INTRODUCTION

The word "geometry" derives from Greek, meaning "earth measurement." Geometry was originally the mathematics describing the shapes of objects and their spatial relationships. Simple geometrical notions and ideas were known to ancient Babylonians and Egyptians 4000 years ago. Starting approximately 2500 years ago, the ancient Greeks developed fundamental geometrical ideas, including some relatively rigorous proofs based on logical reasoning. Dating from this era is Euclid's Elements, which introduced the basis for the axiomatic method and summarizes the knowledge at that time.

Prior to the sixteenth century, geometry and algebra were treated as independent subjects. The notion of combining the two was introduced in 1631 by René Descartes (1596–1650). This led to the field of analytic geometry, which permits the investigation of geometric questions using analytical methods. This area was extensively investigated in the eighteenth century, in particular by Leonhard Euler (1707–1783) and Gaspard Monge (1746–1818). Toward the end of the eighteenth century the use of calculus resulted in the beginnings of differential geometry, studied by Christian Gauss (1777–1855) and others. The introduction by Bernhard Riemann (1826–1866) of the theory of algebraic functions initiated the field of algebraic geometry. In parallel with these developments, the synthetic approach to geometry was extended by Victor Poncelet (1788–1867), who formulated postulates for projective geometry. In the past century and a half, the work of David Hilbert (1862–1943) and others has led to an extension of the scope of geometry to include the study of geometrical relationships between abstract quantities.

This article presents material concerning analytical, differential, projective, and algebraic geometry. The choice of topics and their depth of coverage were dictated primarily by consideration of their importance in applied physics and by limitations of space. In particular, the reader is warned that the weighting assigned to the topics discussed is uncorrelated with their present importance as mathematical fields of research. The treatment is not mathematically rigorous, but introduces sufficient mathematical terminology to make basic textbooks in the subject accessible. Some
of these are listed in the references at the end of the article.

1. ANALYTIC GEOMETRY

The underlying concepts of analytic geometry are the simple geometric elements: points, lines and curves, planes and surfaces, and extensions to higher dimensions. The fundamental method is the use of coordinates to convert geometrical questions into algebraic ones. This is called the "method of coordinates."

To illustrate the basic notion, consider a straight line \( l \). Following the method of coordinates, select one point \( O \) on \( l \) as the origin. This separates \( l \) into two halves. Call one half positive, the other negative. Any point \( P \) on the line can then be labeled by a real number, given by the distance \( OP \) for the positive half and by the negative of the distance \( OP \) for the negative half. There is thus a unique real number \( x \) assigned to every point \( P \) on \( l \), called the Cartesian coordinate of \( P \). Geometrical questions about the line can now be transcribed into analytical ones involving \( x \).

1.1 Plane Analytic Geometry

In two dimensions, basic geometric entities include points, lines, and planes. For a plane \( \pi \) the method of coordinates provides to each point \( P \) an assignment of two real numbers, obtained as follows. Take two straight lines in the plane, and attribute Cartesian coordinates to each line as described above. For simplicity, the lines will be assumed perpendicular and intersecting at their origins. These lines are called rectangular coordinate axes, and the Cartesian coordinates of the first are called abscissae while those of the second are called ordinates. The lines themselves are also referred to as the abscissa and the ordinate. The location of a point \( P \) on \( \pi \) is then specified uniquely by two real numbers, written \((x, y)\). The number \( x \) is defined as the perpendicular distance to the first coordinate axis, while \( y \) is the distance to the second. Using these Cartesian coordinates, geometrical questions about points can be expressed in analytical terms. For example, a formula for the distance \( d \) between two points \( P \) and \( Q \) specified by the coordinates \((x_1, y_1)\) and \((x_2, y_2)\) is

\[
d = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}.
\]

(1)

Given an assignment of Cartesian coordinates on a plane \( \pi \), a curve segment \( s \) in the plane may be analytically specified by providing a set of paired real numbers \((x, y)\) assigned to all points on the curve. In many useful cases, \( s \) can be specified by an equation \( f(x,y)=0 \) between \( x \) and \( y \) that is satisfied by all points \( P \) on \( s \) but not by any other points on \( \pi \). For example, the equation \( x=0 \) describes the straight line consisting of all points having coordinates of the form \((0,y)\), i.e., the ordinate. The method of coordinates thus permits geometrical questions about \( s \) to be transcribed into analytical ones concerning \( f(x,y) \). For example, the set of points lying both on a curve \( f(x,y)=0 \) and on another curve \( g(x,y)=0 \) is specified by values \((x,y)\) satisfying both equations, which can in principle be found by analytical methods.

A simple example of a curve in the plane is a straight line \( l \). The slope \( m \) of \( l \) can be defined in terms of the coordinates \((x_1,y_1)\) and \((x_2,y_2)\) of any two distinct points on \( l \). Provided \( x_1 \neq x_2 \), the slope is given by

\[
m = \frac{y_1-y_2}{x_1-x_2}.
\]

(2)

The slope is zero for lines parallel to the abscissa, and is undefined (infinite) for lines parallel to the ordinate. A line \( l \) with given finite slope \( m \) is uniquely specified by its intersection point \((x,y) = (0,c)\) with the ordinate. The equation for \( l \) is

\[
y = mx + c.
\]

(3)

If \( l \) is parallel to the ordinate instead, it is determined by its intersection point \((x,y) = (a,0)\) with the abscissa, and its equation is simply \( x = a \).

The equation of a straight line \( l \) is also determined entirely by the coordinates \((x_1,y_1)\) and \((x_2,y_2)\) of any two distinct points on \( l \). It can be written

\[
\frac{y-y_1}{y_2-y_1} = \frac{x-x_1}{x_2-x_1}.
\]

(4)

Alternatively, a straight line can be viewed as the curve given by the most general equation linear in the coordinates \( x \) and \( y \):

\[
Ax + By + C = 0,
\]

(5)
where at least one of $A$ and $B$ is nonzero.

Analytical solutions to geometrical problems involving straight lines and points can be obtained using the above results. For example, the equation of the line $l_P$ that is perpendicular to a given line $l$ with equation $y=mx+c$ and that passes through the point $P$ on $l$ with abscissa $x_1$ is

$$y = -\frac{m^2+1}{m}x + x_1 + c.$$  \hspace{1cm} (6)

Another example is the expression for the perpendicular distance $d_p$ between a line $l$ with equation $y=mx+c$ and a point $P$ at $(a,b)$, which is

$$d_p = \frac{|b-ma-c|}{\sqrt{m^2+1}}.$$  \hspace{1cm} (7)

1.2 Conic Sections

An important curve is the circle, denoted by $S^1$, which can be viewed as the set of points in the plane that are equidistant from a specified fixed point. The fixed point $C$ is called the center of the circle, and the distance $r$ between the center and the points on the circle is called the radius. If the Cartesian coordinates of $C$ are $(h,k)$, then the equation of the circle is

$$(x-h)^2+(y-k)^2=r^2.$$  \hspace{1cm} (8)

The circle is a special case of a set of curves called conic sections or conics. These curves include ellipses, parabolas, and hyperbolas. Geometrically, the conics can be introduced as the curves obtained by slicing a right circular cone with a plane. Analytically, they can be viewed as the curves given by the most general expression quadratic in the coordinates $x$ and $y$:

$$Ax^2+2Bxy+Cy^2+Dx+Ey+F=0,$$  \hspace{1cm} (9)

where at least one of the coefficients $A$, $B$, $C$ is nonzero. These are called second-order curves. From this equation it follows that any five points lying on the conic specify it completely. The quantity $B^2-4AC$ is called the discriminant of the conic. If the discriminant is positive, the conic is a hyperbola; if negative, an ellipse; and if zero, a parabola.

A third definition, combining geometrical and analytical notions, is often useful. Consider a straight line $l$, a fixed point $F$ not on $l$, and another point $P$. Denote the distance between $P$ and $F$ by $d_P$ and the perpendicular distance between $P$ and $l$ by $d_l$. Then the conic sections are given by the set of points $P$ that obey the equation

$$d_P=\epsilon d_l,$$  \hspace{1cm} (10)

where $\epsilon>0$ is a constant real number called the eccentricity. The line $l$ is called the directrix and the point $F$ is called the focus. If $\epsilon>1$, the conic is a hyperbola. If $\epsilon=1$, the conic is a parabola. If $0<\epsilon<1$, the conic is an ellipse. The degenerate case $\epsilon=0$ gives a circle; in this case, the directrix is at infinity.

The equation determining a conic has a particularly simple form, called the canonical form, if the focus $F$ is chosen to lie on the abscissa and the directrix $l$ is chosen parallel to the ordinate. The canonical form depends on at most two real positive parameters $a$ and $b$, where $a>b$ is taken for convenience.

For a hyperbola, the canonical form is

$$\left(\frac{x^2}{a^2}\right)-\left(\frac{y^2}{b^2}\right)=1.$$  \hspace{1cm} (11)

The eccentricity of the hyperbola is $e=\sqrt{a^2+b^2}/a$. One focus is the point $(ae,0)$, and the corresponding directrix is the line $x=a/e$. There is a second focus at $(-ae,0)$ and a second directrix at $x=-a/e$. The hyperbola has two branches, each of which asymptotically approaches the lines $y=\pm bx/a$ as $|x|$ becomes large. The distance between the points where the hyperbola intersects the abscissa is $2a$. This is also the difference between the distances from the two foci to any given point on the hyperbola. If $a=b$, the hyperbola is called rectangular.

For a parabola, the canonical form is

$$y^2=4ax.$$  \hspace{1cm} (12)

The eccentricity is $e=1$, the focus is at $(a,0)$, and the directrix is the line $x=-a$.

For an ellipse, the canonical form is

$$\left(\frac{x^2}{a^2}\right)+\left(\frac{y^2}{b^2}\right)=1.$$  \hspace{1cm} (13)

The ellipse has eccentricity $e=\sqrt{a^2-b^2}/a$. There are again two foci, at $(\pm ae,0)$, and two directrices $x=\pm a/e$. The sum of the distances from the two foci to any given point on the ellipse is a constant, $2a$. The line between the points of intersection of the ellipse with the abscissa is called the major axis of the ellipse, and it has length $2a$. Similarly, the minor axis of the ellipse is given by the intersection points with the ordinate and has length $2b$. If
$a=b$ the equation reduces to that of a circle of radius $a$ centered at the origin.

### 1.3 Plane Trigonometry

Consider a point $P$ with coordinates $(x,y)$ lying on a circle of radius $r$ centered at the origin $O$. Denote by $X$ the point $(x,0)$. Call $\theta$ the angle $XOP$ between the line segments $OX$ and $OP$. The choice of a unit of measure for angles permits the assignment of a numerical value to $\theta$. One widely used unit is the degree, defined by the statement that there are 360 degrees in a circle. The SI unit is the radian, of which there are $2\pi$ in a circle.

Certain functions of the angle $\theta$, called trigonometric or circular functions, are of particular use in plane analytic geometry. The ratio $\sin \theta = y/r$ is called the sine of $\theta$ while $\cos \theta = x/r$ is called the cosine of $\theta$. The sine is odd in $\theta$ while the cosine is even, and both functions have period $\pi$ radians. They obey the relations

$$\sin^2 \theta + \cos^2 \theta = 1 \tag{14}$$

following from the Pythagorean theorem, and

$$\sin(\theta \pm \phi) = \sin \theta \cos \phi \pm \sin \phi \cos \theta, \tag{15}$$

$$\cos(\theta \pm \phi) = \cos \theta \cos \phi \mp \sin \theta \sin \phi. \tag{16}$$

The latter two equations are called addition formulas. Other, related functions of $\theta$ include the tangent $\tan \theta = y/x = \sin \theta / \cos \theta$, the cosecant $\csc \theta = r/y = 1/\sin \theta$, the secant $\sec \theta = r/x = 1/\cos \theta$, and the cotangent $\cot \theta = x/y = \cos \theta / \sin \theta$. From these definitions and the above equations many identities can be obtained. Inverse trigonometric functions can also be introduced; for example, if $x = \sin \theta$ then $\sin^{-1} x = \theta$.

Consider a triangle with angles $A$, $B$, $C$ and sides of length $a$, $b$, $c$, where by convention the side labeled by $a$ is opposite the vertex with angle $A$ and there are similar conventions for the other sides. A basic problem in plane trigonometry is to determine one of $a$, $b$, $c$, $A$, $B$, $C$ in terms of the others. This is called solving a triangle. The following relations hold: the law of sines,

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}, \tag{17}$$

the first law of cosines,

$$a = b \cos C + c \cos B; \tag{18}$$

and the second law of cosines,

$$a^2 = b^2 + c^2 - 2bc \cos A. \tag{19}$$

### 1.4 Curvilinear Coordinates

For certain geometrical problems, the analytical details of a calculation may be simplified if a non-Cartesian coordinate system is used. Consider two functions $u = u(x,y)$ and $v = v(x,y)$ of the Cartesian coordinates $x$ and $y$ on a plane $\pi$. Take the functions to be continuous and invertible, except perhaps at certain special points that require separate treatment. Any curve $u = c$ for some constant $c$ is called a coordinate curve, as is any curve $v = c$. A point $P$ on $\pi$ is uniquely specified by two real numbers $(u_1, v_1)$ that are the values of the constants determining the two coordinate curves passing through $P$. This construction generalizes the method of coordinates, and the functions $u$ and $v$ are called curvilinear coordinates. If the coordinate curves meet at right angles, the curvilinear coordinates are called orthogonal. All the analytical geometry described above using Cartesian coordinates can be rephrased using orthogonal curvilinear coordinates.

An important set of orthogonal curvilinear coordinates is generated by the equations

$$x = r \cos \theta, \quad y = r \sin \theta;$$

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}(y/x), \tag{20}$$

where $r > 0$ and $0 < \theta < 2\pi$. In this system, the coordinate curves consist of circles of varying radii centered at the origin and straight lines through the origin at varying angles with respect to the abscissa. The coordinates $(r, \theta)$ of a point $P$ are called plane polar coordinates. As an illustration of their use, consider the conic sections expressed in polar coordinates. In canonical form, with the origin of the polar coordinates placed at the focus at $(ae, 0)$, the equation for a conic section is

$$r = l/(1 + e \cos \theta), \tag{21}$$

where $l$ is called the latus rectum. It is given by $l = b^2/a$ for hyperbolas and ellipses and by $l = 2a$ for parabolas, and it represents the distance from the focus to the curve as measured along a straight line parallel to the ordinate. The quantity $l/e$ is the distance from the focus to the associated directrix.
The conic sections themselves can be used to generate systems of orthogonal curvilinear coordinates. For example, parabolic coordinates can be defined by

\[ x = \frac{1}{2}(u^2 - v^2), \quad y = uv, \]  

(22)

where \( u > 0 \). The coordinate curves are parabolas. Similarly, elliptic coordinates can be defined by

\[ x = a \cosh u \cos v, \quad y = a \sinh u \sin v, \]  

(23)

where \( a > 0 \) and \( 0 < v < 2\pi \). Here, the so-called hyperbolic functions \( \sinh u \) and \( \cosh u \) are defined by

\[ \sinh u = \frac{1}{2}(e^u - e^{-u}), \quad \cosh u = \frac{1}{2}(e^u + e^{-u}). \]  

(24)

The coordinate curves are ellipses and hyperbolas. Another common set is the system of bipolar coordinates, defined by

\[ x = \frac{a \sinh v}{\cosh v - \cos u}, \quad y = \frac{a \sin u}{\cosh v - \cos u}, \]  

(25)

with \( 0 < v < 2\pi \). The coordinate curves are sets of intersecting circles.

1.5 Solid Analytic Geometry

Solid analytic geometry involves the study of geometry in three dimensions rather than two. Many of the ideas of plane analytic geometry extend to three dimensions. For instance, the method of coordinates now provides an assignment of three real numbers \((x, y, z)\) to each point \(P\). A three-dimensional rectangular coordinate system can be introduced by taking three mutually perpendicular straight lines, each given Cartesian coordinates, to form the coordinate axes. The axes are called the abscissa, the ordinate, and the applicate. Each of the values \((x, y, z)\) is defined as the perpendicular distance to the corresponding axis.

A two-dimensional surface \(\sigma\) can now be specified by providing an equation \(f(x, y, z) = 0\) that is satisfied only by points on the surface. The method of coordinates thus converts geometrical questions about \(\sigma\) to analytical questions about \(f(x, y, z)\). Similarly, a curve \(s\) can be viewed as the intersection of two surfaces. If the surfaces are specified by the equations \(f(x, y, z) = 0\) and \(g(x, y, z) = 0\), \(s\) is given analytically by the set of points \((x, y, z)\) obeying both equations simultaneously.

By definition, a surface of the first order is given by the most general equation linear in \(x, y, z\):

\[ Ax + By + Cz + D = 0. \]  

(26)

If at least one of \(A, B, C\) is nonzero, this equation describes a plane. A straight line can be viewed as the intersection of two nonparallel planes and is therefore given analytically by two equations of this form. Just as in the two-dimensional case, the analytical formulation allows solutions to geometrical questions involving planes, lines, and points to be obtained. For example, the perpendicular distance \(d_p\) between a plane given by the above equation and a point \(P\) located at \((a, b, c)\) can be shown to be

\[ d_p = \frac{|Aa + Bb + Cc + D|}{\sqrt{A^2 + B^2 + C^2}}. \]  

(27)

As another example, two planes given by

\[ A_1x + B_1y + C_1z + D_1 = 0, \]
\[ A_2x + B_2y + C_2z + D_2 = 0 \]  

(28)

are parallel if and only if

\[ (A_1, B_1, C_1) = (cA_2, cB_2, cC_2) \]  

(29)

for some constant \(c\).

In analogy to the two-dimensional introduction of conics as curves obeying a quadratic expression in \(x\) and \(y\), a surface of the second order is defined to consist of points satisfying a quadratic expression in \(x, y, z\):

\[ Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0. \]  

(30)

Such surfaces are also called quadrics. An important example is the sphere, denoted by \(S^2\), which can be viewed as the set of points equidistant from a fixed point called the center. The distance from the center to any point on the sphere is called the radius. If the Cartesian coordinates of the center are \((h, k, l)\), the equation of a sphere of radius \(r\) is

\[ (x - h)^2 + (y - k)^2 + (z - l)^2 = r^2. \]  

(31)

The quadrics can be classified. Among the surfaces described are ellipsoids, hyperboloids, paraboloids, cylinders, and cones. Canonical forms of these surfaces are

\[ (x^2/a^2) + (y^2/b^2) + (z^2/c^2) = 1 \]  

(32)
for an ellipsoid;
\[ \left( \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} \right) = 1 \quad (33) \]
for a hyperboloid of one sheet;
\[ \left( \frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} \right) = 1 \quad (34) \]
for a hyperboloid of two sheets;
\[ \left( \frac{x^2}{a^2} + \frac{y^2}{b^2} \right) = 2z \quad (35) \]
for an elliptic paraboloid;
\[ \left( \frac{x^2}{a^2} - \frac{y^2}{b^2} \right) = 2z \quad (36) \]
for a hyperbolic paraboloid;
\[ \left( \frac{x^2}{a^2} + \frac{y^2}{b^2} \right) = 1 \quad (37) \]
for an elliptic cylinder;
\[ \left( \frac{x^2}{a^2} - \frac{y^2}{b^2} \right) = 1 \quad (38) \]
for a hyperbolic cylinder;
\[ \left( \frac{x^2}{a^2} \right) = 2z \quad (39) \]
for a parabolic cylinder; and
\[ \left( \frac{x^2}{a^2} \right) \pm \left( \frac{y^2}{b^2} \right) - \left( \frac{z^2}{c^2} \right) = 0 \quad (40) \]
for a cone. The parameters \( a, b, c \) are called the lengths of the principal axes of the quadric.

The notions of plane trigonometry also extend to three dimensions. A spherical triangle is defined as a portion of a spherical surface that is bounded by three arcs of great circles. Denote by \( A, B, C \) the angles generated by straight lines tangent to the great circles intersecting at the vertices, and call the lengths of the opposite sides \( a, b, c \) as for the planar case. The angles now add up to more than \( \pi \) radians, by an amount called the spherical excess \( E \):
\[ A + B + C = \pi + E. \quad (41) \]
The following relations hold for a spherical triangle: the law of sines,
\[ \frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c}; \quad (42) \]
the first law of cosines,
\[ \cos a = \cos b \cos c + \sin b \sin c \cos A; \quad (43) \]
and the second law of cosines,
\[ \cos A = -\cos B \cos C + \sin B \sin C \cos a. \quad (44) \]

Curvilinear coordinates can be introduced via three locally continuous invertible functions \( u(x,y,z), v(x,y,z), w(x,y,z) \), following the two-dimensional case. A coordinate surface is specified by setting any curvilinear coordinate \( u, v, \) or \( w \) equal to a constant. The coordinate curves are generated by the intersection of the coordinate surfaces, and the system is said to be orthogonal if the surfaces intersect at right angles. Many useful three-dimensional orthogonal curvilinear coordinate systems can be generated from families of quadrics. One particularly useful set is the system of spherical polar coordinates, defined by
\[ x = r \sin \theta \cos \phi, \]
\[ y = r \sin \theta \sin \phi, \]
\[ z = r \cos \theta, \]
where \( r > 0, 0 < \theta < \pi, \) and \( 0 < \phi < 2\pi \). The coordinate surfaces are spheres centered at the origin, right circular cones with axes along the applicate and vertices at the origin, and half-planes with the applicate as one edge. Other common coordinates are the cylindrical coordinates, given by
\[ x = r \cos \theta, \quad y = r \sin \theta, \quad z = z, \]
where \( r > 0 \) and \( 0 < \theta < 2\pi \). This system is generated from plane polar coordinates by translation along the applicate. The coordinate surfaces are right circular cylinders centered at the origin, half-planes with the applicate as one edge, and planes parallel to the plane of the abscissa and ordinate.

The notions of plane and solid analytic geometry can be extended to higher dimensions, too. A space can be defined in which the method of coordinates specifies a point by \( n \) real numbers \( (x^1, x^2, \ldots, x^n) \). This \( n \)-dimensional space, called Euclidean space, is denoted by the symbol \( R^n \). Using coordinates, geometrical questions in \( n \) dimensions can be converted to analytical ones involving functions of \( n \) variables. Surfaces of the first order are \((n-1)\)-dimensional hyperplanes, and surfaces of the second order, or quadric hypersurfaces, can be introduced. An example is the hypersphere of radius \( r \) in \( n \) dimensions, denoted by \( S^{n-1} \), which when centered at the origin satisfies the equation
\[ (x^1)^2 + (x^2)^2 + \cdots + (x^n)^2 = r^2. \quad (47) \]
The notion of curvilinear coordinates also extends to higher dimensions.

A one-dimensional curve \( s \) in \( n \) dimensions can be specified by \( n - 1 \) equations among the \( n \) coordinates \((x^1, \ldots, x^n)\). If \( s \) is continuous, its points can be labeled by a parameter \( t \) that is a real number. Any particular point can be specified by giving the values of the \( n \) coordinates \((x^1, \ldots, x^n)\). As \( t \) varies, so do the coordinates. This means that an alternative specification of \( s \) can be given in terms of the \( n \) expressions

\[ x^j = x^j(t), \quad j = 1, \ldots, n, \tag{48} \]

determining the \( n \) coordinates \((x^1, \ldots, x^n)\) as functions of \( t \). This is called the parametric representation of a curve. Similarly, the points of a continuous two-dimensional surface can be labeled by two real numbers \((t^1, t^2)\). The surface can be specified either in terms of \( n - 2 \) equations among the \( n \) coordinates \((x^1, \ldots, x^n)\) or in parametric form by the \( n \) equations

\[ x^j = x^j(t^1, t^2), \quad j = 1, \ldots, n. \tag{49} \]

A parametric representation can also be given for continuous surfaces of more than two dimensions.

1.6 Example: The Kepler Problem

An example of the appearance of analytic geometry in a physical problem occurs in the study of the classical motion of two bodies under a mutual inverse-square attractive force. Consider for definiteness two bodies of masses \( m_1 \) and \( m_2 \), each acted on by the gravitational field of the other and free to move in three dimensions. This is called the Kepler problem.

The first step is to introduce a convenient coordinate system. For simplicity, the origin can be placed on one mass. The problem can then be reduced to determining the relative position of the second mass and the uniform motion of the center of mass. The latter is neglected here for simplicity. It is natural to select a system of spherical polar coordinates with the applicate along the direction of the angular momentum. Since angular momentum is conserved, the motion of the second mass about the origin must lie in a plane. This means that plane polar coordinates \((r, \theta)\) suffice to describe the position of the second mass relative to the first.

It can be shown that the resulting equations governing the motion of the second mass are precisely those obtained for the behavior of a reduced mass \( m = m_1 m_2/(m_1 + m_2) \) orbiting a fixed center of force. In polar coordinates, the kinetic energy \( T \) of the reduced mass is

\[ T = \frac{1}{2} m (r^2 + r^2 \dot{\theta}^2), \tag{50} \]

where a dot over a letter signifies a derivative with respect to time. The potential energy is

\[ V = -\frac{k}{r}, \tag{51} \]

with \( k = G m_1 m_2 \), where \( G \) is Newton’s gravitational constant.

The equations of motion are

\[ \frac{d}{dt} (m r^2 \dot{\theta}) = 0 \tag{52} \]

and

\[ m \ddot{r} - m r \dot{\theta}^2 + (k/r^2) = 0. \tag{53} \]

The first of these integrates immediately in terms of the constant magnitude \( L \) of the angular momentum:

\[ m r^2 \dot{\theta} = L. \tag{54} \]

This equation can be used to eliminate \( \dot{\theta} \) from Eq. (53) by direct substitution. Also, since

\[ \frac{d}{dt} = \dot{\theta} \frac{d}{d\theta} \tag{55} \]

the independent variable in Eq. (53) can be converted from time \( t \) to angle \( \theta \). An additional change of dependent variable from \( r \) to

\[ s = (1/r) - (mk/L^2) \tag{56} \]

converts Eq. (53) into the simple form

\[ \frac{d^2 s}{d\theta^2} = -s. \tag{57} \]

The solution is readily found. Reconverting \( s \) to the variable \( r \) yields the equation for the orbit as

\[ r = l/(1 + e \cos \theta), \tag{58} \]

where a particular choice for the location of the abscissa relative to the orbit has been made for simplicity. In this equation,
and $E$ can be identified with the energy of the two bodies in the orbit. This demonstrates that the motion of two masses under gravity is described by a conic section; cf. Eq. (21). The energy $E$ determines the shape of the orbit. If $E > 0$, $e > 1$ and the orbit is a hyperbola. If $E = 0$, $e = 1$ and the orbit is a parabola. If $E < 0$, $e < 1$ and the orbit is an ellipse. Finally, if $E = -mk^2/2L^2$, $e = 0$ and the orbit is a circle.

2. DIFFERENTIAL GEOMETRY

The requirement of differentiability provides a restriction on geometrical objects that is sufficiently tight for new and useful results to be obtained and sufficiently loose to include plenty of interesting cases. Differential geometry is of vital importance in physics because many physical problems involve variables that are both continuous and differentiable throughout their range.

2.1 Manifolds

A manifold is an extension of the usual notions of curves and surfaces to arbitrary dimensions. The basic idea is to introduce an $n$-dimensional manifold as a space that is like Euclidean space $R^n$ locally, i.e., near each point. Globally, i.e., taken as a whole, a manifold may be very different from $R^n$. An example of a one-dimensional manifold is a straight line. This is both locally and globally like $R^1$. Another one-dimensional example is a circle $S^1$. The neighborhood of each point on a circle looks like the neighborhood of a point in $R^1$, but globally the two are different. The circle can be constructed by taking two pieces of $R^1$, bending them, and attaching them smoothly at each end. Generalized to $n$ dimensions, this notion of taking pieces of $R^n$ and attaching them smoothly forms the basis for the definition of a manifold.

To define a manifold more rigorously, first introduce the concept of a topological space $T$. This is a set $S$ and a collection $t$ of (open) subsets of $S$ satisfying the following criteria:

1. Both the null set and $S$ itself are in $t$.
2. The intersection of any two subsets of $t$ is in $t$.
3. The union of any collection of subsets of $t$ is in $t$.

Suppose in addition there is a criterion of separability: For any two elements of $S$ there exist two disjoint subsets of $S$, each containing one of the elements. Then $T$ is called a Hausdorff space. The elements of $S$ for a manifold are its points.

Next, define a chart $C$ of the set $S$ as a subset $U$ of $S$, called a neighborhood, together with a continuous invertible map $f: U \rightarrow R^n$ called the coordinate function. For a manifold, the subset $U$ plays the role of a region locally like $R^n$, and the function $f$ represents the introduction of local coordinates in that region. Two charts $C_1$, $C_2$ with overlapping neighborhoods and coordinate functions $f_1$, $f_2$ are called compatible if the composition map $f_2 f_1^{-1}$ is differentiable. The requirement of compatibility basically ensures that the transition from one coordinate patch to another is smooth. A set of compatible charts covering $S$ is called an atlas.

A differentiable manifold $M$ can now be defined as a Hausdorff topological space with an atlas. Given that the range of the coordinate functions is $R^n$, the dimension of $M$ is defined as $n$ and $M$ is sometimes denoted by $M^n$. An example of an $n$-dimensional manifold is the hypersphere $S^n$. An example of an object that is not a manifold is a figure-eight curve, since the neighborhood of the intersection point is not locally like $R^n$ for any $n$.

2.2 Vectors and One-Forms

The usual definition of a vector in a Euclidean space as a directed straight-line segment does not immediately extend to a general manifold. For instance, the circle $S^1$ does not contain any straight-line segments. Instead, vectors at a point of a manifold can be introduced using the notion of the tangents to curves passing through the point.

Consider a curve $s$ through a point $P$. In a neighborhood of $P$, local coordinates $(x^1, ..., x^n)$ can be used to specify $s$ in the parametric representation $x^i = x^i(t)$, $j = 1, ..., n$. A vector tangent to $s$ at $P$ can be specified by the $n$ quantities $dx^i/dt$ forming its components. A familiar example in mechanics is the velocity vector of a moving particle, obtained by differentiation with respect to time of the particle's position vector. If the tangent vectors to
all possible curves in the manifold through $P$ are considered, an $n$-dimensional vector space (see ALGEBRAIC METHODS, Sec. 3) is generated. This is called the tangent space $T_PM$ to $M$ at $P$.

In differential geometry, it is desirable to introduce basic concepts in a manner that is independent of any coordinate choice. For this reason, the differential-geometric definition of a tangent vector is different from the more intuitive one above. Given a curve $s$, introduce an arbitrary differentiable function $f$ assigning a real number to every point $t$ on $s$. The derivative $df/dt$ of $f$ along $s$ is called the directional derivative. In a local coordinate patch,

$$
\frac{df}{dt} = \sum_{i=1}^{n} \frac{dx^i}{dt} \partial_i f, \quad (60)
$$

where $\partial_i f = df/\partial x^i$. This shows that the operator $d/dt$ acting on the space of real functions on $M$ contains all components of the tangent vector, each associated with the corresponding partial derivative $\partial_i$. A tangent vector at $P$ can therefore be defined as the directional-derivative operator $d/dt$, with a natural coordinate basis of vectors for the vector space being the set of partial-derivative operators $\{\partial_i\}$. However, this definition has the disadvantage that it still explicitly includes the parameter $t$.

The formal definition of a tangent vector is therefore slightly more abstract. Given the space $F(M)$ of all differentiable real functions on a manifold $M$, a tangent vector at $P$ is defined as an object $v$ acting on elements of $F(M)$ to produce real numbers,

$$
v : F(M) \to \mathbb{R}, \quad (61)
$$

that satisfies two criteria:

$$
v(af + bg) = av(f) + bv(g),
$$

$$
v(fg) = g(P) v(f) + (f(P)) v(g), \quad (62)
$$

where $f, g \in F(M)$ and $a, b \in \mathbb{R}$. This definition extracts the important properties of the tangent vector without explicit reference to a coordinate system or parametrization. Note that the coordinate realization of a tangent vector at $P$ along $x_i$ as $\partial_i$ acting at $P$ satisfies this definition. The set of all linearly independent tangent vectors at $P$ spans the tangent space $T_PM$ to $M$ at $P$, and the set $\{\partial_i\}$ forms a basis for $T_PM$ called the coordinate basis. An arbitrary vector $v$ can be expanded in this basis as $v = \sum_i v^i \partial_i$. Physicists sometimes say that the components $v^i$ are the contravariant components of a vector. Although in a coordinate basis the intuitive physics notion of a vector and the differential-geometric one contain the same information about components, the latter also contains information about the coordinate basis itself. In the remainder of this article except where noted, the word vector refers to the differential-geometric object.

Since $T_PM$ is a vector space, there exists a dual vector space $\text{Hom}(T_PM, \mathbb{R})$ consisting of linear maps

$$
\omega : T_PM \to \mathbb{R}, \quad (63)
$$

(see ALGEBRAIC METHODS, Sec. 3.7). This space is called the cotangent space at $P$ and is denoted by $T^*M$. Notice that duality also implies $T_PM = \text{Hom}(T^*_M, \mathbb{R})$. Elements of $T^*_M$ are called one-forms. An important example of a one-form is the total differential $df$ of a function $f \in F(M)$, defined as the element of $T^*_M$ satisfying

$$
df(v) = v(f) \quad (64)
$$

for any $v \in T_PM$.

In a chart around $P$, the set $\{dx^i\}$ of total differentials of the coordinates forms a natural coordinate basis for the cotangent space $T^*_M$. It is a dual basis to $\{\partial_i\}$, since

$$
dx^i(\partial_k) = \delta^i_k \quad (65)
$$

An arbitrary one-form $\omega$ can be expanded in the dual basis as $\omega = \sum_i \omega^i dx^i$. Note that for an arbitrary vector $v = \sum_i v^i \partial_i$, the action of $\omega$ on $v$ is then

$$
\omega(v) = \omega^i v^i dx^i(\partial_i) = \omega^i v^i. \quad (66)
$$

In this equation and subsequent ones, the Einstein summation convention is introduced to simplify notation: Repeated indices in the same term are understood to be summed. The vector $v$ is said to be contracted with the one-form $\omega$. Physicists sometimes say the components $\omega^i$ form the covariant components of a vector. As an example, the definitions above can be used to show that

$$
df = \partial_i f dx^i. \quad (67)
$$

a standard result.
2.3 Tensors

The generalization of vectors and one-forms to tensors is straightforward. A tensor $T$ of type $(a, b)$ can be defined at a point $P$ of a manifold $M$ as a multilinear mapping of $a$ one-forms and $b$ vectors giving a real number:

$$ T : T_a^b \mathcal{M} \otimes \cdots \otimes T_a^b \mathcal{M} \otimes T^b_a \mathcal{M} \otimes \cdots \otimes T^b_a \mathcal{M} \to \mathbb{R}, $$

where there are $a$ factors of $T_a^b \mathcal{M}$ and $b$ factors of $T^b_a \mathcal{M}$. The space of tensors of type $(a, b)$ at $P$ is denoted $T^a_b(P)$. Examples introduced above include $T^0_0(P) = T_a^b \mathcal{M}$ and $T^1_1(P) = T_a^b \mathcal{M}$.

A tensor $T$ of type $(a, b)$ can be expanded using a coordinate basis. In the natural basis introduced above,

$$ T = T_{k_1 k_2 \cdots k_a}^{i_1 i_2 \cdots i_b} \partial_{i_1} \cdots \partial_{i_b} dx^{k_1} \cdots dx^{k_a}. $$

Almost all physicists and the older mathematicians call the quantities $T_{k_1 k_2 \cdots k_a}^{i_1 i_2 \cdots i_b}$ the components of an $a$th-rank contravariant and $b$th-rank covariant tensor. Most modern mathematicians by convention interchange the usage of contravariant and covariant. This article uses the physicists' convention.

A tensor is called symmetric with respect to two contravariant or two covariant indices if its components are unaltered when the indices are interchanged. A tensor with indices of only one type is said to be totally symmetric if it is symmetric with respect to all pairs of indices. Similarly, a tensor is antisymmetric with respect to two contravariant or two covariant indices if its components change sign when the indices are interchanged, and a totally antisymmetric tensor is one with pairwise-antisymmetric indices of only one type.

The sum and difference of two tensors of the same type is another tensor of the same type. The tensor product $T_1 \otimes T_2$ of two tensors $T_1$ and $T_2$ of types $(a_1, b_1)$ and $(a_2, b_2)$, respectively, is a tensor of type $(a_1 + a_2, b_1 + b_2)$ with components given by the product of components of $T_1$ and $T_2$ (see ALGEBRAIC METHODS, Sec. 3.8). Various contractions of two tensors can be introduced that generalize the contraction of a vector with a one-form.

A useful concept in physical applications is that of a tensor field of type $(a, b)$, defined as a particular choice of tensor of type $(a, b)$ at each point of $M$. The field is called smooth if the components $T_{k_1 k_2 \cdots k_a}^{i_1 i_2 \cdots i_b}$ of a tensor field are differentiable. Special cases are vector fields and one-form fields. By convention, if $a = b = 0$ the field is called a scalar field and is just an element of $F(M)$, the real-valued functions on $M$.

An example of a tensor that plays a crucial role in physics is the metric tensor $g$. On the manifold $M$, it is a symmetric tensor field of type $(0,2)$ such that if $g(v_1, v_2) = 0$ for any $v_1 \in T^0_0 M$, then $v_2 = 0$. In component form in a coordinate basis near a point $P$

$$ g = g_{ik} dx^i dx^k, $$

where $g_{ik}$ form the components of a symmetric, invertible matrix. The metric tensor $g$ associates any two vectors with a real number. For instance, in the usual geometry in a Euclidean space $\mathbb{R}^n$ the matrix $g_{ik} = \delta_{ik}$ and the real number is the scalar or dot product of the two vectors. In other applications different metrics may be required. For example, in special relativity space-time is taken as a four-dimensional manifold with a Minkowskian metric. If the number $g(v_1, v_2)$ has the same sign for all $v_1, v_2$ at all $P$ on $M$, i.e., if the eigenvalues of the matrix $g_{ik}$ are all of the same sign, the metric is called Riemannian. Manifolds admitting such metrics are called Riemannian manifolds. Other metrics are called pseudo-Riemannian. The special case of a metric with one eigenvalue of different sign is called Lorentzian. By diagonalization and normalization, it is always possible to choose a basis at any given $P$ such that $g_{ik}(P)$ is a diagonal matrix with entries that are $\pm 1$. If the entries are all of the same sign, the metric in this form is called Euclidean. If one entry has a different sign, it is called Minkowskian.

Since $g$ is a map $T^0_0 M \otimes T^0_0 M \to \mathbb{R}$, any given vector $v$ defines a linear map $g(v, \cdot)$ from $T^0_0 M$ to $\mathbb{R}$. This map is evidently a one-form, by definition. The components $v_j$ of this one-form are given by

$$ v_j = g_{jk} v^k. $$

The map is said to lower the index of the vector, and the result is called the associated one-form. An inverse map can be defined that uses the matrix inverse $g^{jk}$ of $g_{ik}$ to raise the index of a form, yielding a vector.

A significant part of the classical literature on differential geometry is concerned with the relationships between different manifolds,
in particular in manifolds endowed with metrics. Consider two manifolds \( M_1 \) and \( M_2 \) of dimensions \( n_1 \) and \( n_2 \). If there exists a smooth and regular map \( f: M_1 \rightarrow M_2 \), then \( M_1 \) is said to be a submanifold of \( M_2 \). The map \( f \) is called an embedding. The notion of a regular map is readily understood in coordinate patches \( \{ x^i \} \) on a chart \( U \) in \( M_1 \) and \( \{ y^j \} \) on a chart \( V \) in \( M_2 \): the matrix with components \( \frac{\partial x^j}{\partial x^i} \) must have maximal rank \( n_1 \) at each point. Intuitively, the requirements for an embedding can be viewed as ensuring for the submanifold its differentiability, the absence of self-intersections, and that curves through a point in \( M_1 \) look locally like their images in \( M_2 \). The references at the end of this article provide details of the methods and results of this subject. A simple example of a question involving the notion of embedding is the determination of equations, called the Frenet-Serret formulas, for a curve in \( R^n \). A more complicated example is the description of the embedding of a hypersurface \( M \) into \( R^n \), which, according to Bonnet's theorem, is determined by the metric tensor \( g \) on \( M \) (which in this context is called the first fundamental form), by another symmetric tensor of type \((0,2)\) called the second fundamental form, and by a set of partial differential equations called the Gauss-Codazzi equations. General results on the possibility of embedding an \( m \)-dimensional manifold into \( R^n \) are also available. An example is Whitney's theorem, which may be viewed as the statement that for compact manifolds such an embedding is possible for \( n = 2m + 1 \).

### 2.4 Differential Forms

A particularly important class of tensors is the set of totally antisymmetric tensors of type \((0,p)\) at a point of \( M^n \). These span a vector space denoted by \( \Lambda^p T^* M \) or just \( \Lambda^p T^* \), and they are called \( p \)-forms. The number \( p < n \) is called the degree of the form. For the case \( p = 0 \), \( \Lambda^0 T^* M \) is chosen as \( F(M) \), the space of real smooth functions on \( M \). The dimension of \( \Lambda^p T^* M \) as a vector space is given by the binomial coefficient \( ^n C_p \). Note that this implies that \( \Lambda^p T^* M \) and \( \Lambda^{(n-p)} T^* M \) have the same dimension.

Introduce the wedge product \( \omega_1 \wedge \omega_2 \) of two one-forms by the definition

\[
\omega_1 \wedge \omega_2 = \omega_1 \otimes \omega_2 - \omega_2 \otimes \omega_1.
\]

(72)

By construction, this is an antisymmetric tensor of type \((0,2)\), i.e., a two-form. It can be shown that a coordinate basis for the two-forms is the set \( \{ dx^1 \wedge dx^2 \} \). In general, antisymmetric tensor products of one-forms can be used to generate \( p \)-forms, and an element \( \omega \in \Lambda^p T^* M \) can be expanded in a coordinate basis as

\[
\omega_p = \sum_{1 \leq i_1 < \cdots < i_p \leq n} \omega_{i_1} \wedge \cdots \wedge \omega_{i_p} dx^{i_1} \cdots dx^{i_p}.
\]

(73)

A natural induced wedge product exists that combines a \( p \)-form \( \omega_1 \) with a \( q \)-form \( \omega_2 \) to give a \((p+q)\)-form. This product obeys

\[
\omega_1 \wedge \omega_2 = (-1)^p \omega_2 \wedge \omega_1.
\]

(74)

A larger vector space \( \Lambda T^* M \) consisting of the direct sum of all the spaces \( \Lambda^p T^* M \) can also be considered. Its dimension is \( 2^n \), and it is called the Cartan exterior algebra of \( T^* M \).

Analogous constructions can be introduced for the case of antisymmetric tensors of type \((p,0)\), called \( p \)-vectors. The totality of these spans a space denoted \( \Lambda^p T \). The \( p \)-forms, \((n-p)\)-forms, \( p \)-vectors, and \((n-p)\)-vectors thus all form vector spaces of dimension \( ^n C_p \) at a point \( P \) of \( M^n \). Various relations can be constructed between these spaces. An important example is the Hodge star map \( * \), defined for manifolds \( M \) that have a metric \( g \). This is a linear map \( *: \Lambda^p T^* M \rightarrow \Lambda^{(n-p)} T^* M \) that is most easily understood by its action on coordinate components. Define the totally antisymmetric symbol by

\[
e_{i_1 \cdots i_n} = \begin{cases} 
+1 & \text{if } (i_1 \ldots i_n) \text{ is an even permutation of } (1 \ldots n) \\
-1 & \text{if } (i_1 \ldots i_n) \text{ is an odd permutation of } (1 \ldots n) \\
0 & \text{otherwise.}
\end{cases}
\]

(75)
If a $p$-form $\omega$ is given in a coordinate basis by Eq. (73), then
\[
\omega = \frac{\sqrt{|g|}}{p!|n-p|!} \sum_{i_1, \ldots, i_p} g^{i_1} \cdots g^{i_p} \omega_{i_1 \cdots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p},
\]
(76)
where $g^{ij}$ is the inverse metric matrix introduced in Sec. 2.3 and $g$ is the determinant of the matrix $g_{ik}$.

From the definition (64), the total differential of a zero-form is a one-form. An extension of the notion of differential can be introduced to obtain a $(p+1)$-form via a $p$-form. Formally, a map $d: \wedge^{p+1}T^* -\wedge^p (\wedge^{p+1}T^*)$ called the exterior derivative can be defined by the following requirements:

1. $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$ for $\omega_1, \omega_2 \in \wedge^p T$;
2. $d(\omega_1 \wedge \omega_2) = (d\omega_1 \wedge \omega_2) + (-1)^p (\omega_1 \wedge d\omega_2)$ for $\omega_1 \in \wedge^p T$ and $\omega_2 \in \wedge^q T$ and
3. $d(d\omega) = 0$ for $\omega \in \wedge^p T$.

It can be shown that the exterior derivative is unique. In a coordinate basis, the exterior derivative of a $p$-form given by Eq. (73) is
\[
d\omega = \frac{1}{p!} \partial_{\omega_{i_1 \cdots i_p}} dx^k \wedge dx^{i_1} \wedge \cdots \wedge dx^r.
\]
(77)
A $p$-form field with vanishing exterior derivative is said to be closed, while one that is obtained as the exterior derivative of a $(p-1)$-form is called exact. The definition of $d$ implies that an exact form is necessarily closed.

The exterior derivative combines in a single notation for manifolds $M^n$ extensions of the gradient, divergence, and curl operations of usual three-dimensional vector calculus. For instance, the gradient of a function $f$ is a covariant vector with components $\partial f$. These are precisely the components of the one-form in Eq. (67). The components of the curl make their appearance in the exterior derivative of a one-form $\omega = \omega_x dx + \omega_y dy + \omega_z dz$:
\[
d\omega = (\partial_\omega_y - \partial_\omega_z) dx \wedge dy + (\partial_\omega_z - \partial_\omega_x) dy \wedge dz + (\partial_\omega_x - \partial_\omega_y) dz \wedge dx.
\]
(78)
The divergence enters the expression for the exterior derivative of a two-form $\omega = \omega_{xy} dx \wedge dy + \omega_{xz} dy \wedge dz + \omega_{yz} dz \wedge dx$:
\[
d\omega = (\partial_\omega_{yz} + \partial_\omega_{zx} + \partial_\omega_{xy}) dy \wedge dz + \partial_\omega_{xz} dz \wedge dx + \partial_\omega_{yz} dx \wedge dy.
\]
(79)
The statement $dd = 0$ contains the usual identities $\text{div}(\text{curl } \nu) = 0$ for a vector $\nu$ and a function $f$.

The existence of the Hodge star map $\star$ makes it possible to define a map from $p$-forms to $(p-1)$-forms by applying first $\star$ [producing an $(n-p)$-form], then $d$ [giving an $(n-p+1)$-form], and finally $\star$ again. This map is called the adjoint exterior derivative and denoted $\delta$.

For Riemannian metrics it is defined as
\[
\delta = (-1)^{np+n+1} d\star
\]
(80)
while for Lorentzian metrics there is an additional factor of $-1$. The adjoint exterior derivative satisfies $\delta \omega = 0$. A $p$-form field with vanishing adjoint exterior derivative is said to be co-closed, while one that is obtained as the adjoint exterior derivative of a $(p+1)$-form is called coexact.

It is possible to express the Laplacian $\Delta$ on a manifold $M^n$ in terms of the maps $d$ and $\delta$:
\[
\Delta = (d + \delta)^2 = d\delta + \delta d.
\]
(81)
For example, acting on a function $f$ in three dimensions, this definition reproduces the standard expression of vector calculus,
\[
\Delta f = (1/\sqrt{|g|}) \partial_i (\sqrt{|g|} g^{ij} \partial_j f).
\]
(82)
A $p$-form $\omega$ is said to be harmonic if $\Delta \omega = 0$. This generalizes the usual notion of harmonic functions.

2.5 Fiber Bundles

In addition to involving a manifold of variables, many physical situations also exhibit symmetry of some kind. The natural geometrical framework in which to formulate such problems is the language of fiber bundles. Here, attention is restricted to a special type of bundle, appearing widely in physics, that involves continuous symmetries. The latter are described mathematically via the theory of Lie groups.

This paragraph presents a few essential definitions involving Lie groups. More details may be found in the articles GROUP THEORY and ALGEBRAIC METHODS. For the present purposes, a Lie group $G$ may be viewed as a group that is also an $r$-dimensional manifold such that for two group elements $g, h \in G$ the map $gh^{-1}: G \times G \to G$ exists and is continuous. Denote coordinates in a chart near some point $P$ of $G$ by $(a^A)$, $A = 1, \ldots, r$. Then the group
composition function \( f : G \times G \rightarrow G \) defined for
\( g(a), h(b), k(c) \in G \) by \( f(h(a))k = gh \) can be written
in terms of \( r \) functions \( \phi^A \) acting on the coordinates as
\[
\mathcal{C}^A = \phi^A(b, a). \tag{83}
\]
The generators \( D_A \) of infinitesimal group transformations on \( G \) span the tangent space \( T_G \) at the
group identity and are given by
\[
D_A = U^a_b \partial_{\theta^b}, \quad U^a_b = \frac{\partial \phi^B}{\partial \phi^A}, \tag{84}
\]
This space is called the Lie algebra associated with the group. The dual basis is spanned by the one-forms
\[
\Omega^A = da^B (U^{-1})^A_B. \tag{85}
\]
As a simple example, consider the group \( U(1) \).
The group manifold is a circle \( S^1 \); if the coordinate is denoted by \( \theta \), the group composition function is \( \theta_2 = \theta_1 + \theta_3 \). The generator \( D_\theta \) is just \( d\theta \) and the dual basis is \( d\theta \).
A fiber bundle is basically a manifold acted on by a symmetry. One important type of bundle, called a principal bundle, looks locally (but not necessarily globally) like a product of a continuous symmetry group
with a manifold. The action of the symmetry provides a natural means of moving around in each local piece of bundle. The idea is to patch together these local pieces in a smooth way to get the whole principal bundle. Globally, the patching can introduce various twists into the overall structure, in which case the bundle is called nontrivial. A trivial bundle is one where no twists arise: the global and local structure are similar.

Here is a more formal definition. Given a manifold \( B \) and a Lie group \( G \), a principal fiber bundle \( E(B, G) \) is a manifold such that

1. \( G \) acts differentiably and without fixed points on \( E \);
2. \( B \) is the quotient space of \( E \) by equivalence under \( G \), and there exists a differentiable map \( \pi : E -\rightarrow B \); and
3. for each chart \( U_j \) in an atlas for \( B \), there exists a differential and invertible mapping \( \phi_j : \pi^{-1}(U_j) \rightarrow U_j \times G \) given by \( E = (\pi(P), f(P)) \) for any point \( P \in E \), where \( f : \pi^{-1}(U_j) \rightarrow G \) satisfies \( f(gP) = g f(P) \) for any \( g \in G \).

The group \( G \) is called the structure group and the manifold \( B \) is called the base manifold. The map \( \pi \) is called the projection. The inverse image of \( \pi \) is the fiber; in effect, each fiber is like a copy of \( G \). A (global) cross section or section \( s \) of a bundle is defined as a smooth map \( s : B \rightarrow E \) such that \( \pi \circ s \) is the identity on \( B \). Local sections, i.e., sections defined only on \( \pi^{-1}(U_j) \), always exist. If the bundle admits a global section, it is called trivial.

### 2.6 Connection and Curvature

Since \( \{ \partial_j \} \) is a basis for the tangent space of the base manifold and \( \{ D_A \} \) is one for the tangent space of the group, a basis for the tangent space to a point in the bundle is the set \( \{ \partial_j, D_A \} \). It has dual basis \( \{ dx^i, \Omega^A \} \). However, linear combinations could also be taken. The existence of this freedom permits the definition of a natural one-form called the connection that contains essential information about the structure of the bundle. The connection is basically a separation of the tangent space of \( E \) into two pieces, one along the group.

Formally, a connection is a choice of a subspace \( T_PH \) of \( T_P E \) at each point \( P \) of \( E \) such that

1. \( T_PE = T_PG \oplus T_PH \), where \( T_PG \) is the space of vectors tangent to the fiber at \( P \);
2. \( T_PH \) is invariant under action by \( G \); and
3. the components in \( T_PG \) and \( T_PH \) of a smooth vector field in \( T_PE \) are also smooth. The spaces \( T_PG \) and \( T_PH \) are called the vertical and horizontal subspaces, respectively.

Some of the implications of this definition are most easily seen in a coordinate basis on the bundle. Let a basis for \( T_PH \) be defined as the linear combination
\[
D_j = \partial_j - h^A_j D_A, \tag{86}
\]
and require that \( D_j \) commute with \( D_A \) (among other consequences, this implies that \( h^A_j \) transforms under a particular representation of the Lie algebra of \( G \), called the adjoint representation). Then the coefficients \( h^A_j \) are called connection coefficients and the basis elements \( \{ D_j \} \) are called the horizontal lifts or the covariant derivatives of the basis elements \( \{ \partial_j \} \). The dual to the basis \( \{ D_A, D_j \} \) for \( T_PE \) is the set \( \{ dx^i, \omega^A \} \), where the \( \omega^A \) are given by
\[
\omega^A = \Omega^A + h^A_j dx^j. \tag{87}
\]
They form the components of a composite one-form \( \omega = \omega^A D_A \), called the connection form.
The connection form $\omega$ encodes many of the interesting properties of the bundle in a concise notation. Its exterior derivative is also an important quantity in physical applications. Introduce a two-form $R$ called the curvature form of the bundle by the definition

$$R = d\omega + \omega \wedge \omega.$$  \hspace{1cm} (88)

The curvature is said to be a horizontal form because its action on any vertical vector vanishes. Its nonzero components are given by the expressions

$$R = R^A_{\alpha} D_A, \quad R^A_{\alpha} = R^A (D_j D_k),$$  \hspace{1cm} (89)

and it follows that

$$[D_p D_k] = R^A_{pk} D_A.$$  \hspace{1cm} (90)

Applying another exterior derivative gives an identity

$$dR = R \wedge \omega - \omega \wedge R = 0$$ \hspace{1cm} (91)

called the Bianchi identity, with components

$$\sum_{jkl} D_j R^A_{kl} = 0,$$  \hspace{1cm} (92)

where the sum is over cyclic permutations of the indices $j, k, l$.

### 2.7 Example: Electromagnetism

An illustration of the role of some of these ideas in physics is provided by the formulation of the theory of electromagnetism in differential-geometric language. First, here is a summary of a few of the key equations of electromagnetism. In this section, a boldfaced symbol denotes a vector viewed as a collection of components. The symbol $\nabla$ is the usual vector gradient operator, while $\cdot$ indicates the vector dot product and $\times$ represents the vector cross product. The Maxwell equations in SI units include Gauss's law,

$$\nabla \cdot E = \rho / \varepsilon_0;$$  \hspace{1cm} (93)

Faraday's law,

$$\nabla \times E + \partial_t B = 0;$$  \hspace{1cm} (94)

equation expressing the absence of magnetic monopoles,

$$\nabla \cdot B = 0;$$  \hspace{1cm} (95)

and the Ampère-Maxwell law,

$$\nabla \times B = \mu_0 J + c^{-2} \partial_t E,$$  \hspace{1cm} (96)

where $\varepsilon_0$ is the absolute permittivity, $\mu_0$ is the absolute permeability, and $c = 1 / \sqrt{\varepsilon_0 \mu_0}$ is the speed of light in vacuo. Although these equations can be solved directly in simple cases, it is often useful to introduce new variables, called potentials, in terms of which the four first-order Maxwell equations are replaced with two second-order equations. The scalar and vector potentials $\phi$ and $A$ are defined by

$$E = -\nabla \phi - \partial_t A, \quad B = \nabla \times A.$$  \hspace{1cm} (97)

With these definitions, the homogeneous equations (94) and (95) are automatically satisfied. The two inhomogeneous Maxwell equations become coupled second-order equations for the potentials:

$$\nabla^2 \phi + \partial_t \nabla \cdot A = -\rho / \varepsilon_0$$  \hspace{1cm} (98)

and

$$\nabla^2 A - c^{-2} \partial_t^2 A - \nabla (\nabla \cdot A + c^{-1} \partial_t \phi) = -\mu_0 J.$$  \hspace{1cm} (99)

There exists a freedom in the definition of $\phi$ and $A$. The electric field $E$ and the magnetic induction $B$ are unchanged by the replacements

$$\phi \rightarrow \phi' = \phi - \partial_t \Lambda$$  \hspace{1cm} (100)

and

$$A \rightarrow A' = A - \nabla \Lambda,$$  \hspace{1cm} (101)

where $\Lambda$ is a function of $x$ and $t$. These replacements are called gauge transformations. Their existence provides sufficient freedom to decouple Eqs. (98) and (99). For more details on the subjects of this paragraph, see ELECTRODYNAMICS, CLASSICAL.

It is easiest to approach the differential-geometric formulation of electromagnetism in stages, each incorporating more aspects of the theory. Here, the Maxwell equations for $E$ and $B$ are first expressed using the language of differential forms. The structure of the theory as a fiber bundle is then described, thereby incorporating the potentials $\phi$ and $A$ and the notion of gauge transformations. To obtain consistent physical dimensionalities within expressions, it is convenient to work with a coordinate $x^0 = ct$ with dimensions of length rather than with the time coordinate $t$. In what follows, the spatial coordinates $(x, y, z)$ are denoted $(x^1, x^2, x^3)$.

Begin with the identification of the space and time dimensions as a four-dimensional smooth manifold $M$. The manifold is often
taken to be $R^4$ but this is not essential. The tangent space to $M$ at a point $P$ is also four-dimensional, and a basis for this space is the set $\{\partial_\mu\}, \mu=0,1,2,3,$ of derivatives with respect to the four coordinates $(x^0, x^1, x^2, x^3)$. An arbitrary vector can be expanded with respect to this basis. One vector, denoted by $j$ and called the four-vector current, has components $\jmath^\mu$ formed from the charge and current densities $\rho, J$:

$$j=\jmath^\mu \partial_\mu = (\mu \varphi/c) \partial_0 + \mu_0 J \cdot \nabla. \quad (102)$$

An important tensor field on $M$ is the Minkowskian metric $g$, defined to have components $g_{\mu\nu}$ in a coordinate basis forming a matrix given by

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (103)$$

This incorporates the essential elements of special relativity.

The Maxwell equations can be expressed in terms of a two-form field $F$ defined on $M$. This antisymmetric tensor of type $(0,2)$ is called the field strength. The components $F_{\mu\nu}$ of $F$ are related to the components of the electric field and magnetic induction, and are given by

$${F_{\mu\nu}} = \begin{pmatrix} 0 & +E^1/c & +E^2/c & +E^3/c \\ -E^1/c & 0 & -B^3 & +B^2 \\ -E^2/c & +B^3 & 0 & -B^1 \\ -E^3/c & -B^2 & +B^1 & 0 \end{pmatrix}. \quad (104)$$

This assignation of $E$ and $B$ is not a priori mathematically unique but establishes agreement of the resulting theory with experiment. In terms of the two-form $F$, the inhomogeneous Maxwell equations can be rewritten as

$$dF=j, \quad (105)$$

and the homogeneous ones become

$$d*F=0. \quad (106)$$

The two-form $*F$ is called the dual field strength. In component form in a coordinate basis, these equations read

$$\partial_\mu F^\mu = \jmath^n. \quad (107)$$

and

$$\varepsilon_{\rho\nu\mu\sigma} \partial_\rho F^{\mu\nu} = 0. \quad (108)$$

Each of these represents four equations, and an inspection shows they reduce to the usual form of the Maxwell equations upon substitution in $F$ and $j$ of $E$, $B$, $\rho$, and $J$.

The discussion so far has excluded the potentials $\phi$ and $A$. These can be combined to form the components $A^\mu$ of a vector, called the gauge potential:

$$A^\mu \partial_\mu = (\phi/c) \partial_0 + A \cdot \nabla. \quad (109)$$

The factor of $c$ is introduced to maintain dimensional consistency. The metric $g$ provides the associated one-form

$$A = A^\mu dx^\mu = g_{\mu\nu} A^\nu dx^\mu, \quad (110)$$

with components obtained by lowering the index. A complete description of the differential-geometric role of the gauge potential in electromagnetism requires a framework in which to place its nonuniqueness under gauge transformations. This freedom can be interpreted as a symmetry of Eqs. (98) and (99) expressing electromagnetism in terms of the potentials. It can be shown that this symmetry is a Lie group, called $U(1)$. A natural geometrical framework to express this is a fiber bundle, as is discussed next. For simplicity in what follows, the charge and current densities are taken to vanish. Nonzero distributions can be incorporated consistently with the addition of some extra structure.

The bundle of interest is a principal fiber bundle with the four-dimensional space-time manifold taken as the base manifold $B$ and the symmetry group $U(1)$ of gauge transformations taken for the structure group $G$. Since the manifold of the group $U(1)$ is a circle $S^1$, the principal bundle is five-dimensional. Denote the coordinate on $S^1$ by $\theta$. The introduction of a connection separates the tangent space to a point $P$ in the bundle into a four-dimensional horizontal subspace spanned by the basis $\{\partial_\mu = \partial_\mu\}$ and a one-dimensional vertical subspace spanned by the generator $D_\theta = \partial_\theta$ of the Lie algebra of $U(1)$. The dual basis is the set $\{dx^\mu, \Omega^\theta = d\theta\}$. The composite connection form $\omega$ is $\omega = \Omega^\theta D_\theta = d\theta \partial_\theta$.

The gauge potential $A$ can be identified with the value of the one-form $\Omega^\theta$ on a section $s$ of the bundle. Suppose that the surface $s$ through the bundle $E$ is specified in a chart $U$ by choosing the group coordinate $\theta$ as a
function of the coordinates \( (x^\mu) \) provided by \( U \). Then the dual form becomes

\[
\Omega^\theta = d\theta = \partial_\mu \beta (x) dx^\mu = A_\mu (x) dx^\mu,  \tag{111}
\]

where the identification of the components of the one-form \( \Omega^\theta \) with the components of the gauge-potential one-form has been made. Under a change of cross section, which is equivalent to the action of a group element with a parameter \( \Lambda \), say, the potentials \( A_\mu \) change by an amount \( \partial_\alpha \Lambda \). This provides the geometrical interpretation for the gauge transformations (100) and (101).

The curvature two-form \( dw + \omega \wedge \omega \) derived from the connection form \( \omega \) is denoted by \( F \). Evaluated on a local section, it has components

\[
F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.  \tag{112}
\]

In terms of the scalar and vector potentials, this equation reproduces the definitions of Eq. (97). The Bianchi identity in component form in this case can be written

\[
d*F = 0,  \tag{113}
\]

thereby reproducing the homogeneous Maxwell equations. To complete the specification of the bundle, additional equations are needed that explicitly determine in each section the connection and the curvature. These are called equations of motion. Requiring these to transform as usual under Lorentz transformations and to be second-order differential equations for the connection or first-order equations for the curvature significantly restricts the options. An inspection of the Lorentz representation content of the general first-order term \( \partial_\nu F_{\mu\nu} \) shows that the simplest choice is \( \partial_\nu F_{\mu\nu} = 0 \) or its form equivalent

\[
dF = 0.  \tag{114}
\]

This reveals the geometrical role of the remaining two equations in Maxwell's theory.

In the presence of monopoles, the homogeneous Maxwell equations are modified by the introduction of sources. A geometrical setting for the equations describing the field of a monopole is provided by a nontrivial principal bundle. It can be shown that the essential physics is contained in a bundle with base space \( S^2 \) and structure group \( O(1) \). The bundle space \( E \) looks like \( S^3 \) and the projection map \( \pi \) is called the Hopf map.

### 2.8 Complex Manifolds

Just as the requirement of differentiability for manifolds introduces many useful structures, a further restriction imposing complex analyticity is of considerable interest. The resulting manifolds, called complex manifolds, look locally like the complex plane. Some of their main features are outlined in this section. Basic methods of complex analysis are assumed here. See Analytical Methods, Sec. 1, for more details.

The formal definition of a complex manifold \( M \) parallels that for a real differentiable manifold presented in Sec. 2.1. The key difference is that the local charts now contain maps \( f \) taking neighborhoods \( U \) into \( C^n \), the product of \( n \) complex planes \( C \), and that the composition map \( f_1 \circ f_2^{-1} \) is required to be holomorphic rather than differentiable. This ensures that the methods of complex analysis can be used on \( M \) independently of any choice of chart. The number \( n \) is called the complex dimension of \( M \); the real dimension is \( 2n \). An important feature is that a complex manifold may have two or more incompatible atlases, i.e., the union of two atlases may not satisfy the requirements for an atlas. In this case the atlases are said to define different complex structures. An example is the torus \( T^2 \) with two real dimensions; it can be shown that the complex structures on the torus are distinguished by a complex number called the modular parameter.

Denote the \( n \) complex coordinates on \( M \) in a chart \( U \) by \( z^j = x^j + iy^j, \ j = 1, \ldots, n \), with \( \bar{z}^j = x^j - iy^j \). The tangent space \( T_p M \) at a point \( P \) of \( M^n \) is spanned by a \( 2n \)-dimensional coordinate basis \( \{ \partial/\partial x^j, \partial/\partial y^j \} \). It is useful to define

\[
\partial_j = \frac{\partial}{\partial x^j} = \frac{1}{2} \left( \frac{\partial}{\partial x^j} - i \frac{\partial}{\partial y^j} \right),
\]

\[
\partial_{\bar{j}} = \frac{\partial}{\partial \bar{x}^j} = \frac{1}{2} \left( \frac{\partial}{\partial \bar{x}^j} + i \frac{\partial}{\partial \bar{y}^j} \right).  \tag{115}
\]

The cotangent space is spanned by the dual basis \( \{ dx^j, dy^j \} \), or equivalently by

\[
\{ dz^j = dx^j + idy^j, \ \bar{dz}^j = dx^j - idy^j \}.  \tag{116}
\]

Define the linear map \( J: T_p M \rightarrow T_p M \) by

\[
J \partial_j = i \partial_j, \quad J \partial_{\bar{j}} = -i \partial_{\bar{j}}.  \tag{117}
\]

Note that \( J^2 = -I \). This map is smooth and globally defined on any complex manifold \( M \).
It is called the almost complex structure of \( M \). The action of \( J \) separates \( T_pM \) into two separate vector spaces, one spanned by vectors \( v \) such that \( Jv = iv \) and the other by vectors such that \( Jv = -iv \). It follows that a vector in \( T_pM \) can be uniquely decomposed into two pieces, called the holomorphic and antiholomorphic parts. The cotangent space \( T^*_pM \) can be separated into two corresponding pieces.

Complex differential forms of degree \((p,q)\) can also be introduced. These are elements of a vector space denoted by \( \Lambda^{(p,q)}T \). In local coordinates, \( \Lambda^{(p,q)}T \) is spanned by a coordinate basis containing \( p \) factors of \( dz^i \) and \( q \) factors of \( d\bar{z}^j \). The exterior derivative \( d \) naturally separates into the sum of two pieces,

\[
    d = \partial + \bar{\partial},
\]

called the Dolbeault operators. They satisfy

\[
    \bar{\partial} \partial = \bar{\partial} \partial = \partial \bar{\partial} + \bar{\partial} \partial = 0.
\]

All complex manifolds admit a Hermitian metric. A Riemannian metric \( g \) on \( M \) is said to be Hermitian if

\[
    g(Jv_1, Jv_2) = g(v_1, v_2)
\]

for all vectors \( v_1, v_2 \in T_pM \) at all points \( P \). In a coordinate basis, \( g \) can be shown to have the form

\[
    g = g_{ik} dz^i \wedge d\bar{z}^k + g_{k\bar{l}} d\bar{z}^k \wedge dz^\ell.
\]

One can also define a two-form \( \Omega \) called the Kähler form by

\[
    \Omega(v_1, v_2) = g(Jv_1, v_2).
\]

If the Kähler form is closed, \( d\Omega = 0 \), the manifold is called a Kähler manifold and the metric \( g \) is said to be a Kähler metric. In a chart, the components of a Kähler metric can be written as

\[
    g_{ik} = \partial_i \partial_k K,
\]

where \( K \) is a scalar function called the Kähler potential. Compact Kähler manifolds in one complex dimension are called Riemann surfaces and are of great importance in certain branches of physics, notably string theory. Examples of Riemann surfaces are the two-sphere \( S^2 \) and the two-torus \( T^2 \).

### 2.9 Global Considerations

Essentially all the differential geometry considered above has involved local concepts. It is also of interest to address the issue of the extent to which the local properties of a manifold determine its global ones. The study of global properties of a manifold forms part of the branch of mathematics called topology (\( q.v. \)) and as such is tangential to the scope of this article. This section provides a sketch of some connections between the two subjects. Details may be found in the references provided at the end of the article.

One link between the geometry and topology of a differentiable manifold \( M \) can be introduced by considering the space of all closed \( p \)-forms on \( M \). This space can be separated into classes, each containing closed forms differing from one another only by exact ones. The set of all classes is a vector space called the \( p \)th de Rham cohomology group of \( M \) and denoted \( H^p(M) \). This vector space contains topological information about \( M \). For example, the dimension of \( H^p \), called the \( p \)th Betti number, is a topological invariant of \( M \) that contains information about the holes in \( M \). The Betti numbers also determine the number of harmonic forms on \( M \).

There are relationships between the number of critical points of functions on a manifold \( M \) and the topology of \( M \). This is the subject of the calculus of variations in the large, or Morse theory. Among the results obtained are the Morse inequalities, which relate the number of certain types of critical points of a function to combinations of the Betti numbers on \( M \).

The presence of a metric on \( M \) permits other types of global information to be obtained. An example is the Hodge decomposition theorem. This can be viewed as the statement that on a compact orientable Riemannian manifold \( M \) without boundary, any \( p \)-form can be uniquely decomposed into the sum of an exact form, a coexact form, and a harmonic form.

The issue of describing the global structure of a bundle (not necessarily principal) provides another link between geometry and topology. It is possible to develop measures of the ways in which a given bundle differs from the trivial bundle. The relevant mathematical objects are called characteristic classes. They are elements of the cohomology classes of the...
base manifold, and are given different names depending on the type of bundle being considered. Among these are Pontrjagin, Euler, and Chern classes, corresponding to orthogonal, special orthogonal, and unitary structure groups. Elements in these classes can be expressed in terms of the curvature two-form of the bundle. Another set of characteristic classes, the Steifel-Whitney classes, determines the orientability of a manifold and whether a spinor field can be consistently defined on it.

There are also relations between certain aspects of differential operators on bundles and the topology of the bundles. These are given by index theorems. An important example is the Gauss-Bonnet theorem, which connects the number of harmonic forms on a manifold (this is a property of the exterior derivative operator) to an integral over the Euler class (this is a topological quantity). Another important example is the Riemann-Roch theorem for complex manifolds. These are special cases of the Atiyah-Singer index theorem.

2.10 Further Examples

Many sets of smooth physical variables can be viewed as differentiable manifolds, and so differential-geometric concepts such as vectors, tensors, forms, and bundles play key roles in much of theoretical physics. Examples can be found in every major branch of physics. For instance, the modern formulation of the Hamiltonian dynamics of a system proceeds via the investigation of a manifold \( M \) called the phase space, with local coordinates corresponding to the generalized coordinates and momenta of the system. A closed nondegenerate two-form called the symplectic form is defined on \( M \), making the phase space a symplectic manifold. The study of the properties of the phase space using the methods of differential geometry provides information about the behavior of the system. An extension of this example occurs in quantum mechanics (see QUANTUM MECHANICS). Quantization of a system involves the introduction of complex structure on the symplectic manifold. The study of this procedure is called geometric quantization.

Differential geometry is particularly crucial in the development of theories of fundamental interactions and particles. The geometrical constructions presented above for electromagnetism can readily be extended to other theories of fundamental forces. For example, the equations believed to describe the underlying physics of the strong interactions (see the article STRONG INTERACTIONS) form a theory called chromodynamics. This theory can be expressed geometrically using a principal bundle over space-time but where the structure group is the eight-dimensional Lie group called \( SU(3) \) rather than \( U(1) \). The presence of a multidimensional group manifold with a nontrivial group composition law means that, unlike the electrodynamics case, the horizontal lifts are inequivalent to the basis for the tangent space to the base manifold. As a result, the structure of the Bianchi identities and the equations of motion are somewhat more complicated in detail. The essential construction, however, remains the same.

Another important physical theory is general relativity, which provides a good description of the gravitational interactions at the classical level (see the article GRAVITATION AND GENERAL RELATIVITY). This theory can also be given a geometrical interpretation as a fiber bundle, but it is of a somewhat different kind, called a bundle of frames. Each point on a fiber of this bundle consists of a choice of basis vectors for the tangent space to the space-time manifold, and the symmetry group that plays the role of the structure group of a principal bundle now acts to rotate these bases into one another. A connection form and an associated curvature still exist, and they are closely related to the Christoffel symbols and the Riemann space-time curvature tensor of general relativity. In addition, there exists new freedom arising from the choice of basis vector on the base manifold, which leads to the existence of a second natural one-form on the bundle called the solder form or vierbein. This also has an associated two-form, called the torsion. In Einstein's general relativity the torsion form is specified to be zero, although other possibilities can be envisaged.

Attempts to unify the known fundamental forces and particles (see UNIFIED FIELD THEORIES) make wide use of geometrical constructions. Examples of such theories in four dimensions are the grand unified theories, describing the strong, weak, and electromagnetic forces in a single framework. The geo-
metrical structures discussed above can be extended to more complicated symmetry groups large enough so that the connection forms include all the force fields needed for these theories. Certain elementary particles (see PARTICLES, ELEMENTARY) play the role of sources for these fields and can also be incorporated in bundles called associated bundles. Many unified theories involve higher-dimensional manifolds, in which physical spacetime is a submanifold. These include the so-called Kaluza-Klein theories. Often, the symmetries of the extra dimensions permit them to play the role of the structure group in a principal bundle.

Generalizations of the geometrical framework of gravitation are also possible. For example, if the base manifold for a bundle of frames is generalized in a certain way, it is possible to specify bundles describing extensions of general relativity that include fundamental particles and forces other than gravity and that incorporate enlarged symmetries called supersymmetries. The resulting theories are called supergravities.

String theories are candidate unified theories including gravity that are believed to be consistent with quantum mechanics. In these theories, the fundamental forces and particles are interpreted as objects that are extended in one dimension (hence the name string). As a string propagates in space-time, it sweeps out a two-dimensional surface called the world sheet. A description of the world sheet involves the study of complex manifolds, in particular Riemann surfaces, as well as the notions of global differential geometry.

3. PROJECTIVE GEOMETRY

In its basic form, projective geometry is essentially the theory of perspective, i.e., the study of those features of geometrical objects that remain the same when the objects are projected from a point onto a line or plane. The elements of projective geometry are implicitly used by artistic painters, designers, and other people who represent three-dimensional objects on a two-dimensional medium. In its generalized form, the subject is fundamental in axiomatic geometry. It can be viewed as subsuming the classical Euclidean and non-Euclidean geometries.

There are two approaches to projective geometry. Synthetic projective geometry seeks to develop the subject as a series of deductions starting from certain axioms, in the Euclidean tradition. Analytical projective geometry introduces homogeneous coordinates and uses analytical techniques to obtain results. The two approaches are complementary, although projective geometries exist for which coordinates cannot be introduced.

A key feature of projective geometry is that parallel lines are assumed to meet in a single point, called the point at infinity, and that parallel planes meet in a single line, called the line at infinity. One advantage of these assumptions is that geometrical statements do not require exceptions for parallelism. For example, it is now true that any two lines in the plane determine a point, and any two planes in three dimensions determine a line.

In a plane, the statement that two lines determine a point is strikingly similar to the statement that two points determine a line. In general, projective-geometric statements involving points and lines in the plane remain valid when the roles of the points and lines are interchanged. In the plane, points are said to be dual to lines. In three dimensions the notion of duality applies between points and planes, or between lines and lines. A similar concept exists in higher dimensions.

With these ideas, a set of axioms for synthetic projective geometry can be formulated in terms of three basic notions: point, line, and incidence. The latter is meant in the sense of intersection: for example, a point is incident to a line if it lies on the line. The axioms can be expressed in dual pairs, so that propositions deduced necessarily have valid duals.

3.1 Some Theorems

There are several theorems that play a central role both in the development of the basic theory and in its extension to more abstract situations. A key result is Desargues's theorem: Given six distinct points in two sets, \( \{A_1, A_2, A_3\} \) and \( \{B_1, B_2, B_3\} \) (i.e., the vertices of two triangles), if the lines \( A_1B_1, A_2B_2, A_3B_3 \) meet at a point, then the three points \( C_1, C_2, C_3 \) given respectively by the pairwise line intersections \( A_1B_2 \) and \( A_2B_1 \), \( A_2B_3 \), and \( A_3B_2 \), and \( A_1B_3 \) are collinear. This theorem holds in all projective geometries in three dimensions or more and in certain two-dimensional cases,
including the usual plane projective geometry. However, in two dimensions non-Desarguesian geometries also exist.

Another important result that holds for a large class of projective geometries including the usual plane and solid ones is Pappus’s theorem: Given two lines $a$ and $b$ lying in a plane and two sets of three distinct points $(A_1, A_2, A_3)$ incident to $a$ and $(B_1, B_2, B_3)$ incident to $b$, then the three points $C_1, C_2, C_3$ given respectively by the pairwise line intersections $A_1B_2$ and $A_2B_1$, $A_2B_3$ and $A_3B_2$, $A_3B_1$ and $A_1B_3$ are collinear. Non-Pappian geometries also exist.

A pencil of lines about a point $P$ is defined as the set of all lines lying in a plane and incident with $P$. A line $s$ in the plane not incident with $P$ is called a section of the pencil, and the pencil is said to project the section from $P$. Two pencils can be projectively related through a common section. Two distinct sections are said to be related by a projective transformation from the point $P$. The fundamental theorem of projective geometry states that a projective transformation is specified when three collinear points and their images are given. The theorem generalizes to projective transformations of higher-dimensional figures.

Conic sections (see Sec. 1.2) have a natural construction in projective geometry, and their theory can be developed entirely within this subject. Since all conics can be generated by projection of a circle from a point onto a plane, the projective approach gives them a unified treatment and consequently several results of analytical geometry can follow from a single projective theorem. Plane-projective definitions also play an important role. For example, the locus of intersections of corresponding lines in two projectively related pencils is a conic. A well-known result in this branch of the subject is Pascal’s theorem: Given six points $(A_1, A_2, A_3, A_4, A_5, A_6)$ incident to a conic, then the three points $B_1, B_2, B_3$ given respectively by the pairwise line intersections $A_1A_2$ and $A_4A_5$, $A_2A_3$ and $A_5A_6$, $A_3A_4$ and $A_6A_1$, are collinear. The dual to Pascal’s theorem is sometimes called Brianchon’s theorem. These methods of projective geometry can also be extended to the study of quadrics and higher-dimensional hypersurfaces.

### 3.2 Homogeneous Coordinates

In analytical projective geometry, a set of coordinates called homogeneous coordinates is introduced. Consider first homogeneous coordinates on the line. A Cartesian coordinate system assigns a single real number $x$ to each point $P$. In contrast, a homogeneous coordinate system assigns two real numbers $(x_0, x_1)$ to each point, where $x = x_1/x_0$ and at least one of $(x_0, x_1)$ is nonzero. Evidently, the homogeneous coordinates $(x_0, x_1)$ and $(cx_0, cx_1)$, where $c$ is a constant, both represent $P$. The advantage of homogeneous coordinates is that the point $(0, 1)$ at infinity is treated on the same footing as, say, the origin $(1, 0)$. It also makes any polynomial equation $f(x) = 0$ homogeneous in $(x_0, x_1)$ without affecting the degree of the equation.

In the plane, the homogeneous coordinates of a point $P$ specified in Cartesian coordinates by $(x, y)$ are three real numbers $(x_0, x_1, x_2)$, not all zero, for which $x = x_1/x_0$, $y = x_2/x_0$. A line in Cartesian coordinates is given by the linear equation $Ax + By + C = 0$. In homogeneous coordinates this becomes the homogeneous linear equation

$$Ax_1 + Bx_2 + Cx_0 = 0.$$  \( \text{(124)} \)

The line at infinity has equation $x_0 = 0$ and is thereby treated on a similar footing to other lines; for example, the $x$ and $y$ coordinate axes have equations $x_2 = 0$ and $x_1 = 0$, respectively. All these ideas generalize to higher dimensions.

In addition to providing a framework in which analytical calculations can be developed, the homogeneous coordinate system offers a simple setting for duality. For example, given Eq. (124), the three numbers $(A, B, C)$ can be viewed as homogeneous coordinates for a line in the plane. Then, coordinate statements about a point are expressed in terms of three numbers $(x_0, x_1, x_2)$, while statements about a line are expressed in terms of a dual set of three numbers $(A, B, C)$. A single equation thus represents a line or a point depending on which three numbers are considered variables.

Any set of three coordinates $(\tilde{x}_0, \tilde{x}_1, \tilde{x}_2)$, obtained from the homogeneous coordinate system $(x_0, x_1, x_2)$ in the plane by an invertible linear transformation

$$\tilde{x}_j = A_{jk}x_k$$  \( \text{(125)} \)
3.3 Group of Projective Transformations

Instead of being taken as a change of coordinates for a fixed point \( P \), the linear transformation (125) can be interpreted as a mapping from a point \( P \) at \((x_0, x_1, x_2)\) to another point \( \tilde{P} \) at \((\tilde{x}_0, \tilde{x}_1, \tilde{x}_2)\). This provides a mapping of the projective plane onto itself. Such mappings form a group \( G \) called the group of projective transformations for the plane. Similarly, groups of projective transformations can be introduced for higher-dimensional cases.

According to the so-called Erlangen program, projective geometry can be viewed as a study of properties of figures that are invariant under the action of \( G \). Various other geometries can be obtained by requiring invariance under a subgroup of \( G \). They include the regular Euclidean geometry, as well as affine geometry and the non-Euclidean elliptic and hyperbolic geometries.

Extensions of projective geometry to fields other than the real numbers exist. For example, complex projective geometry is defined over the complex numbers. The field may be finite or even noncommutative (see Algebraic Methods, Sec. 2). For example, a finite geometry in the plane called PG(2, 5) can be constructed using 31 points and 31 lines, with six points on each line and six lines through each point. Details of these generalized projective geometries may be found in the references at the end of this article.

4. Algebraic Geometry

Algebraic geometry involves the study of mathematical objects called varieties, which are generalized curves, surfaces, and hypersurfaces. The subject has several levels of abstraction, in each of which the precise meaning of the word variety is different. For the purposes of this article a relatively simple level of sophistication suffices, in which a variety can roughly be viewed as the solution to a set of polynomial equations for variables in a space. Note, however, that the modern definition of variety is considerably more abstract. It uses a branch of mathematics called the theory of schemes, about which more can be found in the references at the end of this article.

This section presents a few simple notions of algebraic geometry in the framework of polynomial equations. The discussion refers to several concepts (e.g., field, polynomial ring, rational functions) that are defined and described in the article Algebraic Methods.

4.1 Affine Varieties

Here is a more precise definition of one important type of variety. Consider an algebraically closed field \( F \). An \( n \)-dimensional affine space \( \mathbb{A}^n \) over \( F \) is defined as the set of points specified by the coordinates \((f_1, ..., f_n)\) with \( f_j \in F \). Denote by \( F[f_1, ..., f_n] \) the polynomial ring in \( n \) variables over \( F \). An affine variety \( V \) is a subset of \( \mathbb{A}^n \) given by the common zeros of a set \( S \) of polynomials in \( F[f_1, ..., f_n] \). If \( S \) contains only one polynomial, \( V \) is called an affine curve for \( n = 2 \), an affine surface for \( n = 3 \), and an affine hypersurface for \( n > 3 \).

A subset of \( V \) satisfying the definition of a variety is called a subvariety. If \( V \) is the union of two subvarieties, it is called reducible; otherwise, it is irreducible. For example, an irreducible affine curve is one for which the defining polynomial is irreducible (i.e., cannot be factored). An irreducible component of \( V \) is defined as a maximal irreducible subvariety of \( V \). One result in this subject is that any variety \( V \) can be written uniquely as the union of finitely many distinct irreducible components.

Starting with a variety \( V \), a sequence of irreducible varieties can be constructed such that each member of the sequence is a subvariety of the preceding one. This sequence is of finite length, and the number of subvarieties in it is called the dimension of \( V \).

The unions and finite intersections of subvarieties of \( V \) are also subvarieties. This means that the complements of the subvarieties of \( V \) can be used as the collection \( T \) of subsets for a topological space (see Sec. 2.1). Therefore, \( \mathbb{A}^n \) and hence also \( V \) can be endowed with a topology, called the Zariski topology. This topology is not Hausdorff but, unlike the usual Hausdorff topology on \( C^n \), it is defined for all affine varieties over \( F \).
4.2 Projective Varieties

Several extensions of the notion of affine variety to more general varieties exist. One generalization uses an approach similar to that taken in the construction of differentiable manifolds: The meaning of variety is extended to include objects constructed by patching together affine varieties. This generalization then looks locally like an affine variety but globally is different. An important result in algebraic geometry is that certain subsets of projective spaces form varieties of this sort, called projective varieties.

An $n$-dimensional projective space $P^n$ over $F$ can be introduced as the set of points specified by the homogeneous coordinates $(f_0, f_1, ..., f_n)$ with $f_j \in F$ not all zero, subject to the restriction that two such sets of homogeneous coordinates related via a single nonzero constant $c \in F$ as

$$(f_0, f_1, ..., f_n) = (cf_0, cf_1, ..., cf_n)$$

specify the same point (cf. Sec. 3.2). Denote by $H[f_1, ..., f_n]$ the ring of homogeneous polynomials in $n$ variables over $F$. A projective variety $V$ is a subset of $P^n$ given by the common zeros of a set $S$ of polynomials in $H[f_1, ..., f_n]$. If $S$ contains only one polynomial, $V$ is called a projective curve for $n=2$, a projective surface for $n=3$, and a projective hypersurface for $n>3$.

4.3 Classification

The ultimate aims of algebraic geometry are the classification and characterization of varieties. These are difficult and unsolved problems in the generic case. To attack the classification problem, a means of relating varieties to one another is needed. This is provided by the notion of a rational map.

A rational map $f:V \to A^n$ from an affine variety to $n$-dimensional affine space is basically a set of $n$ rational functions $f_j$. The domain of $f$ is by definition taken as the union of the domains of the $n$ functions $f_j$. A rational map $f:V_1 \to V_2$ between two affine varieties $V_1 \subset A^n$ and $V_2 \subset A^m$ is defined to be a rational map $f:V_1 \to A^m$ such that the range of $f$ lies in $V_2$. If the map $f$ also has a rational inverse, it is called a birational equivalence.

The classification problem is approached by seeking a classification up to birational equivalence. Ideally, this means providing discrete and/or continuous numerical quantities that are invariant under birational equivalence and that characterize inequivalent varieties. Then, given a birationally equivalent set of varieties, a standard subset with desirable features (e.g., no singularities) can be sought and a classification attempted. Finally, one can seek some means of measuring the deviation from this standard subset of the remaining members of the birational-equivalence class.

An example is provided by the special case of the algebraic curves over $F$. For these varieties, a discrete quantity called the genus $g$ can be introduced, which is a nonnegative real number that is invariant under birational equivalence. Curves with $g=1$ are sometimes called elliptic curves. For each nonzero $g$ the birational-equivalence classes can be labeled by a one-dimensional continuous variable if $g=1$ and by a $(3g-3)$-dimensional set of continuous variables otherwise. The continuous variables are called moduli. They also form an irreducible variety, called moduli space, that can in turn be studied with the methods of algebraic geometry.

If the field $F$ is the field $C$ of complex numbers, the resulting algebraic curves are the Riemann surfaces. A curve with $g=0$ is topologically a sphere, while one with $g=1$ is topologically a torus. The set of Riemann surfaces plays an important role in string theories (see Sec. 2.10). For example, at a particular order in perturbation theory for a scattering process the string world sheet is topologically a Riemann surface with punctures for the incoming and outgoing strings. The methods of differential and algebraic geometry play a significant role in the evaluation of such contributions to the scattering amplitudes.

GLOSSARY

Considerations of space prevent an extensive glossary being provided for this article. Instead, the following is a list incorporating important concepts together with the number of the section in which the concept appears.

Abscissa: See Sec. 1.1.
Adjoint Exterior Derivative: See Sec. 2.4.
Affine Space: See Sec. 4.1.
Affine Variety: See Sec. 4.1.
Almost Complex Structure: See Sec. 2.8.
Antisymmetric Symbol: See Sec. 2.4.
Antisymmetric Tensor: See Sec. 2.3.
Applicative: See Sec. 1.5.
Atlas: See Sec. 2.1.
Base Manifold: See Sec. 2.5.
Betti Number: See Sec. 2.9.
Blanch Identity: See Sec. 2.6.
Birational Equivalence: See Sec. 4.3.
Brianchon's Theorem: See Sec. 3.1.
Bundle of Frames: See Sec. 2.10.
Cartan Exterior Algebra: See Sec. 2.4.
Cartesian Coordinates: See Sec. 1.1.
Characteristic Class: See Sec. 2.9.
Chart: See Sec. 2.1.
Circle: See Sec. 1.2.
Closed Form: See Sec. 2.4.
Coclosed Form: See Sec. 2.4.
Coexact Form: See Sec. 2.4.
Complex Manifold: See Sec. 2.8.
Complex Structure: See Sec. 2.8.
Cone: See Sec. 1.5.
Conic Section: See Sec. 1.2.
Connection: See Sec. 2.6.
Contraction: See Sec. 2.2.
Contravariant Components: See Sec. 2.2.
Coordinate Basis: See Sec. 2.2.
Cosine: See Sec. 1.3.
Cotangent Space: See Sec. 2.2.
Covariant Components: See Sec. 2.2.
Covariant Derivative: See Sec. 2.6.
Cross Section: See Sec. 2.5.
Curvature Form: See Sec. 2.6.
Curvilinear Coordinates: See Sec. 1.4.
Cylinder: See Sec. 1.5.
Cylindrical Coordinates: See Sec. 1.5.
De Rham Cohomology: See Sec. 2.9.
Desargues's Theorem: See Sec. 3.1.
Differential Forms: See Sec. 2.4.
Directrix: See Sec. 1.2.
Discriminant: See Sec. 1.2.
Dolbeault Operator: See Sec. 2.8.
Dual Basis: See Sec. 2.2.
Dual Vector Space: See Sec. 2.2.
Duality, Projective: See Sec. 3.
Eccentricity: See Sec. 1.2.
Einstein Summation Convention: See Sec. 2.2.
Ellipse: See Sec. 1.2.
Ellipsoid: See Sec. 1.5.
Embedding: See Sec. 2.3.
Erlangen Program: See Sec. 3.3.
Euclidean Space: See Sec. 1.5.
Exact Form: See Sec. 2.4.
Exterior Derivative: See Sec. 2.4.
Fiber: See Sec. 2.5.
Fiber Bundle: See Sec. 2.5.
Finite Geometry: See Sec. 3.3.
Focus: See Sec. 1.2.
Fundamental Theorem of Projective Geometry: See Sec. 3.1.
Genus: See Sec. 4.3.
Geometric Quantization: See Sec. 2.10.
Group of Projective Transformations: See Sec. 3.3.
Harmonic Form: See Sec. 2.4.
Hausdorff: See Sec. 2.1.
Hermitian Metric: See Sec. 2.8.
Hodge Decomposition Theorem: See Sec. 2.9.
Hodge Star: See Sec. 2.4.
Homogeneous Coordinates: See Sec. 3.2.
Hopf Map: See Sec. 2.7.
Horizontal Lift: See Sec. 2.6.
Horizontal Subspace: See Sec. 2.6.
Hyperbola: See Sec. 1.2.
Hyperboloid: See Sec. 1.5.
Hypersphere: See Sec. 1.5.
Hypersurface: See Sec. 1.5.
Incidence: See Sec. 3.
Index Theorem: See Sec. 2.9.
Irreducible Variety: See Sec. 4.1.
Kähler Metric: See Sec. 2.8.
Kepler Problem: See Sec. 1.6.
Lie Algebra: See Sec. 2.5.
Lie Group: See Sec. 2.5.
Line At Infinity: See Sec. 3.
Manifold: See Sec. 2.1.
Maxwell Equations: See Sec. 2.7.
Method of Coordinates: See Sec. 1.
Metric: See Sec. 2.3.
Modular Parameter: See Sec. 2.8.
Moduli: See Sec. 4.3.
Morse Theory: See Sec. 2.9.
Neighborhood: See Sec. 2.1.
Ordinate: See Sec. 1.1.
Pappus's Theorem: See Sec. 3.1.
Parabola: See Sec. 1.2.
Paraboloid: See Sec. 1.5.
Parametric Representation: See Sec. 1.5.
Pascal's Theorem: See Sec. 3.1.
Pencil: See Sec. 3.1.
Plane Analytic Geometry: See Sec. 1.5.
Plane Polar Coordinates: See Sec. 1.4.
Point at Infinity: See Sec. 3.
Potentials: See Sec. 2.7.
Principal Bundle: See Sec. 2.5.
Projective Geometry: See Sec. 3.
Projection Map: See Sec. 2.5.
Projective Transformation: See Sec. 3.1.
Projective Variety: See Sec. 4.2.
Quadric: See Sec. 1.5.
Riemann Surface: See Secs. 2.8, 2.10, 4.3.
Riemannian Manifold: See Sec. 2.3.
Scalar Field: See Sec. 2.3.
Sine: See Sec. 1.3.
Slope: See Sec. 1.1.
Solid Analytic Geometry: See Sec. 1.5.
Sphere: See Sec. 1.5.
Spherical Polar Coordinates: See Sec. 1.5.
Spherical Triangle: See Sec. 1.5.
Structure Group: See Sec. 2.5.
Submanifold: See Sec. 2.3.
Symmetric Tensor: See Sec. 2.3.
Symplectic Form: See Sec. 2.10.
Tangent: See Sec. 1.3.
Tangent Space: See Sec. 2.2.
Tensor: See Sec. 2.3.
Tensor Field: See Sec. 2.3.
Topological Space: See Sec. 2.1.
Variety: See Sec. 4.
Vector: See Sec. 2.2.
Vector Field: See Sec. 2.3.
Vertical Subspace: See Sec. 2.6.
Wedge Product: See Sec. 2.4.
Zariski Topology: See Sec. 4.1.

Further Reading


Frampton, P. (1987), Gauge Field Theories, Reading, MA: Benjamin Cummings.

**GEOMETRICAL OPTICS**

See OPTICS, GEOMETRICAL