Geometries for Spatial Data

The aim of this discussion is to introduce methods of positioning objects in spatial reference systems, local and global. This includes questions of measuring distance and facilitating access to spatial data, as well as different approaches to the representation of artificial and natural objects, especially by contrasting the Euclidean, topological and fractal geometries.

Different Geometries

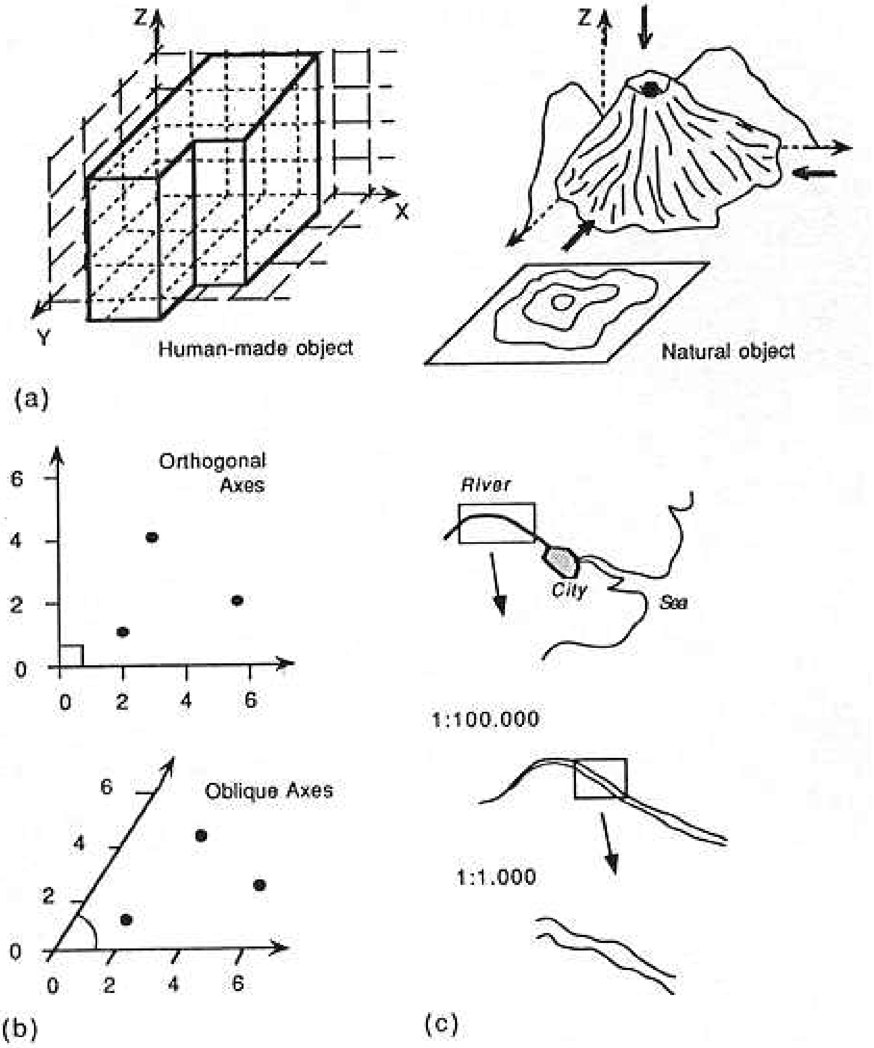
In the context of spatial information systems there are needs for dealing with many aspects of space:

1. two or three dimensions;
2. planar or non-planar situations;
3. continuous or discrete referencing;
4. isotropic or anisotropic conditions;
5. forms and distances that change over time;
6. measurements of qualitative or quantitative properties of space;
7. smooth or rough objects.

Spatial indicates pertaining to space, the void in which material elements exist. Geometry is that branch of mathematics which deals with spatial quantities and the shape of spatial forms. **Shape** (or **configuration)** refers to the structure of those forms. **Form** is the mode of arrangement related to function, or the appropriateness and effectiveness of purpose (like the hexagonal pattern of the bee's honeycomb results from the application of both a principle of economy of subdividing space and a principle of stability of structure achieved by the hexagon shape).

Descriptive geometry is the measurement of properties of objects in space or relationships between objects, that enables the true lengths, angles, lines of intersection and other elements to be determined by graphical means. **Topological** geometry relates to the relationships of components of form. **Fractal** geometry deals with the fragmentation and non-smoothness of form and the dimensionality of objects. **Computational** geometry refers to those concepts, principles, and tools used in the numeric, in contrast to the algebraic practice of geometry.

Geometry provides many tools, and formalisms to work with spatial concepts. Consider first that material objects have length, width, and height, and can be placed in a framework, a **spatial reference system:**



that allows measurements of those three fundamental properties (a). Generally, the three principal axes are orthogonal (at right angles), but in some applications of statistical methods, entities may be viewed in oblique coordinate space (b), reflecting the correlation. Some material objects may be clearly demarcated. The line representing the edge of a (wide) highway may be appropriate for a viewing position from a long distance, but a walk along this 'fine line' reveals that it is very irregular at this scale of observation (c). The **field** view of spatial objects has this concept of indeterminancy of boundaries, and the **geometry of fractals** provides some ways to represent **the unevenness** of objects.

Earth scientists may not always be able to use the rigid planar assumptions of solid geometry, and so use coordinate systems with curvature. Map-making, analogue or digital, automated or manual, has long been practised with the aid of conversions from curved to plane surfaces. **Map projections** were devised to allow representation of the spherical world on flat paper.

Material objects in the one, two, or three spatial dimensions have different properties. Shape implies the use of absolute coordinate values as a basis for measuring length, perimeter and subsequent variation from a regular figure like a circle; but **many** spatial relationships utilize topological properties of connectivity, containment and contiguity. Spatial information systems are often structured on the basis of explicit topology, while some kinds of representations, for example, map models, show those properties implicitly - the human eye is very good at seeing relationships between places and phenomena shown cartographically.

Positioning Objects in Spatial Referencing Systems

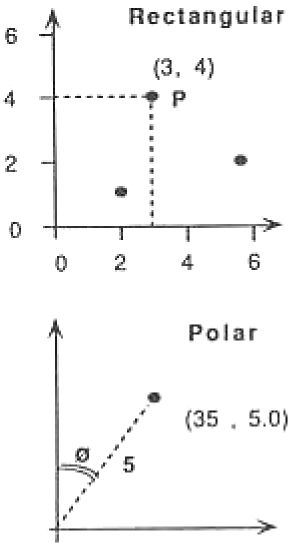
Positioning objects, fundamental activities of fields as different as geodesy or land surveying and thematic cartography, referring to the real world and map model domains, respectively, involve considerations of:

1. The geometric character of the reference system;
2. Measurement metrics;
3. Types of reference system: Cartesian or polar;
4. Nature of the origin;
5. Discrete or continuous references.

For most purposes, people relate to the length and width **dimensions,** although occasionally, for some people, the height dimension, associated with mountain climbing and skyscraper elevators can be painfully obvious. Today many spatial information systems have data for only the *{x, y, a),* not the z dimension, reflecting origins in natural resources mapping and management, the emulation of map overlay modelling techniques, and, originally, difficulties in programming computers to work with data for three dimensions. Often the z variable is treated as an attribute of objects, not a positioning coordinate.

Continuous space referencing

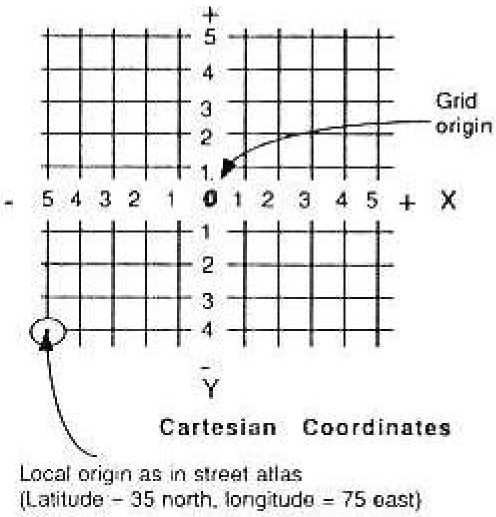
Within the two- or three-dimensional void, referencing may be Cartesian or polar:



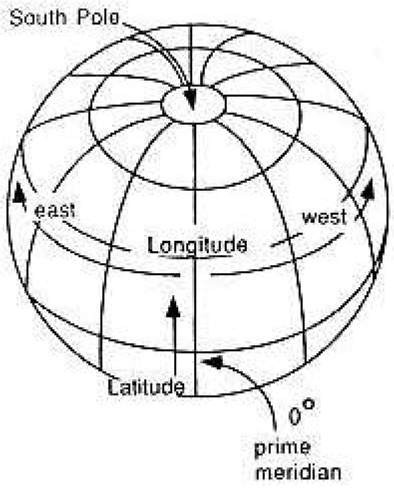
(a)

This **polar** type of reference is useful for problems dealing with direction of travel, as a personal reference system, or for mapping phenomena like journey to work or migration that use a spatial dyad as a basic unit. Radar based displays or ground surveys are other examples of the use of distance-direction combinations.

Positioning objects in space generally uses Euclidean (rectangular or plane coordinates) geometry, measuring distances from a specified origin , either global (tied in to the earth) or local, as is generally used for city street atlases:

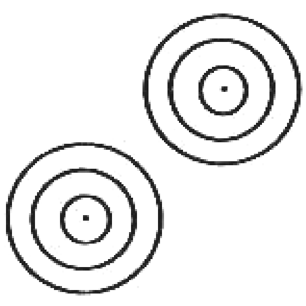


Greenwich Meridian or the Equator are used for global referencing:

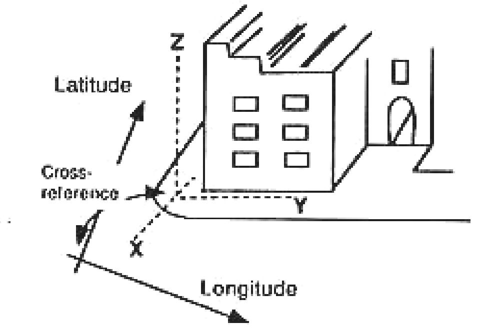


Information systems for the whole world generally need to tie placements into the geographic reference system of latitude and longitude. Accurate measures for two or three control points can be used to translate from local into geographic coordinates; and spherical coordinate systems may be used for the earth rather than projecting to planar coordinate systems.

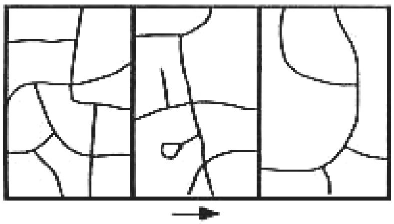
Moreover, **origins** may not be singular, may not be stationary in space, and may be at different elevations:



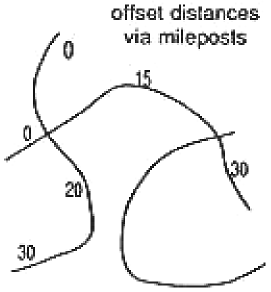
(a) Two air traffic control reference systems



(c)



(b) Three successive maps on an in-vehicle navigation aid monitor



(d)

Studies of flow phenomena, channelled or not, may require several origin points, for example, the individual towers for air traffic control systems, each usually associated with a polar map (a). Navigation aids generally work from the current origin of a traveller, meaning that instructions for wayfinding, or maps, are based on a succession of reference points (b). Individual entities may have more than one reference system. For example, a building may be seen from the perspective of its own three dimensional coordinate axes, and by reference to an external frame, perhaps geographic coordinates (c). Civil engineers or transportation planners usually refer to the location of objects or events on highways as an **offset distance** from a specific origin, perhaps the beginning of a highway, an intersection, or a national boundary (d).

The referencing may not be consistent with identification of objects. For example, a piece of highway with particular properties like the number of lanes or surface condition can be demarcated by how far from an origin the combined attribute of lanes and surface changes. However, a topologically based definition of a one-dimensional object will not recognize that segmentation into pieces based on the attributes.

But, alas, spatial objects move, like the continents which positions are quite different due to plate tectonics. In the modern world the distance between North America and Europe increases every year by a few centimetres.

Another problem is the existence of earthquakes, implying the movement of pieces of land, and therefore, of landmarks. The question arises for such events, as to whether we consider the landmark coordinates as fixed, or moved, do we have to change the coordinates in a spatial information system?

Particular entities may have their coordinates obtained from a variety of sources. Ground surveys and mensuration have traditionally been used to establish a control system, or, if it meets certain conditions of accuracy and precision, a **geodetic reference system.** The control system is generally regarded as a set of physical objects, called monuments, put into the ground, and the data describing their positions, which in combination, form a basis for establishing the position of entities on the earth's surface. The geodetic system provides a universal basis for what would otherwise be separate measures. For instance, the principal global meridian (zero degrees) is clearly marked at Greenwich Observatory in London, and Sweden maintains a set of monuments marking the location of the polar circle.

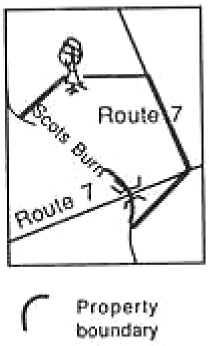
Moreover, the officially accepted **horizontal datum,** the standard global framework for establishing latitude and longitude, may change in time. For example, the change in the USA from the North American Datum of 1927 to a new one, North American Datum of 1983, caused shifts of up to 100 metres in some coordinates as a result of going to a new spheroid to represent the globe from the ellipsoid used in 1927 which was designed to fit only the shape of the conterminous United States.

Referencing for discrete entities

Discrete referencing, the means of establishing the location of separate entities, other than by coordinates in continuous space, has several possibilities:

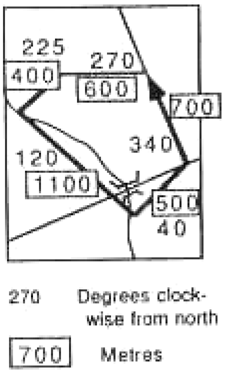
1. Narrative descriptions;
2. Street addresses;
3. Blocks of earth space.

In many parts of countries legal specifications of property have historically been given as a narrative description of boundaries. Descriptions would identify landscape objects, and possibly give some indication of distance. Reference is made to supposedly permanent objects:

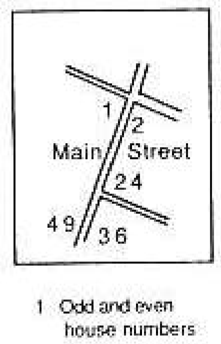


like the large oak tree or the bridge over the river. Descriptions are as precise as the objects used as landmarks or boundary paths, and as practical only to the extent that these features still exist, and are recognizable, and have not moved over the years.

Modern equivalents use coordinate geometry based direction and distance measures:



Street addresses provide an almost universally adopted mechanism of locating discrete properties, such as a dwelling or a plot of land:



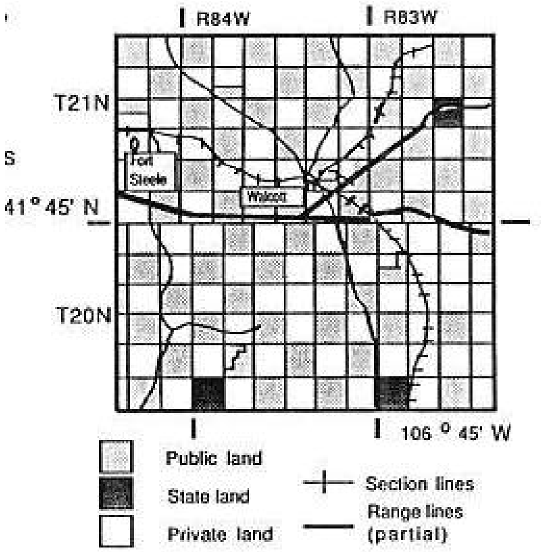
The United States Census Bureau DIME system, incorporated street numbers for left and right sides of streets into the basic set of data for the line segments, as well as the indication of the area units (city blocks) on either side of the lines. Street addressing systems are, though, not without problems. Difficulties can arise for numbering streets bordered on one side by a river, and when properties are subdivided. Street address based referencing has its own practical utility because many data, particularly administrative or taxation type, are recorded by street address. While there may be some shortcomings in precision of location relative to latitude and longitude, and inconsistencies or duplications in naming streets or assigning numbers, the street addressing system is widely accepted throughout the world.

While there are many possible criteria for creating areal units for referencing, worldwide the most common are:

- zones of postal services,
- administrative units,
- statistical reporting units for census purposes.

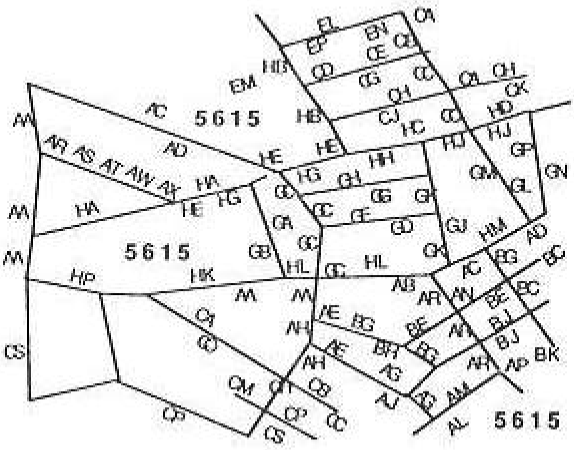
As such, these **irregular tiles,** are useful for accumulating aggregates by particular attributes, and for retrieval purposes using names or other identifiers for those units, although they may lack positional precision.

A **regular tiling** system, that is a subdivision of space into areas equal in shape and size, may also be used for reference to places on the earth. Most street atlases provide tables of street names and a number or code letter for the row and column in which the street can be found:

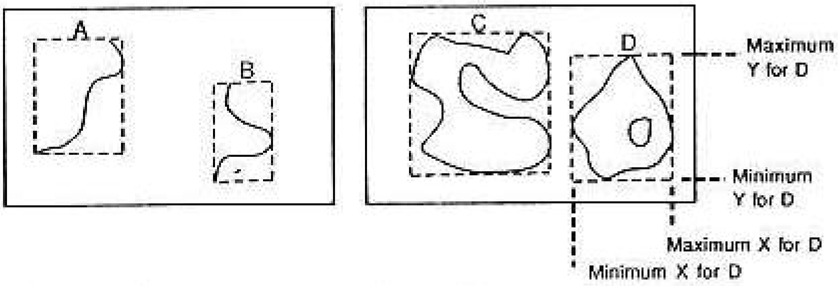


Spatial information system databases created from existing map sheets may use the particular sheets for reference purposes or for simplifying management of a large database.

The postal code, one of few universal reference systems useful for spatial data, is ostensibly associated with areal units. In reality, the basic unit is the postal delivery route, a concept associated with streets, linear features. Many countries have postal codes reflecting several levels of organization. For example the Netherlands, similarly to other countries, have postal code hierarchical, referring to large territorial divisions, then streets, and routes. The last-named often follow only one side of a street, and often do not break at road junctions:



For several purposes, especially for the access to and retrieval of spatial data, line and area entities are referred to by some limiting conditions. Most usually these are the coordinates for the ends of lines or the furthest **extents** of lines or polygons in orthogonal x- and y-axes. The extents established by the minimum and maximum coordinates establish the corners of boxes known as **minimum bounding rectangles.** This means that entities will often be represented by two kinds of data - the specifics of their location and/or form, and their positional extent:



(a) Lines (b) Polygons

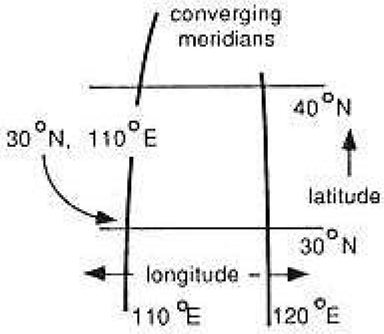
By way of summary, the term **locator** is being used to refer to any device for indicating relative or absolute position in some spatial reference system. Locators may be Cartesian or polar coordinates, distances from specified but arbitrary origins measured along lines, through space, in one or more dimensions; they may be discrete addresses, or special locational codes known as Peano and Hilbert keys.

Global Reference Systems

The domain of three-dimensional space is very important for the development of global databases such as for environmental management and research into global warming and its impact.

Global referencing

The basic concept of global referencing consists of using a set of two coordinate values for the position on the real (curved, but flat) surface:



Latitude is generally measured north and south from the equator by positive and negative values as a y-axis; longitude is measured east or west by positive and negative numbers of x from an arbitrary prime meridian, usually the Greenwich Meridian but not necessarily. Indeed, in the 1880s there was much political activity before agreement was reached to use the line passing through London, and several countries today have their own prime meridians for mapping purposes.

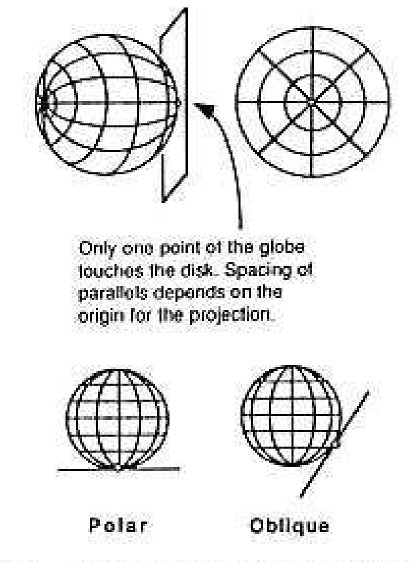
Currently there exists in space a set of satellite based electronic signal generators that can be used to very precisely and accurately pinpoint a location on the earth's surface. This **global positioning system** (GPS) developed by the United States government, allows accurate spatial referencing in continuous space to a precision of just a few metres, or about the width of an average city street, or even better as electronic equipment improves.

Map projections

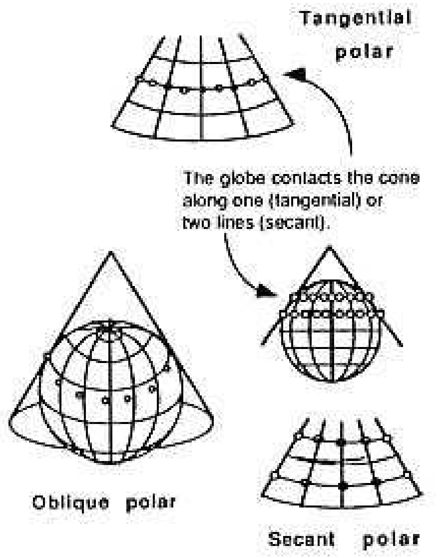
World maps in an atlas provide a spatial referencing system, but the same pieces of the earth can look different according to the map projection used for the map. A map projection is a device for **producing** all or part of a sphere on a flat sheet. Some projections preserve distance or true direction, some maintain correct shapes, and others preserve the property of areal size. Map projections can be studied from the point of view of:

1. The form of the surface used for the projection.
2. The particular viewing origin and the standard points and lines used.
3. Spatial properties, preserved or distorted.
4. The number of points used to transform from a sphere to a flat surface.
5. The formulae used in mapping from a sphere to a noncurved surface.

In manner of construction projections have been classified first of all based on the form of the piece of paper theoretically wrapped around or otherwise touching the globe, principally the three types of cylinder, cone, and disk:

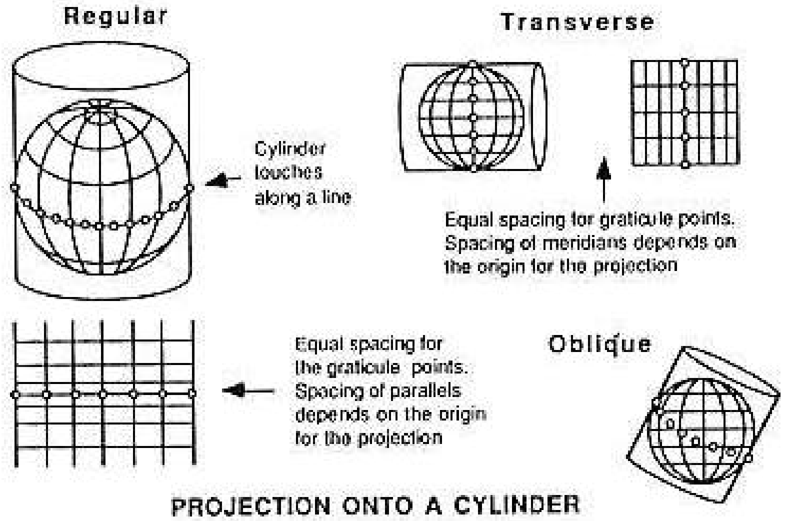


(a) PROJECTION ONTO A DISK



(b) PROJECTION ONTO A CONE

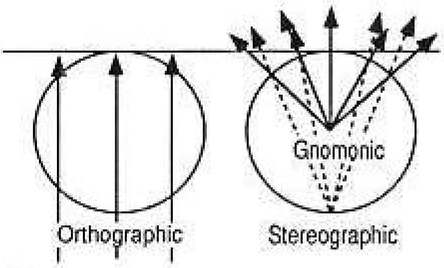
The chosen graticule and earth features are mapped from the sphere to a disk tangential to the globe at one point, possibly a pole (a), or any other point suitable as the origin. The cone is tangential to the surface along the path of one circle or two circles, if the secant projection is being used (b). The cylinder object may be wrapped around the earth in different locations, but touches only along a circle (c):



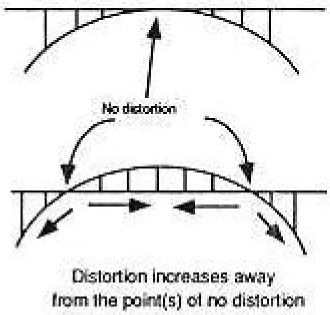
(c) PROJECTION ONTO A CYLINDER

In all cases there is a single line of points (or a single point in the case of the azimuthal surface) for which coordinate values are correct to scale, as suggested by the small circles in the figure.

The projection surfaces may also cut the earth surface; such **secant projections** will have two principal axes, as shown for the conic projection, thereby reducing average distortions compared to the tangential form. The secant type is more common for conic rather than cylindrical or azimuthal projections. Cylindrical projections that are oriented to meridians rather than parallels are known as transverse projections.



(d)



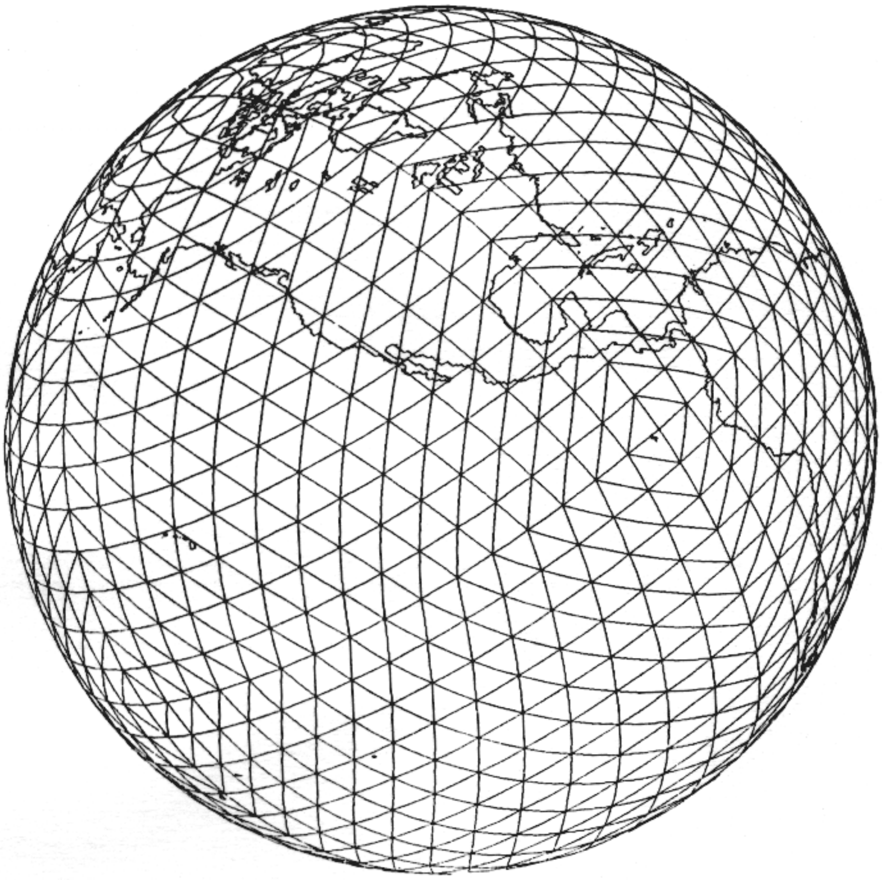
(e)

Different projection origins may be used. Illustrated only for projecting to a disk (d), we can imagine a base of infinity, represented by parallel projection lines, or an origin at the pole opposite to the centre of the azimuthal projection, or located at the centre of the globe. The particular point of origin will cause the graticule of equal degree increments for latitude to be represented by different spacing, as shown.

Planar projections may be further classified as to the particular properties preserved. Not all properties can remain undisturbed in a single projection. Some guarantee one factor; others are compromises. The existence of many particular projections is not surprising, given the four different properties: distance, direction, shape, and area, that may be preserved, and their many uses. Engineers, meteorologists and **navigators** are particularly interested in distances and compass directions; ecologists, geologists and geographers often need area preserving maps.

Geometric structure is apparent in the three-dimensional world. The faceted spherical surface of a football or basketball consists of mostly hexagonal patches, but also pentagonal patches set at predetermined positions. It can be shown mathematically that only hexagons, or their equivalent, triangles, will not work if curved lines are not allowed in the graticule or wireframe model. A quarternary triangular mesh for a globe, a set of triangles initiated from an octahedron model has recently been proposed as an effective global reference system. The intent is to have:

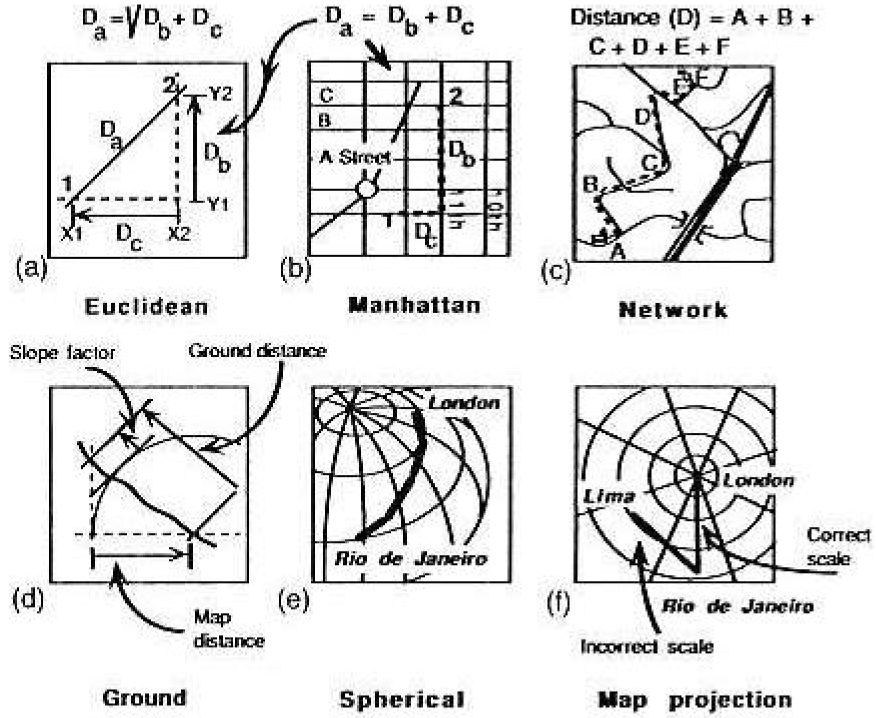
1. Basic units almost equal in size;
2. Basic units almost equal in shape.



The Distance

The distance is a building block so important in spatial information systems that it deserves some extra treatment. There are many aspects to distance:

- Metric for measurement
- Type of geometry
- Planar or spherical
- Simple or composite
- Singular or accumulated
- Isotropic or anisotropic
- Distance in a graph
- Number of dimensions
- Precision and accuracy

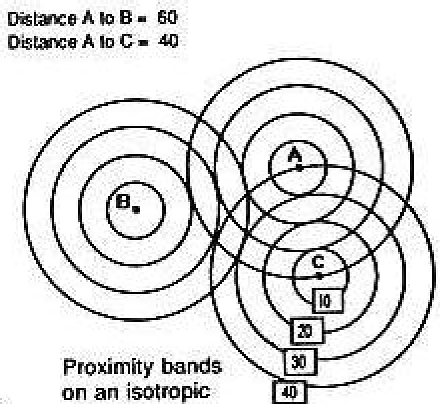


Distances in planar Cartesian coordinate systems use the data for coordinate pairs, often computing the straight line distance as the length of a diagonal of a right angled triangle (a). However, there are other possibilities, of which the most intriguingly named is the Manhattan, taxicab or city block distance (b). This computes distance by summing the lengths of the sides of the triangle. Algebraically, it is based on absolute distances rather than on the squared distances for the straight line between two points based on the Pythagorean theorem.

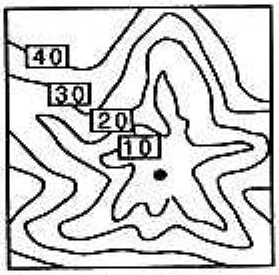
The distance measures may be made over a surface or may be computed for the assumed real channels of movement. Real **network distances** or driving routes (c) are almost always greater than the direct straight line distance.

Determination of accurate **non-planar distances** requires elevation coordinates so that measures may recognize gradients of the earth's surface. On-ground distances will also be underestimated if the terrain conditions are not considered (d). False data or impressions will be obtained for global distances if wrong map projections are used (e and f).

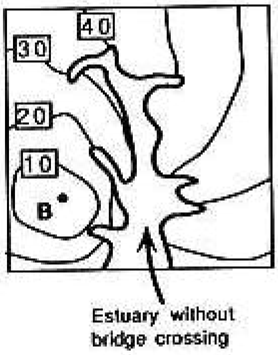
In the human world the effort to move over the earth's surface is often thought of in terms of cost, either **in** time or money. Transportation planners, urban geographers, and others have developed many measures of **spatial impedance,** and cartographers have devised various techniques to represent the relative positions of places in different metrics. Indeed unless a single composite arithmetic measure is designed for combining different metrics, it is difficult to deal with the inherent conflict of placement of points according to different measures.



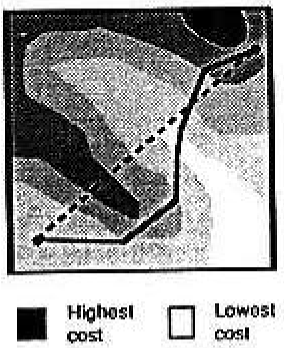
(a) Proximity bands on an isotropic surface



(b) Travel time isolines from central point in an urban area



(c) Distances affected by anisotropic conditions



(d) Minimum cost path over terrain divided into travel cost zones.

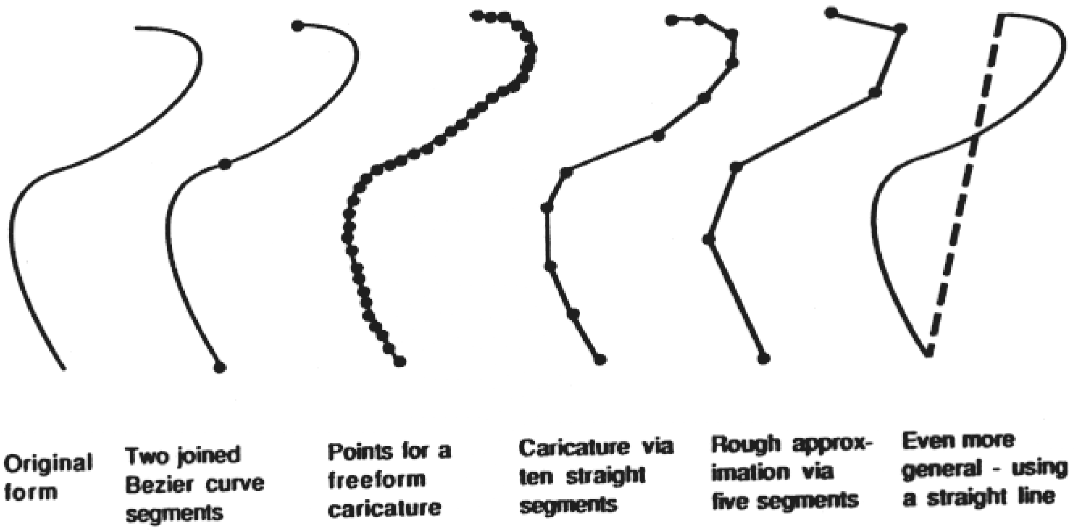
Sometimes distance is dealt with as **proximity zones,** generally for simplicity assuming an azimuthal base (a). **Accumulations** of distance from one point to reach all others in a set are useful for producing maps of accessibility (b). In the anisotropic case, **barriers,** permeable or impermeable, create distortions to the otherwise even patterns (c). Permeable barriers allow some movement, perhaps representing a real condition of limits to the number of vehicles able to pass a particular point or journey on a selected route. They are similar to aggregations of several weighted segments, each associated with a chunk of terrain of different spatial impedance (d).

Coordinates And Splines: The Representation Of Lines

Not only are there interesting questions about distance, but there are practical issues about the representation of one- and two-dimensional spatial elements such as:

1. The amount of generalization in the representation of lines.
2. The mathematical form of the representation for lines.
3. The need for minimizing the amount of data.
4. The amount of regularity in the entity.
5. The accuracy and precision of representation.

Simply put, an irregular line joining fixed and known end points, can be assumed to be straight, as complete as the original data, or represented by some condition between these two extremes:



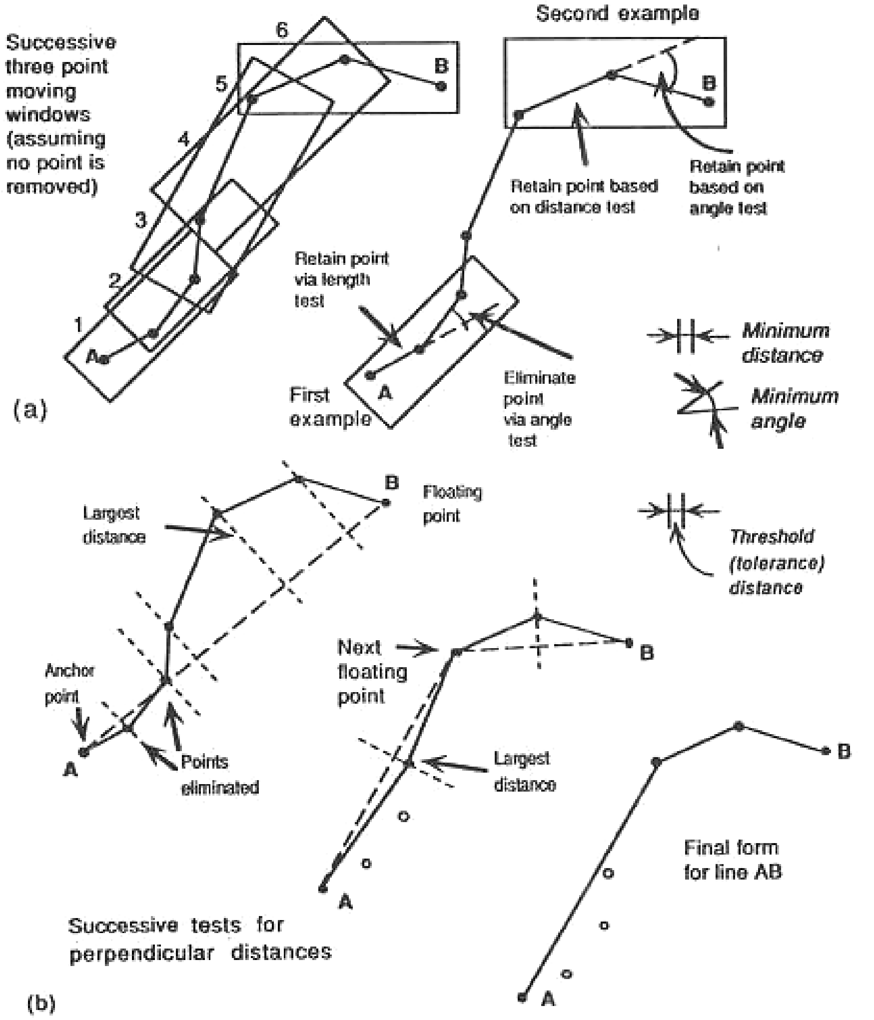
The reduction in the number of pieces of information needed to be able to draw the line can be effected in different ways, systematic or otherwise. Several aspects of generalization, with smoothing and line thinning are considered below.

Line simplification

Cartographers often use procedures of line simplification that systematically eliminate certain points from those originally recorded or captured for a linear feature. Ideally, a representation should both preserve the main element of shape, and recognize topological properties. That is to say, for a set of boundary lines it is important to use the topological junctions (line intersections) as inviolate points. Then the undulations and indentations in the shape of the line between the topological end points may be altered depending on how much of the original shape is important for a given purpose. Map users may feel more comfortable if the recognizable features are preserved, even if they are not necessary for a given specific analytical purpose.

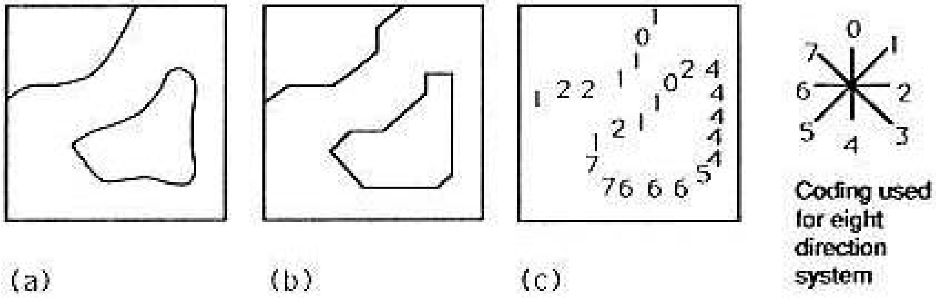
The simplest procedures for line generalization consist of retaining only a fraction of points at systematic intervals: for example, every fifth point is kept and all others discarded. This procedure does not guarantee shape preservation. Out of the very large number of procedures that try to preserve shape, especially the major turning points, just two are introduced to demonstrate certain basic principles.

First of all, for an already selected line feature, we can simplify by removing points that are too close to neighbours or have a small angle of separation between vectors. Either of these conditions may be examined separately, or they may be combined in one procedure and a three point moving window can be applied to a line from a topological (or arbitrary) origin. The first three vertices caught in the window are examined with regard to-two criteria (a). One is the distance of a vertex from the previous vertex if that distance is too short then the second point can be **removed.** The second criteria relates to the existence of turning points. If the angle between a projected line joining the first two points and the line from second to third is small, then the third point can be discarded. This process continues for every succeeding set of three points in the window.



Perhaps the most well known procedure, used in several commercial software systems, uses a slightly different approach to retain the main shape features of a line (b). The Douglas-Peucker procedure first joins the beginning and ending vertices of a line feature by a straight line, and then examines perpendicular distances to the individual vertices. Those that are closer than a selected threshold distance can be removed. The point furthest away is selected as a new end point for repetition of the process until there are no points closer to a line than the threshold.

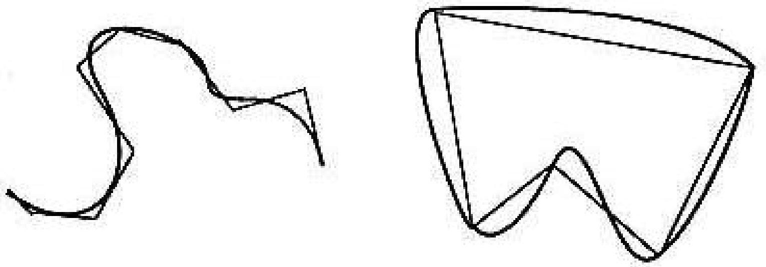
In addition to generalizing by point elimination techniques, lines may be represented as a series of compass orientations for successive straight line segments, known as **chain codes.** If lines are constructed on a regular arithmetic grid, the dominant direction may be coded using one of several cardinal directions:



More precise coding is possible with more directions. Actual coordinates are needed for only one point, either the beginning of a polyline or an arbitrary point for a closed line; and special techniques can be used for junctions. However, it is more difficult to compare or to apply alterations to lines generalized by this incremental type of encoding than it is to lines represented by coordinates.

Smoothed lines

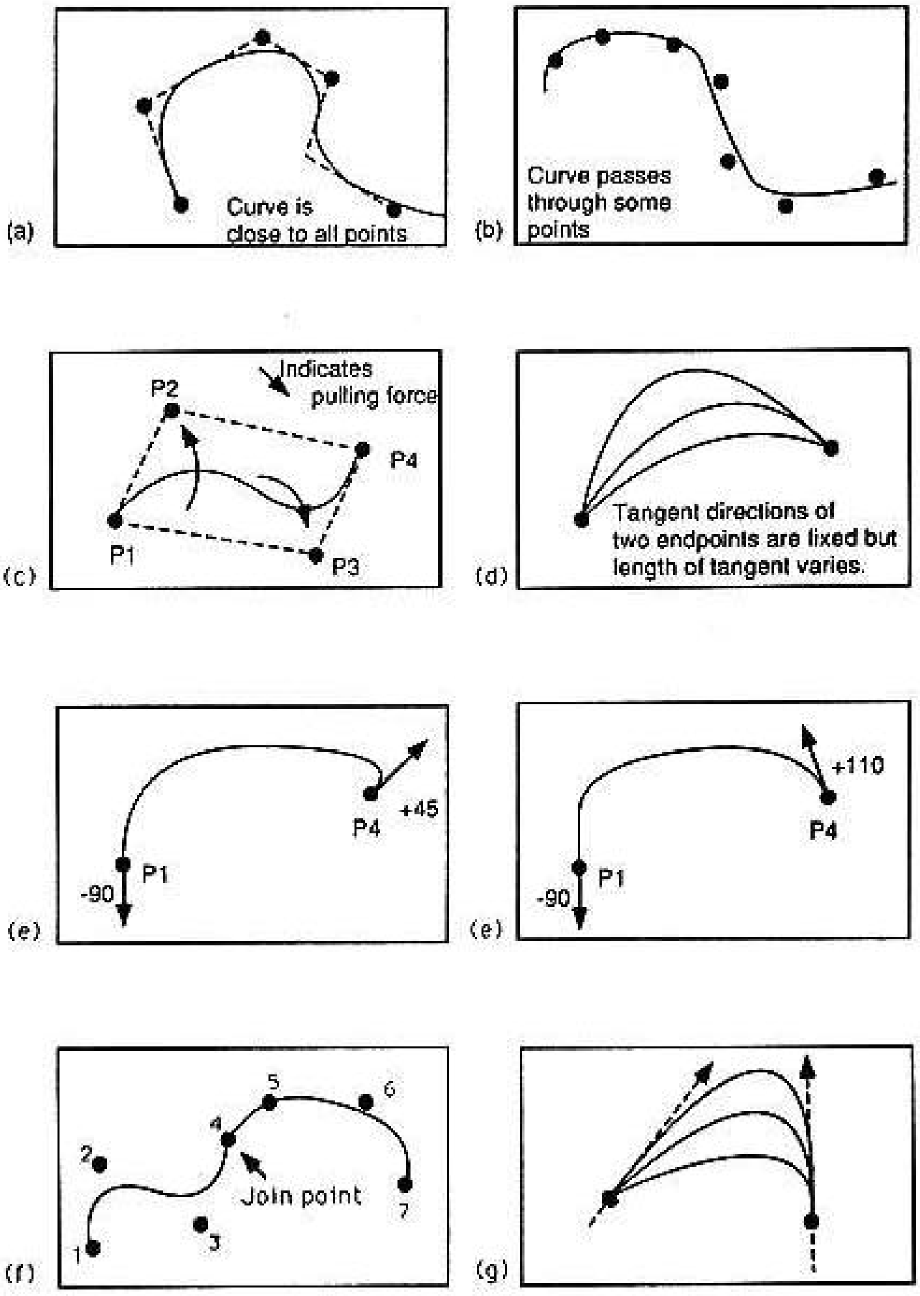
With the same objective of reducing the number of pieces of information, lines may be represented as a single curve, or as a combination of **smooth curves.** Essentially, a smooth curve is fitted, according to certain criteria, to a shape for an irregular polyline or polygon represented by a set of specific points, generally passing through all or possibly just a selection of the specified points (a). In the case of a line feature, the curve may cross intermediate segments on its route from one end to the other. In the case of a polygon, separate curve pieces generally are fitted to the distinct line segments (b):



(a) Polyline smoothing (b) Polygon smoothing

Important criteria that may be established for an ideal curve are:

1. It should pass through specific end (or mid) points.
2. It should be tangential to only one point on an individual line segment, or cross at only one point.
3. It should have no discontinuity where separate curves join when making up a series to represent a complex polyline or irregular polygon.
4. It should have compact mathematical representation.



Line smoothing may use statistical averaging techniques or mathematical equivalents to the French curve, the flexible drafting tool used by graphic artists and engineers. In practice, lines (or surfaces for the three-dimensional context) are often represented explicitly by curves from the polynomial family:

$$x(t) = a(t)x_1 + b(t)x_2 + c(t)x_3 + d(t)x_4,$$
$$y(t) = e(t)y_1 + f(t)y_2 + g(t)y_3 + h(t)y_4,$$
$$z(t) = p(t)z_1 + q(t)z_2 + r(t)z_3 + s(t)z_4$$

Functions from *a(t)* to *s(t)* are special cubic polynomials called **blended** functions. They influence several end and control points. With different influences from the end and intermediate points, many composite paths may be determined.

Generally the curve-fitting process is guided by several conditions:

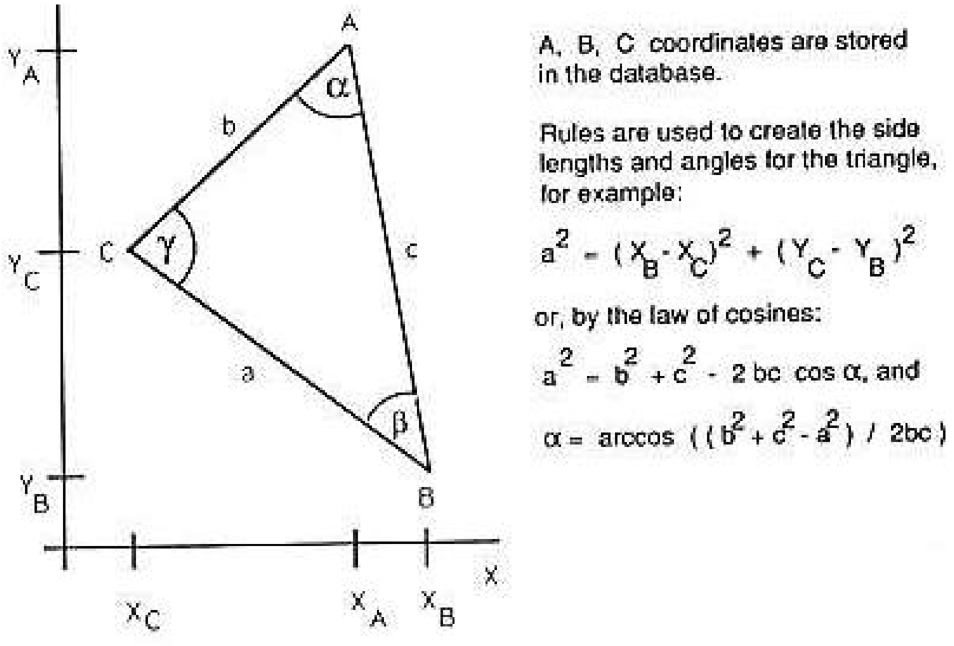
- the number of points of influence on the path of the curve;
- the number of points the curve passes through;
- tangents at the origin and at the extremity;
- if continuities are desired at the location where the different pieces of the curve come together.

Intensional and extensional representation of objects

In general, the individual entities may be represented in practical terms in two contrasting ways. For example, a straight line can be constructed from an ordered series of points or **may** be created via t mathematical formula. The arc or spline type of entity implies the latter case. So we recognize these two situations:

1. The extensional character.
2. The intensional character.

For the **intensional** state, we deal with a small amount of data explicitly, and store the rules, procedure, or method for obtaining other information. That is, then, for the river the coefficients of a mathematical equation are stored along with the end points of the line feature. A more simple **example** is **the representation** of a triangle by three vertices and rules to compute segment lengths and/or angles. This can be not restricted to geometric objects, but can also deal with chunks of space in this way.



A, B, C coordinates are stored in the database.

Rules are used to create the side lengths and angles for the triangle, for example:

$$a^2 = (X_B - X_C)^2 + (Y_C - Y_B)^2$$

or, by the law of cosines:

$$a^2 = b^2 + c^2 - 2bc \cos\alpha, \text{ and}$$

$$\alpha = \arccos\left((b^2 + c^2 - a^2) / 2bc\right)$$

For the moment, we summarize that intensional data require:

1. The storage of some privileged elements of data for the object.
2. The specification of a rule for generating all possible elements, the generative rule.

3. The definition of a rule for testing if an element, for example, a line segment, is a **member** of the object, perhaps a polygon.

The third aspect is required for, among other things, ascertaining if a particular coordinate position falls inside an object (say a polygon) or outside, as well as on the boundary. Thus, there is a way to circumvent the practical difficulty in storing the infinite number of points that lie inside a polygon.