^2 is the result of a rotation followed by two orthogonal stretchings.*

Stretch T changes the area of a triangle by factor $\lambda_1\lambda_2$, and det $(T) = \lambda_1\lambda_2$. Hence det $(A) = \det(AQ) = \det(A)\ \det(Q) = \lambda_1\lambda_2$.

As an example consider the deformation or mapping performed by

$$A = \begin{bmatrix} 1 & k \\ & 1 \end{bmatrix} \tag{4.205}$$

called *shear*. Only one vector, $u = [1\ 0]^T$ is mapped colinearly by A and has the eigenvalue $\lambda = 1$. For $k = 1$ we compute from eq. (4.202) $\phi = -26.565^o$ and $\theta = 31.718^o$, and from $A = TQ$ we have

$$u_1^Ta_1 = \lambda_1 u_1^Tq_1,\ u_2^Ta_1 = \lambda_2 u_2^Tq_1 \tag{4.206}$$

so that $\lambda_1 = 1.618,\ \lambda_2 = 0.618$.

Figure 62 shows the decomposition of shear into a rotation and two perpendicular stretchings (actually a compression and a stretch).

Shear has the physical significance of involving the relative sliding movement of material layers–a deformation with friction. Such friction between layers of flowing fluid is termed *viscosity*.

In R^3, if the mapping $x' = Ax$ is such that $x^Tx = 1$, then $1 = x'(AA^T)^{-1}x'$ describes an ellipsoid. Every plane through the origin cuts it in an ellipse. Among all ellipses there is one with longest major axis and shortest minor axis. These are the maximum and minimum of ${x'}^Tx'$, respectively. It is a simple argument to show that the extrema of ${x'}^Tx'$ occur in orthogonal directions. Let the maximum of ${x'}^Tx'$ happen in the direction of x'_3 and the minimum in the direction of x'_1. They are the image of the orthogonal x_3, x_1. By the gradient argument of Section 4.8, $Cx'_1 = \lambda_1^{-2}x'_1$ and $Cx'_3 = \lambda_3^{-2}x'_3, C = (AA^T)^{-1}$. If x'_2 is a vector orthogonal to both x'_1 and x'_3, then $Cx'_2 = \lambda_2^{-2}x'_2$. We have that

$$x_1^Tx_2 = {x_1'}^T A^{-T}A^{-1}x'_2 = 0 \quad , \quad x_1^Tx_3 = {x_1'}^T A^{-T}A^{-1}x'_3 = 0 \tag{4.207}$$

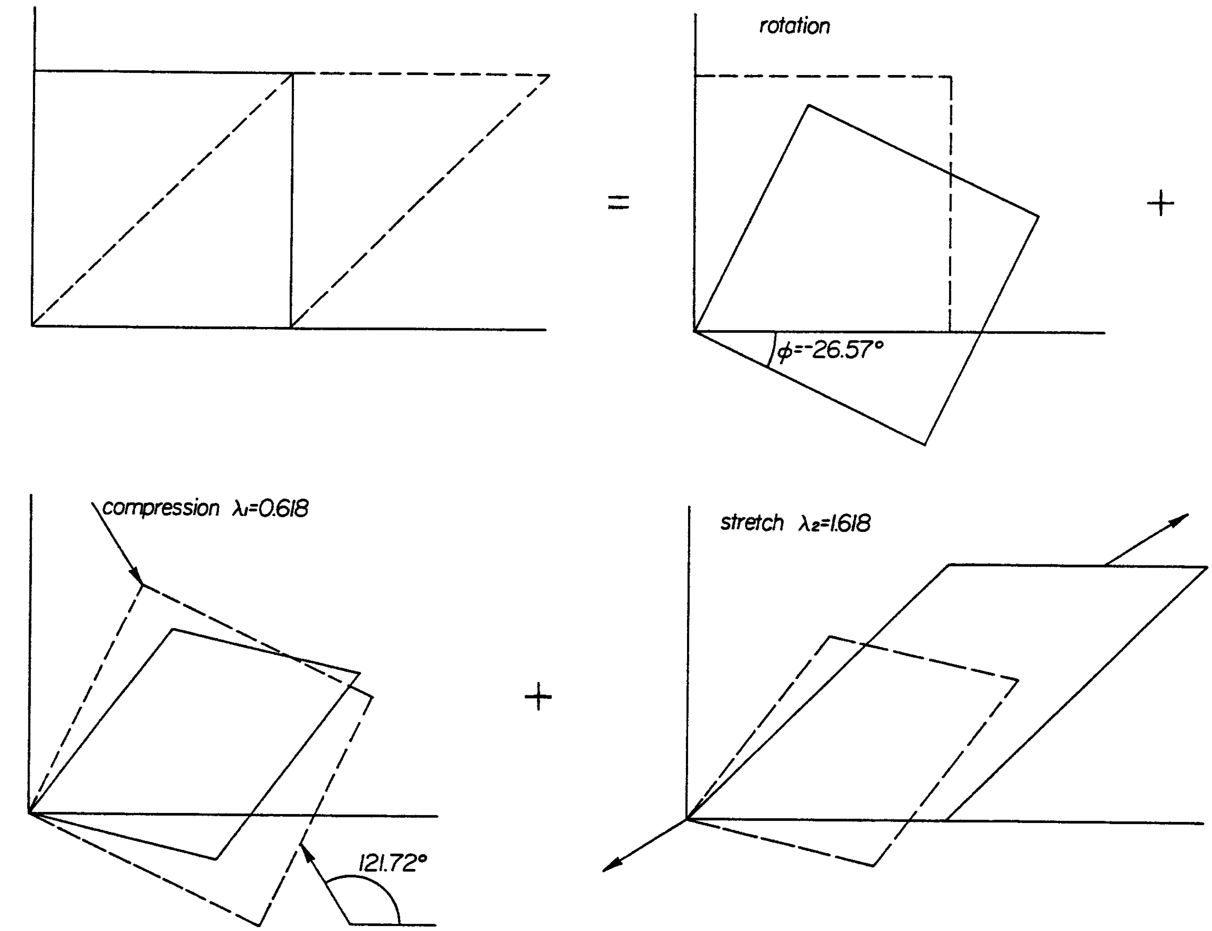


Fig. 4.62

and the orthonormal x_1, x_2, x_3 is mapped into the orthogonal $x_1' x_2' x_3'$; the eigenvectors of $(AA^T)^{-1}$ or AA^T. If λ_j^{-2} is such that $(AA^T)^{-1}x_j' = \lambda_j^{-2}x_j'$, then ${x_j'}^T x_j' = \lambda_j^2$. In sum

Theorem 4.39. *Every nonsingular linear mapping of R^3 is the product of three orthogonal stretchings (positive or negative) and a rotation. Matrix $A = A(3 \times 3)$ can be written as $A = TQ$ where $T = T^T$ and where Q is a rotation.*

Now let A be symmetric, $A = A^T$. There exists an orthonormal system u_1, u_2, u_3 in R^3 that is mapped by $u' = Au$ into the orthogonal u_1', u_2', u_3' such that $A^2 u_j' = \lambda_j^2 u_j'$, or $A^2 u_j = \lambda_j^2 u_j$, and $A^2(u_j - u_j') = \lambda_j^2(u_j - u_j')$. If the principal axes of the ellipsoid are unique, then u_j is colinear with u_j' and we have that $u_j' = \mu_j u_j$ or $Au_j = \mu_j u_j$. No rotation of the mapped u_1, u_2, u_3 takes place, $A = T$, and we have

Theorem 4.40. *Every nonsingular symmetric mapping of R^3 is the product of three orthogonal stretchings.*

Any symmetric $A = A(3 \times 3)$ matrix, not necessarily positive definite, has three real eigenvalues and three orthonormal eigenvectors. If the eigenvalues are distinct, the eigenvectors are unique up to sense.

Computation of the eigenvalues and eigenvectors is related to the minimization of a function of several variables and is more involved than the solution of a system of linear equations. The algebraic eigenproblem is theoretically interesting and physically significant. It constitutes the better part of theoretical and computational linear algebra and we shall return to the subject in Chapters 6 and 8 for a thorough discussion.

Meanwhile we consider the two dimensional eigenproblem det $(A - \lambda I) = 0$ that can be given a simple geometrical meaning. Writing $A = TQ$, where T is the symmetric stretch matrix and Q a rotation, we determine that

$$\begin{aligned} \det(A - \lambda I) &= \det(TQ - \lambda Q^T Q) = \det(T - \lambda Q^T) \\ &= \det(T - \lambda Q) \end{aligned} \tag{4.208}$$

and det $(A - \lambda I) = 0$ if and only if vector $u \neq o$ exists so that $Tu = \lambda Qu$.

Mapping $u' = Tu$ stretches and rotates unit vector u. Let $T = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T$ be with positive λ's. Then, as in Fig. 63, rotation is zero when u is colinear with u_1 or u_2, and is

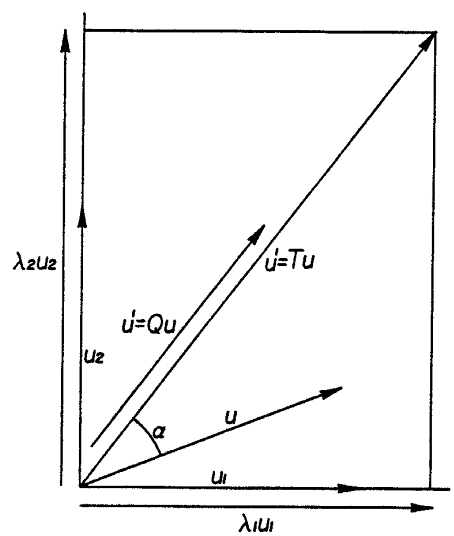


Fig. 4.63

maximal for some intermediate direction. If λ_1 is negative then Tu_1 rotates u_1 by 180^o. The isometry $u' = Qu$ rotates u through a *constant* angle α. Colinearity of Tu and Qu, that is,

$Tu = \lambda Qu$, can be realized if and only if the angle of rotation by Qu is less than or equal to the maximum rotation by Tu. And we see that if there is one solution to $Tu = \lambda Qu$, then there are two.

Consider for example

$$A = \begin{bmatrix} 1 & \\ & 2 \end{bmatrix} \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = TQ. \tag{4.209}$$

Matrix T rotates and elongates unit matrix $u = [\cos\phi,\ \sin\phi]^T$. Figure 64 shows the angle of rotation θ between u and its image $u' = Tu = [\cos\phi,\ 2\sin\phi]$

$$\cos\theta = \frac{u^T u'}{u'^T u'^T} \tag{4.210}$$

as a function of $0 \le \phi \le \pi/2$.

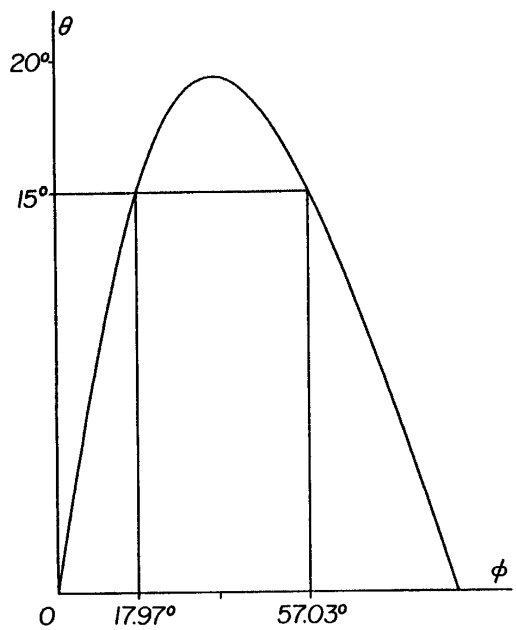


Fig. 4.64

As expected, $\theta = 0$ at $\phi = 0$ and $\phi = \pi/2$, and it reaches a maximum of $\theta = 19.47^o$ at $\phi = 35.28^o$. Hence matrix A has two (real) eigenvalues if $\alpha \le 19.47^o$, and no real eigenvalue if $\alpha > 19.47^o$. Take $\alpha = 15^o$ and look again at Fig. 64. For this α matrix A has two noncolinear eigenvectors $u_1 = [\cos\phi_1\ \ \sin\phi_1]^T$ and $u_2 = [\cos\phi_2\ \ \sin\phi_2]^T$ at $\phi_1 = -32.97^o$ and $\phi_2 = -72.03^o$, with corresponding eigenvalues $\lambda_1 = 1.134$ and $\lambda_2 = 1.764$. The reader may verify that $Au_1 = \lambda_1 u_1$ and $Au_2 = \lambda_2 u_2$.

More generally, if

$$\det\left(\begin{bmatrix}\lambda_1 & \\ & \lambda_2\end{bmatrix} - \lambda\begin{bmatrix}\cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha\end{bmatrix}\right) = 0 \tag{4.211}$$

written out as

$$\lambda^2 - \lambda(\lambda_1 + \lambda_2)\cos\alpha + \lambda_1\lambda_2 = 0$$

is with $\lambda_1\lambda_2 < 0$, that is if Tu can perform a half-turn, then the quadratic equation is soluble for any α and

$$A = \begin{bmatrix}\lambda_1 & \\ & \lambda_2\end{bmatrix}\begin{bmatrix}\cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha\end{bmatrix} = TQ \tag{4.212}$$

has two real eigenvalues for any α.

The three-dimensional picture is more involved, but we readily conclude that since in R^3 det $(A - \lambda I) = 0$ is a *cubic* equation in λ, one real solution to it always exists.

Exercises

4.10.1. Show that the three roots of $\det(R - \lambda I) = 0$, $R^TR = I$, $\det(R) = 1$, are $\lambda_1 = 1$, $\lambda_2 = \cos\theta + i\sin\theta$, $\lambda_3 = \cos\theta - i\sin\theta$.

4.10.2. Solve

$$det(A - \lambda I) = 0\ ,\ A = \begin{bmatrix}1 & -1 \\ -1 & 2\end{bmatrix}$$

and write the matrix as $A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T$ for orthonormal u_1, u_2.

4.10.3. Express

$$A = \begin{bmatrix}1 & 1 \\ & -1\end{bmatrix}$$

as $A = TQ$ in accordance with Theorem 4.38.

4.10.4. Let the matrix in transformation (4.129) be A,

$$A = \begin{bmatrix}1 & -\sigma_1\sin\phi \\ 0 & \cos\phi - \sigma_2\sin\phi\end{bmatrix}\ ,\ A^TA = \begin{bmatrix}1 & -\sigma_1\sin\phi \\ -\sigma_1\sin\phi & \sigma_1^2\sin^2\phi + (\cos\phi - \sigma_2\sin\phi)^2\end{bmatrix}.$$

Show that the two eigenvalues λ_1^2, λ_2^2 of A^TA are limited by

$$\lambda_1^2 + \lambda_2^2 - \lambda_1^2\lambda_2^2 \geq 1.$$

4.10.5. Prove that the 3×3 symmetric

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \lambda_3 u_3 u_3^T \ , \ u_i^T u_i = 1, u_i^T u_j = 0$$

is positive definite if and only if its three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ are positive.

4.10.6. Prove that for every $A = A^T$ there exists a symmetric orthogonal matrix $Q = Q^T, Q^2 = I$, so that AQ is positive definite.

4.10.7. Prove that if $B = -B^T$, then $Q = (I - B)(I + B)^{-1}$ is orthogonal. Show that $\det(Q) = 1$.

4.11 Vector and matrix norms

Motivated by geometrical considerations, we introduced in Section 4.3 the concept of vector magnitude or *norm* as the length of the directed segment the list of numbers represents. For vectors with more than three (real) components, for vectors in vector space R^n, the geometrical meaning of length is lost but the concept of size remains useful and is naturally extended. We take $\|a\| = \sqrt{a^T a}$ to be the norm of $a \in R^n$, and have for any $a \in R^n$ and $b \in R^n$,

Theorem 4.41.

1. $\|\alpha a\| > 0$ if $a \neq o$, and $\|a\| = 0$ if $a = o$,
2. $\|\alpha a\| = |\alpha| \, \|a\|$, and
3. $\|a + b\| \leq \|a\| + \|b\|$,
4. $|a^T b| \leq \|a\| \, \|b\|$.

Proof. Statements 1 and 2 are obvious. To prove the Cauchy-Schwarz inequality of statement 4 we use the fact that whatever α, a, and b,

$$(a + \alpha b)^T (a + \alpha b) \geq 0 \tag{4.213}$$

or

$$a^T a + 2\alpha a^T b + \alpha^2 b^T b \geq 0 \tag{4.214}$$

and conclude that since this quadratic form in α is non-negative its discriminant is not positive, or

$$(a^Tb)^2 - (a^Ta)(b^Tb) \leq 0. \tag{4.215}$$

Upon taking the positive square root of $(a^Tb)^2$ we obtain the desired inequality

$$|a^Tb| \leq \|a\| \, \|b\|. \tag{4.216}$$

To prove the triangular inequality of statement 3, we start with

$$\|a+b\|^2 = (a+b)^T(a+b) = \|a\|^2 + 2a^Ta + \|b\|^2 \tag{4.217}$$

and get

$$\|a+b\|^2 \leq \|a\|^2 + 2|a^Tb| + \|b\|^2. \tag{4.218}$$

Replacement of $|a^Tb|$ by the not less $\|a\|\|b\|$ results in

$$\|a+b\|^2 \leq \|a\|^2 + 2\|a\| \, \|b\| + \|b\|^2 = (\|a\| + \|b\|)^2 \tag{4.219}$$

and the triangular inequality is obtained by taking the positive square root on both sides. End of proof.

The next inequality is important enough to be put as

Corollary 4.42. *Let $a \in R^n$ and $b \in R^n$. Then*

$$\|a-b\| \geq | \, \|a\| - \|b\| \, |. \tag{4.220}$$

Proof. Writing the triangular inequality as

$$\|a+b-a\| \leq \|a\| + \|b-a\|, \text{ and } \|a-b+b\| \leq \|a-b\| + \|b\| \tag{4.221}$$

results in $\|b-a\| \geq \|b\| - \|a\|$ and $\|a-b\| \geq \|a\| - \|b\|$. Hence the absolute value. End of proof.

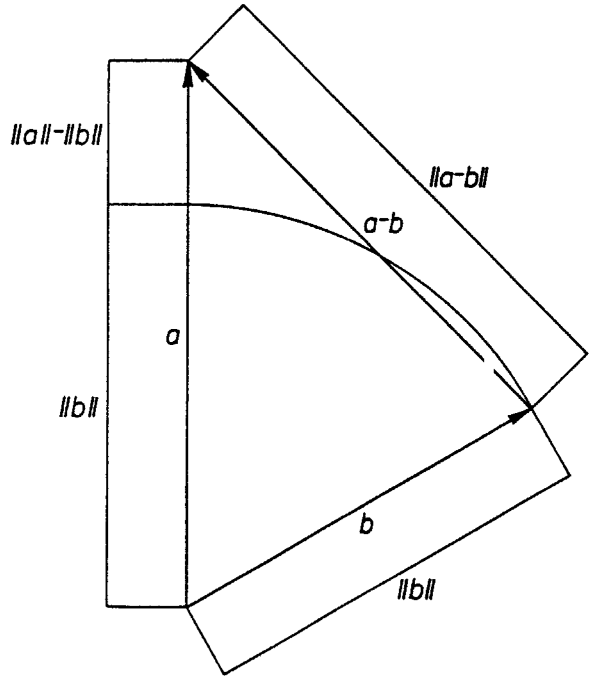


Fig. 4.65

For a geometrical interpretation to Corollary 4.42 see Fig. 65.

Norms introduce *analysis* into linear algebra in giving sense to the idea of *closeness* or *distance* between two vectors. If it is true for the three vectors a, x, x' that

$$\|a - x\| \leq \|a - x'\| \tag{4.222}$$

then vector x is closer, or nearer to vector a than vector x', and we concisely qualify

$$\lim_{n\to\infty} x_n = a \tag{4.223}$$

as

$$\lim_{n\to\infty} \|x_n - a\| = 0 \tag{4.224}$$

instead of saying that every component of x_n tends to the corresponding component of a as $n \to \infty$.

In R^n, $n > 3$, we may relinquish any vestige of geometry and seek other convenient or appropriate, analytically defined, norms that satisfy the basic requirements 1,2,3 of Theorem 4.41.

The choice

$$\|a\|_p = (|a_1|^p + |a_2|^p + \cdots + |a_n|^p)^{\frac{1}{p}} \quad p \geq 1 \tag{4.225}$$

often referred to as the ℓ_p norm, is common, not only for $p = 2$, but also for $p = 1$, for which it becomes

$$\|a\|_1 = |a_1| + |a_2| + \cdots + |a_n| \tag{4.226}$$

and $p = \infty$, for which it becomes

$$\|a\|_\infty = \max_i |a_i|. \tag{4.227}$$

Obviously $\|a\|_p > 0$ if $a \neq o$, $\|a\|_p = 0$ only if $a = o$, and $\|\alpha a\|_p = |\alpha|\|a\|_p$. Also

$$\|a + b\|_1 = \sum_{i=1}^{n} |a_i + b_i| \leq \sum_{i=1}^{n} |a_i| + |b_i| = \|a\|_1 + \|b\|_1 \tag{4.228}$$

and

$$\|a + b\|_\infty = \max_i |(a + b)_i| \leq \max_i |a_i| + \max_j |b_j| = \|a\|_\infty + \|b\|_\infty. \tag{4.229}$$

As an exercise the reader may prove that

$$\|a\|_\infty \leq \|a\|_2 \leq \|a\|_1 \tag{4.230}$$

and compare the inequalities with Fig. 66.

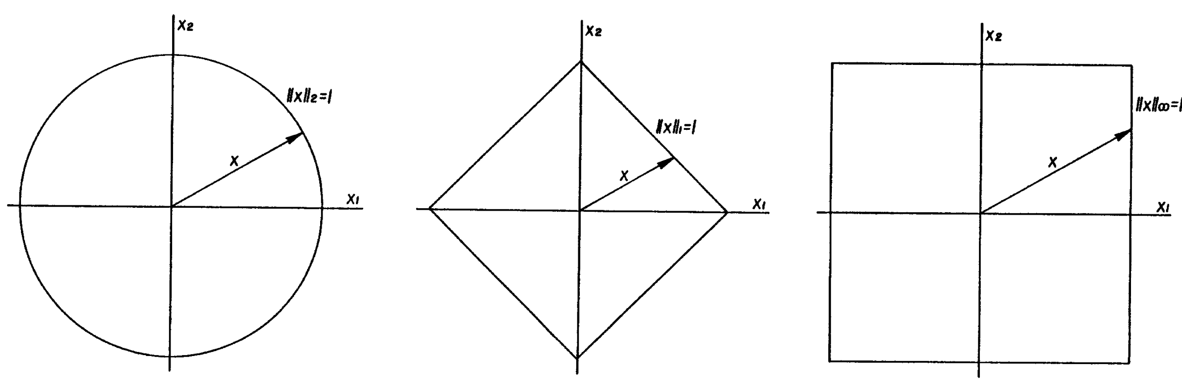


Fig. 4.66

From vector norms we turn to matrix norms and find it helpful to think of matrix A as being a linear operator that acts on x to produce $x' = Ax$. The size or magnitude of the operator is taken to be its largest action —the greatest relative magnification it gives x:

$$\|A\| = \max \|Ax\|, \ \|x\| = 1 \tag{4.231}$$

where max stands for the superior or utmost value. Since the linear transformation is *bounded* such value exists. When $\|Ax\|$ is a continuous function of the components of x, as for $\|Ax\|_2$, the least and largest values of $\|Ax\|_2$ under the condition that $\|x\|_2 = 1$ are true local minima and maxima.

Equivalently

$$\|A\| = \max \frac{\|Ax\|}{\|x\|} \quad x \neq o \tag{4.232}$$

from which the fundamental norm inequality

$$\|Ax\| \leq \|A\| \ \|x\| \tag{4.233}$$

quickly results.

Theorem 4.43. *If $\|A\| < 1$, then for any vector x, $\|A^n x\| \rightarrow 0$ as $n \rightarrow \infty$.*

Proof.

$$\|A^n x\| = \|AA^{n-1}x\| \leq \|A\| \, \|A^{n-1}x\| \leq \|A\|^n \|x\| \tag{4.234}$$

and as $n \rightarrow \infty$ $\|A\|^n \rightarrow 0$. End of proof.

Matrix A is said in this case to be *convergent*.

Matrix norm $\|A\|$ is defined through vector norm $\|Ax\|$, and the matrix norm depends on the choice of the vector norm so that we have $\|A\|_1$, $\|A\|_2$, $\|A\|_p$, and $\|A\|_\infty$. In case the subscript is omitted, reference is to any accepted norm.

Unlike vector norms, matrix norms are not always easily computed, let alone explicitly written in terms of the entries, but $\|A\|_\infty$ is an exception. From the definition

$$\|A\|_\infty = \max \|Ax\|_\infty, \ \|x\|_\infty = 1 \tag{4.235}$$

it follows that

$$\|A\|_\infty = \max_i \sum_j |A_{ij} x_j| \ , \ \|x\|_\infty = 1 \tag{4.236}$$

and the maximum is seen to be achieved with $x_j = \pm 1$ so as to make each term in the row sum positive, and

$$\|A\|_\infty = \max_i \sum_j |A_{ij}|. \tag{4.237}$$

We return to the ℓ_2 norm of Ax, for which

$$\|Ax\|_2^2 = x^T A^T Ax. \tag{4.238}$$

From the previous section we know that in R^3 A^TA has three non-negative eigenvalues $\lambda_1^2 \leq \lambda_2^2 \leq \lambda_3^2$, and

$$\|A\|_2 = |\lambda_3|. \tag{4.239}$$

Matrix A need not be square, but if A is singular, then some of the eigenvalues of A^TA are zero, but as long as $A \neq O$ at least one eigenvalue of A^TA is greater than zero.

The symmetric case is most interesting. When $A = A^T$, matrix A itself has three (real) eigenvalues which are the principal stretches of $x' = Ax$.

Theorem 4.44. *Let the three eigenvalues of $A = A^T$ be arranged as $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$. Then $\|A\|_2 = |\lambda_3|$ and $\|A^{-1}\|_2 = |\lambda_1^{-1}|$.*

Proof. If $Ax_1 = \lambda_1 x_1$ then $A^{-1}x_1 = \lambda_1^{-1}x_1$, and hence the eigenvalue magnitudes of A^{-1} are $|\lambda_3^{-1}| \leq |\lambda_2^{-1}| \leq |\lambda_1^{-1}|$. End of proof.

We have limited our arguments of $\|A\|_2$ to R^3 in the geometrical spirit of this chapter, but the incisive reader should see the ready extensions to R^n. How to actually compute the maximum of $\|Ax\|_2/\|x\|_2$ over all $x \neq o$ must wait until Chapter 8.

Computation of $\|A\|_2$ is costly, but it has a remarkable property.

Theorem 4.45. *If $A = A^T$, then $\|A\|_2 \leq \|A\|$.*

Proof. Let the eigenvalues of A be ordered as $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$, and write $Ax_3 = \lambda_3 x_3$. Then

$$\|Ax_3\| = |\lambda_3| \, \|x_3\| = \|A\|_2 \tag{4.240}$$

if we choose $\|x_3\| = 1$. Since $\|Ax_3\| \leq \|A\|$ the inequality of the theorem is established. End of proof.

The *spectral norm* $\|A\|_2$ of $A = A^T$ is the least of all norms.

Theorem 4.45 asserts that if $A = A^T$, then a *necessary and sufficient* condition for $A^n \rightarrow O$ as $n \rightarrow \infty$, is that $\max |\lambda_j| < 1$. For any other norm the condition is sufficient but not necessary. If A is not symmetric, then the condition that $\|A\|_2 < 1$ is again sufficient but not necessary. We define therefore the *spectral radius* of A as the function ρ of A such that in any norm

$$\lim_{k\to\infty} \|(\frac{1}{\rho}A)^k\| = 1 \tag{4.241}$$

or

$$\rho = \lim_{k\to\infty} \|A^k\|^{\frac{1}{k}},\ \rho = \lim_{k\to\infty} \frac{\|A^{k+1}\|}{\|A^k\|}. \tag{4.242}$$

Then $\rho \leq \max|\lambda_j|$, λ_j^2 being the eigenvalues of A^TA, with equality holding if $A = A^T$. The purpose of factor $1/\rho$ is to make sure that as $k \to \infty, (1/\rho A)^k$ tends to a nonzero matrix of finite entries.

Theorem 4.46. *For matrix* A:

1. $\|A\| > 0$ *if* $A \neq O$, $\|A\| = 0$ *if* $A = O$.
2. $\|\alpha A\| = |\alpha|\,\|A\|$.
3. $\|AB\| \leq \|A\|\,\|B\|,\ \|A^2\| \leq \|A\|^2$.
4. $\|A+B\| \leq \|A\| + \|B\|$.
5. $\|A-B\| \geq |\,\|A\| - \|B\|\,|$.
6. $\|A^{-1}\| \geq \|A\|^{-1}$.

Proof.

1. Since $x \neq o$, $\|Ax\| = 0$ only if $A = O$.
2. $\|\alpha Ax\| = |\alpha|\ \|Ax\|$.
3. Let $x,\ \|x\| = 1$, be such that $\|AB\| = \|ABx\|$.

Then $\|ABx\| \leq \|A\|\,\|Bx\| \leq \|A\|\,\|B\|$.

4. Let $x,\ \|x\| = 1$, be such that $\|A+B\| = \|(A+B)x\|$. Then $\|(A+B)x\| \leq \|Ax\| + \|Bx\| \leq \|A\| + \|B\|$.
5. Same as Corollary 4.22.
6. Since $AA^{-1} = I$, and $\|I\| = 1$, it results that $\|AA^{-1}\| = 1$ and $\|A\|\,\|A^{-1}\| \geq \|AA^{-1}\| = 1$.

End of proof.

As an example of how the concept of norm enters analytical considerations, particularly the notion of *convergence*, into linear algebra consider the following. Suppose that matrix A can be split into $A = I - E$ with $\|E\| < 1$. Writing

$$A^{-1} = (I - E)^{-1} = I + E + E^2 + \cdots + E^k + R_k \tag{4.243}$$

where R_k is a residual matrix, leads to

$$I = (I - E)(I + E + E^2 + \cdots + E^k + R_k) \tag{4.244}$$

and the residual matrix is obtained as

$$R_k = (I - E)^{-1}E^{k+1} = A^{-1}E^{k+1}. \tag{4.245}$$

In norms

$$\|R_k\| = \|A^{-1}E^{k+1}\| \leq \|A^{-1}\| \; \|E^{k+1}\| \leq \|A^{-1}\| \; \|E\|^{k+1} \tag{4.246}$$

and $\|R_k\| \to 0$ as $k \to \infty$.

Otherwise stated, if

$$B_k = I + E + E^2 + \cdots + E^k \tag{4.247}$$

then

$$\|A^{-1} - B_k\| = \|R_k\| \tag{4.248}$$

and $\|R_k\| \to 0$ as $k \to \infty$. The *iterative* algorithm

$$B_{k+1} = I + EB_k \; , \; B_0 = I \tag{4.249}$$

generates a sequence of matrices $B_0, B_1, \ldots, B_k$ that *converges* to A^{-1}.

All this opens up fascinating prospects for the iterative solution of the linear system $Ax = f$. If we succeed in writing $A = I - E$ with $\|E\| < 1$, then the iterative algorithm

$$x_{k+1} = f + Ex_k, \; x_0 = f \tag{4.250}$$

generates the vector sequence $x_0, x_1, \ldots, x_k$ that is assured to converge to the solution x of $Ax = f$. In other words, for any preset tolerance ϵ, index k exists so that $\|x - x_k\| \leq \epsilon$.

With the *residual vector* $r_k = f - Ax_k$, and by writing $E = I - A$, the iterative algorithm to solve $Ax = f$ becomes simply

$$x_{k+1} = x_k + r_k \tag{4.251}$$

with x_0 being arbitrary. Subtraction of x from both sides of this equation results in

$$e_{k+1} = Ee_k \tag{4.252}$$

where $e_k = x_k - x$ is the error vector of the kth iteration. Iterative algorithms that exhibit such an error relationship are said to converge *linearly*.

For a feel as to how effective the iterative scheme for the solution of $Ax = f$ can be we undertake the solution of

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3/2 \end{bmatrix}, \ x = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \tag{4.253}$$

for which we compute $\|E\|_2 = \|E\|_\infty = 1/2$. Starting with $x_0 = [1\ 0]^T$ we iteratively generate the sequence

$$x_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \ x_2 = \begin{bmatrix} 1 \\ 3/2 \end{bmatrix}, \ x_3 = \begin{bmatrix} 3/4 \\ 2 \end{bmatrix}, \ x_4 = \begin{bmatrix} 1 \\ 15/8 \end{bmatrix}, \ x_6 = \begin{bmatrix} 15/16 \\ 2 \end{bmatrix} \tag{4.254}$$

and decide to pause.

For a larger system with

$$A = \begin{bmatrix} 1 & -1/2 & & & & \\ -1/2 & 1 & -1/2 & & & \\ & -1/2 & 1 & -1/2 & & \\ & & -1/2 & 1 & -1/2 & \\ & & & -1/2 & 1 & -1/2 \\ & & & & -1/2 & 1 \end{bmatrix}, \ E = \frac{1}{2} \begin{bmatrix} & 1 & & & & \\ 1 & & 1 & & & \\ & 1 & & 1 & & \\ & & 1 & & 1 & \\ & & & 1 & & 1 \\ & & & & 1 & \end{bmatrix} \tag{4.255}$$

we readily establish that $\|E\|_\infty = 1$, but it is still possible that $\|E\|_2 < 1$. We compute

$$E^2 = \frac{1}{4} \begin{bmatrix} 1 & & 1 & & & \\ & 2 & & 1 & & \\ 1 & & 2 & & 1 & \\ & 1 & & 2 & & 1 \\ & & 1 & & 2 & \\ & & & 1 & & 1 \end{bmatrix}, \ E^3 = \frac{1}{8} \begin{bmatrix} & 2 & & 1 & & \\ 2 & & 3 & & 1 & \\ & 3 & & 3 & & 1 \\ 1 & & 3 & & 3 & \\ & 1 & & 3 & & 2 \\ & & 1 & & 2 & \end{bmatrix}$$

$$
E^4 = \frac{1}{16}\begin{bmatrix} 2 & & 3 & & 1 & \\ & 5 & & 4 & & 1 \\ 3 & & 6 & & 4 & \\ & 4 & & 6 & & 3 \\ 1 & & 4 & & 5 & \\ & 1 & & 3 & & 2 \end{bmatrix} , \; E^5 = \frac{1}{32}\begin{bmatrix} & 5 & & 4 & & 1 \\ 5 & & 9 & & 5 & \\ & 9 & & 10 & & 4 \\ 4 & & 10 & & 9 & \\ & 5 & & 9 & & 5 \\ 1 & & 4 & & 5 & \end{bmatrix}
$$

$$
E^6 = \frac{1}{64}\begin{bmatrix} 5 & & 9 & & 5 & \\ & 14 & & 14 & & 5 \\ 9 & & 19 & & 14 & \\ & 14 & & 19 & & 9 \\ 5 & & 14 & & 14 & \\ & 5 & & 9 & & 5 \end{bmatrix} , \; E^7 = \frac{1}{128}\begin{bmatrix} & 14 & & 14 & & 5 \\ 14 & & 28 & & 19 & \\ & 28 & & 33 & & 14 \\ 14 & & 33 & & 28 & \\ & 19 & & 28 & & 14 \\ 5 & & 14 & & 14 & \end{bmatrix} \tag{4.256}
$$

and obtain the sequence $\|E\|_\infty = 1$, $\|E^2\|_\infty^{\frac{1}{2}} = 0.866$, $\|E^3\|_\infty^{\frac{1}{3}} = 0.956$, $\|E^4\|_\infty^{\frac{1}{4}} = 0.949$, $\|E^5\|_\infty^{\frac{1}{5}} = 0.936$, $\|E^6\|_\infty^{\frac{1}{6}} = 0.932$, $\|E^7\|_\infty^{\frac{1}{7}} = 0.926, \ldots, \|E^{14}\|_\infty^{\frac{1}{14}} = 0.914$, that converges to $\|E\| = 0.901$. Iterative scheme (4.248) converges with this E, albeit more slowly.

In fact, for $E = E(n \times n)$ at large n, $\|E\| = 1 - 4/n^2$, nearing to 1 as n increases. The iterative method becomes slower and less efficient exactly when its application becomes most desirable.

The practical lesson drawn from this limited example is that even though iterative methods are theoretically intriguing, they can be a practical disappointment. Repetitive correction of an arbitrary initial guess has great computational appeal but is costly and raises troublesome issues on start, finish and error estimates.

What attracts us to iterative methods is the central need of computational linear algebra to deal with vast *sparse* systems of linear equations. Iterative methods do not require matrix A explicitly, only Ax_k, which for a sparse A can be done very efficiently regardless of the sparseness pattern. Considerable effort has gone and is still going into the search and analysis of efficient iterative methods for the solution of the large linear system. As matters now stand no iterative method exists that *generally* outperforms Gauss elimination, certainly not for small dense matrices. The only exception to this is the method of conjugate gradients, to be discussed in Chapter 7, that for special large sparse systems does raise a serious challenge to the Gauss elimination algorithm.

Exercises

4.11.1. Under what conditions on vectors a and b is inequality $\|a-b\| \geq |\|a\|-\|b\||$ best, and under what conditions is it worst?

4.11.2. Prove that if $\|a\| = \sqrt{a^Ta}$, then

$$\|a+b\|^2 + \|a-b\|^2 = 2(\|a\|^2 + \|b\|^2)$$

and

$$a^Tb = \frac{1}{4}\|a+b\|^2 - \frac{1}{4}\|a-b\|^2.$$

4.11.3. Prove that for any four vectors p, q, r, s

$$\|p-r\| + \|q-s\| \geq |\ \|p-q\| - \|r-s\||\ .$$

4.11.4. Prove the Hölder inequality

$$|a^Tb| \leq \|a\|_p\|b\|_q \ \ \frac{1}{p} + \frac{1}{q} = 1,\ p > 1, q > 1.$$

4.11.5. Prove that if u and v are unit vectors and $\eta \geq 0$, then

$$\|u-v\| \leq 2\|u-\eta v\|.$$

4.11.6. Prove that

$$\|a-b\| \geq \frac{1}{4}(\|a\| + \|b\|)\|u-v\|\ ,\ u = a/\|a\|,\ v = b/\|b\|.$$

4.11.7. For $A = a^Tb$, write $\|A\|_\infty$.

4.11.8. Can $\|A\| = \max_{ij}|A_{ij}|$ serve as a norm for matrix A? Show that, in fact,

$$\|A\| > 0 \text{ if } A \neq O,\ \|A\| = 0 \text{ if } A = O,\ \|\alpha A\| = |\alpha|\,\|A\|,\ \|A+B\| < \|A\| + \|B\|,$$

but that generally $\|AB\|$ is not $<\|A\|\,\|B\|$.

4.11.9. Let $A = [a_1 \ a_2 \ a_3]$. Show that the *Frobenius* norm of matrix A,

$$\|A\|_F = (\sum_{i,j} A_{ij}^2)^{1/2} = (\|a_1\|^2 + \|a_2\|^2 + \|a_3\|^2)^{1/2}$$

satisfies the basic norm provisions $\|A\|_F > 0$ if $A \neq O$, $\|A\|_F = 0$ if $A = O$, $\|\alpha A\|_F = |\alpha| \, \|A\|_F$, $\|A+B\|_F \leq \|A\|_F + \|B\|_F$, and $\|AB\|_F \leq \|A\|_F \|B\|_F$. Notice that $\|x\|_F = \sqrt{x^T x}$.

4.11.10. Prove that if $\|U\| < 1$, then $I - U$ is nonsingular and

$$\|(I-U)^{-1}\| \leq \frac{1}{1 - \|U\|}.$$

Hint: use the fact that $(I-U)^{-1} = I + U(I-U)^{-1}$.

4.11.11. Let $U = I - A^{-1}B$. Show that if $\|U\| < 1$, then

$$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|U\|} \ .$$

4.11.12. Show that if $\|U\| < 1$, then

$$\|(I-U)^{-1}\| \geq 1 + \|U\|.$$

4.11.13. Let A be nonsingular. Prove that if $\|A^{-1}B\| < 1$, then $A + B$ is nonsingular.

4.11.14. The *condition number* of A is defined as

$$\kappa(A) = \|A\| \ \|A^{-1}\|.$$

Show that $\kappa(I) = 1, \ \kappa(A) \geq 1$, and that $\kappa(\alpha A) = \kappa(A), \ \alpha \neq 0$.

Also that if D is diagonal, then

$$\|D\|_2 = \|D\|_\infty = \max_i |D_{ii}|$$

and

$$\kappa_2(D) = \kappa_\infty(D) = \max_i |D_{ii}| / \min_j |D_{jj}|.$$

4.11.15. Prove that

$$\frac{\|B^{-1}-A^{-1}\|}{\|B^{-1}\|} \le \kappa(A)\frac{\|B-A\|}{\|A\|}.$$

Hint: start with $A^{-1}-B^{-1}=A^{-1}(I-AB^{-1})$.

4.11.16. If $Ax=f$ and $Ax'=f'$, show that

$$\frac{\|x-x'\|}{\|x\|} \le \kappa(A)\frac{\|f-f'\|}{\|f\|}.$$

Hint: start with $x-x'=A^{-1}(f-f')$.

4.11.17. If $AB=I+E$, show that

$$\frac{\|A^{-1}-B\|}{\|A^{-1}\|} \le \|E\|,\ \|AB-I\|=\|E\|,\ \|BA-I\| \le \kappa(A)\|E\|.$$

4.11.18. If $Ax=f$ and $A'x'=f$, show that

$$\frac{\|x-x'\|}{\|x\|} \le \frac{\epsilon\kappa(A)}{1-\epsilon\kappa(A)},\quad \epsilon=\frac{\|A-A'\|}{\|A\|}$$

if $\epsilon\kappa(A)<1$.

4.11.19. Let $Ax=f$ and $Ax'=f+r$ be with the symmetric matrix

$$A=\lambda_1u_1u_1^T+\lambda_2u_2u_2^T+\lambda_3u_3u_3^T \quad 0<|\lambda_1|\le|\lambda_2|\le|\lambda_3|$$

for orthonormal u_1,u_2,u_3. Show that if $r=\alpha_1u_1+\alpha_2u_2+\alpha_3u_3$, then

$$x'-x=A^{-1}r=\frac{\alpha_1}{\lambda_1}u_1+\frac{\alpha_2}{\lambda_2}u_2+\frac{\alpha_3}{\lambda_3}u_3.$$

4.11.20. Show that if $A=\lambda_1u_1u_1^T+\lambda_2u_2u_2^T+\lambda_3u_3u_3^T$ for the orthonormal u_1,u_2,u_3, and $\lambda_j>0$, then

$$A^{\frac{1}{2}}=\lambda_1^{\frac{1}{2}}u_1u_1^T+\lambda_2^{\frac{1}{2}}u_2u_2^T+\lambda_3^{\frac{1}{2}}u_3u_3^T,\quad A^{\frac{1}{2}}A^{\frac{1}{2}}=A.$$

4.11.21. Implicit in the vector norm is the requirement that $\|x\|$ is finite as long as components x_i of x are finite. Prove that if $\max_i|x_i|\to\infty$, then in any norm $\|x\|\to\infty$. Also, that if entries A_{ij} of matrix A are all finite, then so is $\|A\|$, and that if $\|A\|$ is finite, then so are all A_{ij}.

4.11.22. Use the fact that $\|A^k\| \leq \|A\|^k$ to prove that the spectral radius ρ of A, as defined in eq.(4.241), is such that $\rho \leq \|A\|$.

4.11.23. Give a thorough analytical consideration to eqs.(4.241), (4.242) and compute a good approximation to the spectral radius of

$$A = \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} = 2 \begin{bmatrix} \frac{1}{2} & -1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

using both the $\|A\|_F$ and the $\|A\|_\infty$ norms of A. The exact value is $\rho = \sqrt{3}$. What is $\lim(1/\rho A)^k,\ k \rightarrow \infty$.

Explain why the result is independent of the norm used.

Hint: say the two spectral radii $\rho_2 > \rho_1$ are obtained by two different norms

$$\lim_{k\rightarrow\infty} \|(\frac{1}{\rho_1}A)^k\|_1 = 1, \quad \lim_{k\rightarrow\infty} \|(\frac{1}{\rho_2}A)^k\|_2 = 1.$$

This essentially means that for a very high value of k,

$$(\frac{1}{\rho_1}A)^k = B_1 \neq O, \quad (\frac{1}{\rho_2}A)^k = B_2 \neq O$$

with finite B_1 and B_2 that are independent of k. Or

$$B_1 = (\rho_2/\rho_1)^k B_2, \quad \rho_2/\rho_1 > 1.$$

Answers

Section 4.3

4.3.1. $\alpha = \pm 1$.

4.3.3. Vectors not coplanar.

4.3.4. $\alpha = 2\gamma, \beta = 3\gamma, \gamma = \gamma$.

4.3.5. $\alpha = 3$.

4.3.6. $\alpha = 2, \beta = 3$.

Section 4.4

4.4.1. $x^T a = a^T x = 0,\ x = \alpha[1\ 1]^T$ for arbitrary α.

4.4.2. $x = \pm(\sqrt{2}/2)[1\ 1]^T$.

4.4.3. $x^T a = x^T b = 0, x = o$.

4.4.4. $x^T a = 0, x = \alpha[1\ 1\ 0]^T + \beta[-2\ 0\ 1]$ for arbitrary α and β.

4.4.5. $x^T a = x^T b = 0, x = \alpha[1\ 0\ \ -1]^T$ for arbitrary α.

4.4.6. $\cos\phi = -\sqrt{6}/3, \phi = 144.74^o$.

4.4.7. $x = \alpha a + \beta b, x^T c = 0, x = \gamma[-1\ 2\ \ -7]^T$.

4.4.8. $x = \alpha a + \beta b, x^T c = \sqrt{x^T x}\sqrt{c^T c}\cos(60^o)$. Since the condition is on angle we may assume first that $x^T x = 1$. Then $6\alpha^2 + 5\beta^2 = 1, \alpha = 3/14, \beta = \pm(1/7)\sqrt{71/10}$.

Section 4.6

4.6.1. The nullspace of $A = [a_1\ a_2\ a_3]$ consists of all vectors $x = \alpha[1\ \ -1\ 1]^T$ for arbitrary α from which we conclude that the columns of A are coplanar. In fact, $a_1 - a_2 + a_3 = o$. The range of the transformation consists of all vectors $x' = x_1 a_1 + x_2 a_2 + x_3 a_3$ for arbitrary scalar variables x_1, x_2, x_3. But since $a_1 - a_2 + a_3 = o$ the range is only two dimensional, $x = \alpha_1 a_1 + \alpha_2 a_2$ where $\alpha_1 = x_1 - x_3, \alpha_2 = x_2 + x_3$.

4.6.3. $x = [x_1 \ x_2]^T, x_2 = \pm 2x_1$.

4.6.4. Does a linear mapping $x' = Ax$ of R^2 onto itself exist by which the orthonormal x_1, x_2 are mapped such that $Ax_1 = \lambda_1 x_2,\ Ax_2 = \lambda_2 x_1,\ \lambda_1 \neq \lambda_2$? Hint: $x_1^T A x_1 = x_2^T A x_2 = 0$.

Section 4.7

4.7.4. $\alpha = 1/3$.

Section 4.8

4.8.1.

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

4.8.7.

$$x'_1 = \frac{2x_1 + x_2}{4x_1 + 4x_2 - 2},\ x'_2 = \frac{x_1 + 2x_2}{4x_1 + 4x_2 - 2}\ .$$