[FC-05-042] Distance Operations

Abstract

Distance is a central concept in geography, and consequently, there are various types of operations that leverage the concept of distance. This short article introduces common distance measures, the purpose of distance operations, different types of operations and considerations, as well as sample applications in the physical and social domains. Distance operations can be performed on both vector or raster data, but the operations and results may differ. While performing distance operations, it is important to remember how distance is conceptualized while performing the operation.

Keywords: cost distance, distance decay, distance matrix, Euclidean distance, network distance

Author & citation

Sarkar, D. (2019). Distance Operations. The Geographic Information Science & Technology Body of Knowledge (3rd Quarter 2019 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2019.3.3

This Topic is also available in the following editions:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Distances and Lengths. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Explanation

1. Definitions
2. Common Distance Measures in GIS
3. Distance Operations

1. Definitions

Space: Space in this context is broadly defined as boundless three-dimensional extent in which objects and events have relative position and direction**.**

Euclidean space: Euclidean space is a metric space, that is, a space in which the notion of distance between two points and its properties are axiomatically defined and quantifiable; in Euclidean space, it is possible to know the (x, y) location of an object and it is also possible to quantify the distance and orientation between them.

Point: A point in space denotes a location using (x, y) coordinates.

Feature: A feature is an object or event located in space; the object may be a single location (like a well), which can be denoted using a point, may represent a more expansive entity like a mountain range, or may represent more complex phenomenon, such as the

location of diseased trees in a pine forest.

2. Common Distance Measures in GIS

The First Law of Geography states that “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). Distance is the quantified embodiment of the concept of near and far. Consequently, ‘distance’ is a central concept in geography on which many further concepts are built. Simply put, a distance is a metric that allows you to measure the traversable space between two features. Figure 1 shows four different ways of traversing between two points. From the figure, it is clear that given two points in space, there are several ways one might reach from one to the other. The different ways in which the space is traversed produce different methods of calculating distance. In this topic, we first introduce the common methods of distance calculations in GISc before introducing the common distance operations. Note that, to perform most of the distance operations, several methods of measuring distance may be used. The choice of the distance measure used may significantly alter the results of the operation.

2.1 Euclidean Distance

Euclidean Distance is the most commonly used distance measure. Given two points, $P_1(x_1, y_1), P_2(x_2, y_2)$, in two-dimensional Euclidean space, the Euclidean distance is the shortest, ‘as the crow flies’ distance between them. Mathematically, Euclidean distance (E) is defined as:

$$E(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

In addition to the two-dimensional location information, if elevation is also present, then the two points are defined by 3 coordinates (x, y, z), and the formula for Euclidean distance between them is as follows:

$$E(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

2.2 Manhattan Distance

The space between the two points P1 and P2 is rarely empty. Given the placement of features between the two points, it may not be possible to take a direct ‘as the crow flies’ route. In the case of planned cities, the roads often meet each other at right angles, dividing the city into a collection of blocks, forcing a traveler to travel only in east-west, or north-south directions (Figure 1). In such a case, the shortest distance between the two points is shown using red in Figure 1. This distance is called the Manhattan distance (M), or the taxi cab distance and is measured as follows:

$$M(P_1, P_2) = |x_1 - x_2| + |y_1 - y_2|$$

Interestingly, unlike Euclidean distance which has only one shortest path between two points P1 and P2, there can be multiple shortest paths between the two points when using Manhattan Distance. In Figure 1, the lines the red, yellow, and blue paths all have the same shortest path length of 12, while the Euclidean shortest path distance shown in green has a

length of 8.5.

Strictly speaking, Manhattan distance is a two-dimensional metric defined in a different geometry to Euclidean space, in which movement is restricted to north-south, east-west directions only.

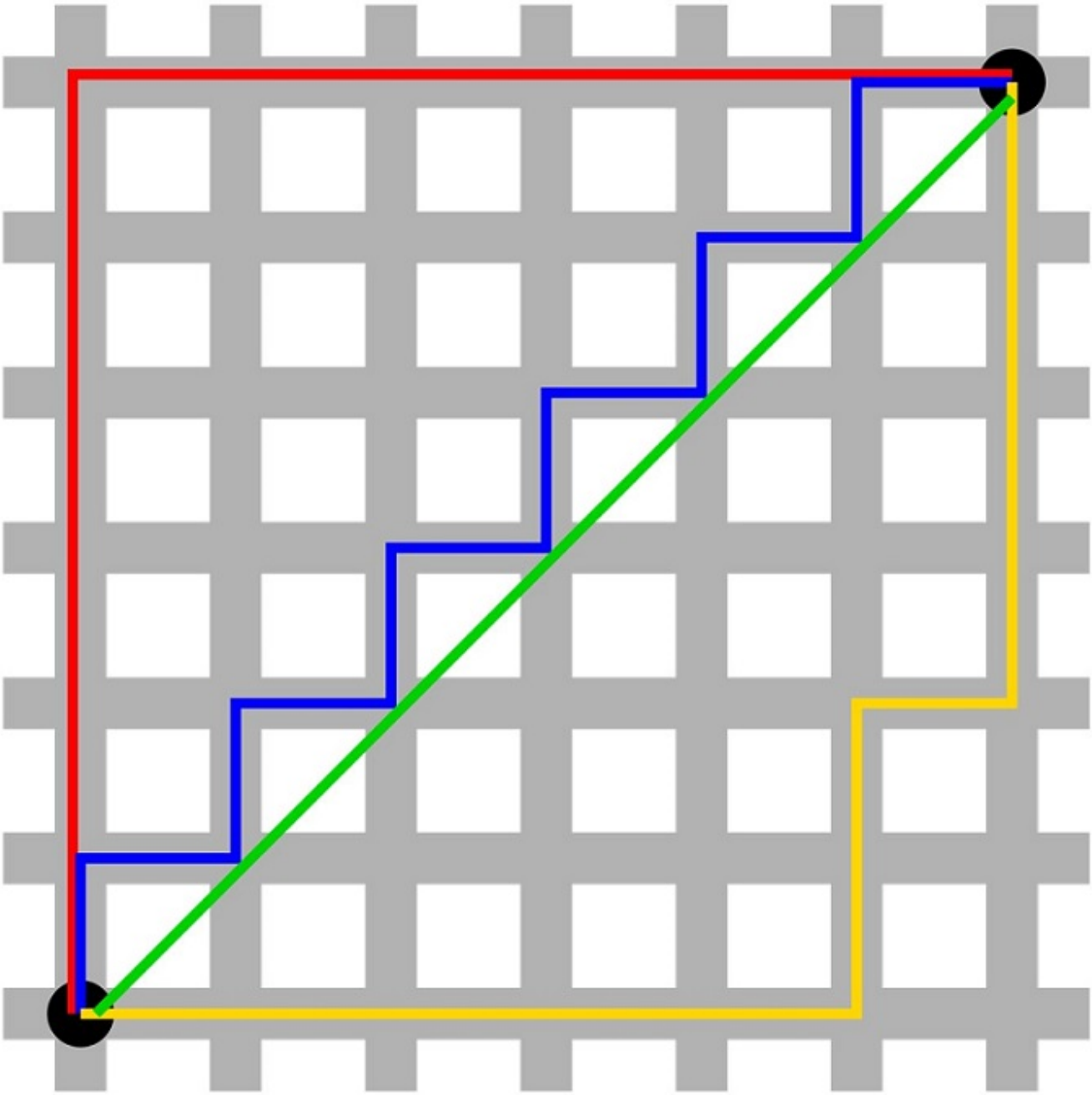


Figure 1. Euclidean Distance and Manhattan Distance. The green line shows the Euclidean path between the two points while the red, blue, and yellow lines show the Manhattan Paths between the two points. All the three Manhattan paths have the same Manhattan Distance. (Image from Wikimedia Commons)

2.3 Cost Distance

In a landscape, the ease of traversing every part of the landscape may not be uniform. For example, the topography of the landscape may make traversing certain parts harder than others (Li, Larson, & Rex, 2005). Furthermore, there may be barriers (e.g. mountains, fences), which may render certain parts of the landscape not reachable from others. Thus, for such cases, the distance between features can be measured in 'cost' of traversal (Douglas, 1994; Tomlin, 2010). Cost distance is often available in the form of a raster referred to as a cost surface, where the value of each pixel represents the cost of traversing through that pixel. An example of such a raster is a slope raster calculated from a Digital Elevation Model (DEM). The value of this cost surface represents the slope, which provides a measure of how hard it is to travel along particular routes. In order to find the cost of traversal between two points on this surface, one needs to add all the values of the pixels through which the path traverses. The final cost value should be multiplied with the actual distance traversed between the two points to get the Cost Distance.

2.4 Geodesic Distance

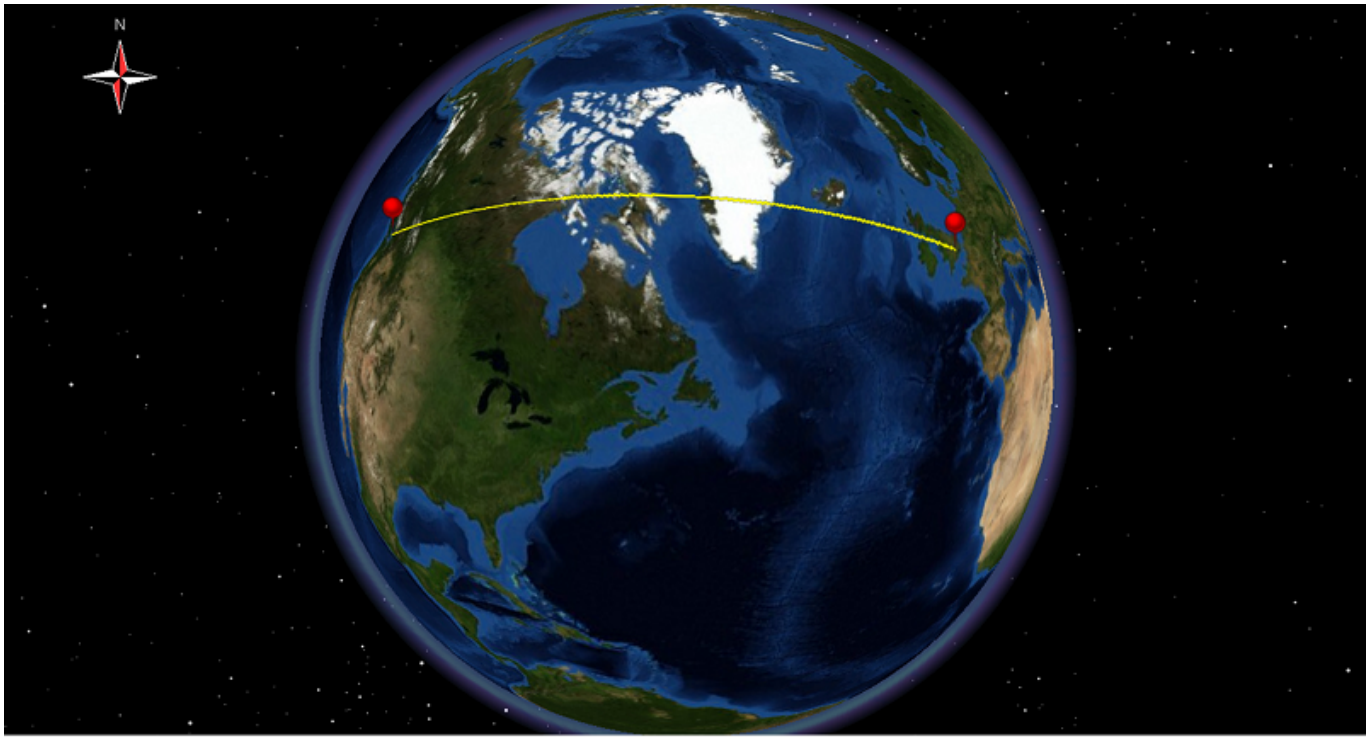
While the Euclidean and Manhattan distances are suitable for measurement of distance on projected data, they are not suitable for non-Euclidean surfaces like a sphere, which is often used to represent the earth. The shortest distance between two points P1 and P2 on the surface of a sphere, with latitude and longitude of ($^{\Phi_1, \lambda_1}$) and ($^{\Phi_2, \lambda_2}$) measured in radians, that is the Geodesic Distance (G), is given by the spherical law of cosines as:

$$G = r\Delta\sigma$$

$$where, \Delta\sigma = \{sin(sin(\varphi_1)) \cdot sin(sin(\varphi_2)) + cos(cos(\varphi_1)) \cdot cos(cos(\varphi_2)) \cdot cos(cos(\Delta\lambda))\}$$

$$and, \Delta\lambda = |\lambda_1 - \lambda_2|$$

Let $^{x_2, y_2}$ with coordinates ($^{x_2, y_2}$) and ($^{x_2, y_2}$) be the projection of the points P1 and P2 on a flat surface. Let e be the distance between the points p1 and p2 on the plane. There is difference between the distance measure G and e. This deviation becomes more significant when the two places are further apart on the surface of the sphere (Figure 2). Thus, when using distance measures, it is important to consider the location of the points on the surface of the earth and the projection used on the data. if the places are located at higher latitudes, inappropriate projections may cause analysis errors when using distance operations. The term Great Circle Distance is sometimes used to refer to Geodesic Distance when the earth is modeled as a sphere. The term Geodetic Distance is more general as it represents the shortest path on a curved surface (e.g. sphere or ellipsoid) where as Great Circle Distance is the shortest path on a sphere, but not on all ellipsoids. Since Great Circle Distance measures the distance between two points on the surface of the sphere, it represents the shortest route between two places on earth (Karney, 2013). Thus, when looking at a flat map, an airplane may seem to be flying a longer route, but it is in fact following the geodesic path which is getting distorted when projected on a map.


N

A


GREENLAND
N
7591.41 Kms
13660.47 Kms
CANADA
UNITED KINGDOM
London
Paris
FRANCE
Toronto
Chicago
New York
UNITED STATES
Madrid
San Francisco
Los Angeles
ALGERIA
Atlantic Ocean
Legend
Airports
Planar Distance
Geodesic Distance
Esri, HERE, Garmin, © OpenStreetMap contributors, and the GIS user community

B

Figure 2. The top figure (A) shows the Geodesic Distance between Heathrow Airport, London and Vancouver International Airport, Canada. The Image below (B) shows the projection of the Geodesic Distance on a map along with the Planar Distance between the two places.

2.5 Network Distance

Not all spaces can be traversed in a continuous manner between two points. For example, on a road network, the movement is from point to point along the roads. Such a structure is represented using a network consisting of nodes (intersections), connected together by edges (roads). Social networks also use the node and edge structure to depict relationships (as edges) between entities (denoted by nodes). Distance is measured by counting the number of intermediate nodes traversed as result of hoping from node to node along the edges. However, to calculate the shortest distance between two nodes in a network, a simple formula does not suffice. Algorithms, such as Dijkstra's algorithm are required to calculate the shortest path on network. This complexity arises because of the constraints of the network data model in how nodes and edges can be traversed (Sarkar, Sieber, & Sengupta, 2016).

Note that, in the case of road networks, the movement is not necessarily from node to node (denoting intersections) but may be continuous along the edges representing the roads. In such a case, the distance between two points on the road network can be calculated by adding up the distance, segment by segment. This measure is often called Road Network Distance, or Network Distance in short.

3. Distance Operations

This section describes the various types of Distance Operation commonly used in GIS. Many of the topics are covered in detail in other chapters listed in the ‘Related Topics’ section. Readers are encouraged to look at these chapters for a more comprehensive treatment.

3.1 Buffers

Buffers delineate proximal regions by creating a boundary of a given distance around a feature. One of the most common tasks in GIS is to find out what is proximal to or within the neighborhood of a location. For example, a conservation scientist wants to create riparian buffer zones to protect streams from the impact of adjacent land uses (Xiang, 1993). This question can be handled by creating buffers of the required distance around the streams to delineate protected areas. The buffer of a geographic feature is an area where the shortest distance from any location within the area to the feature is less than or equal to a threshold distance.

3.2 Isochrones

Distance can also be defined in terms of time. Given a fixed time interval, it is possible to draw buffer-like rings around a feature that denote the maximum distance an object may have covered in that interval if the direction of travel is not known or does not matter

(Figure 3). These time-based distance rings are known as isochrones (Miller & Bridwell, 2009; O'Sullivan, Morrison, & Shearer, 2000). The isochrones shown in Figure 3 account for the underlying road network. Thus, the distance is calculated as network distance and each polygon represents the potential distance that could have been traveled from the center at every 5-minute interval, if the speed of the vehicle was 100 Km/h at most. Isochrones are useful in scenarios such as catching a burglar. After a crime has been reported, the police may have a general idea of how fast a car can travel in the area. With the knowledge of the speed and the location of the crime, the police may calculate isochrones to help determine where to set up barriers. Note that, the isochrones may not be perfect circles. As the movement of the car is restricted to the road network, the distance may be calculated along the roads (road network distance) and not 'as the crow flies' as measured by Euclidean distance (Figure 3).

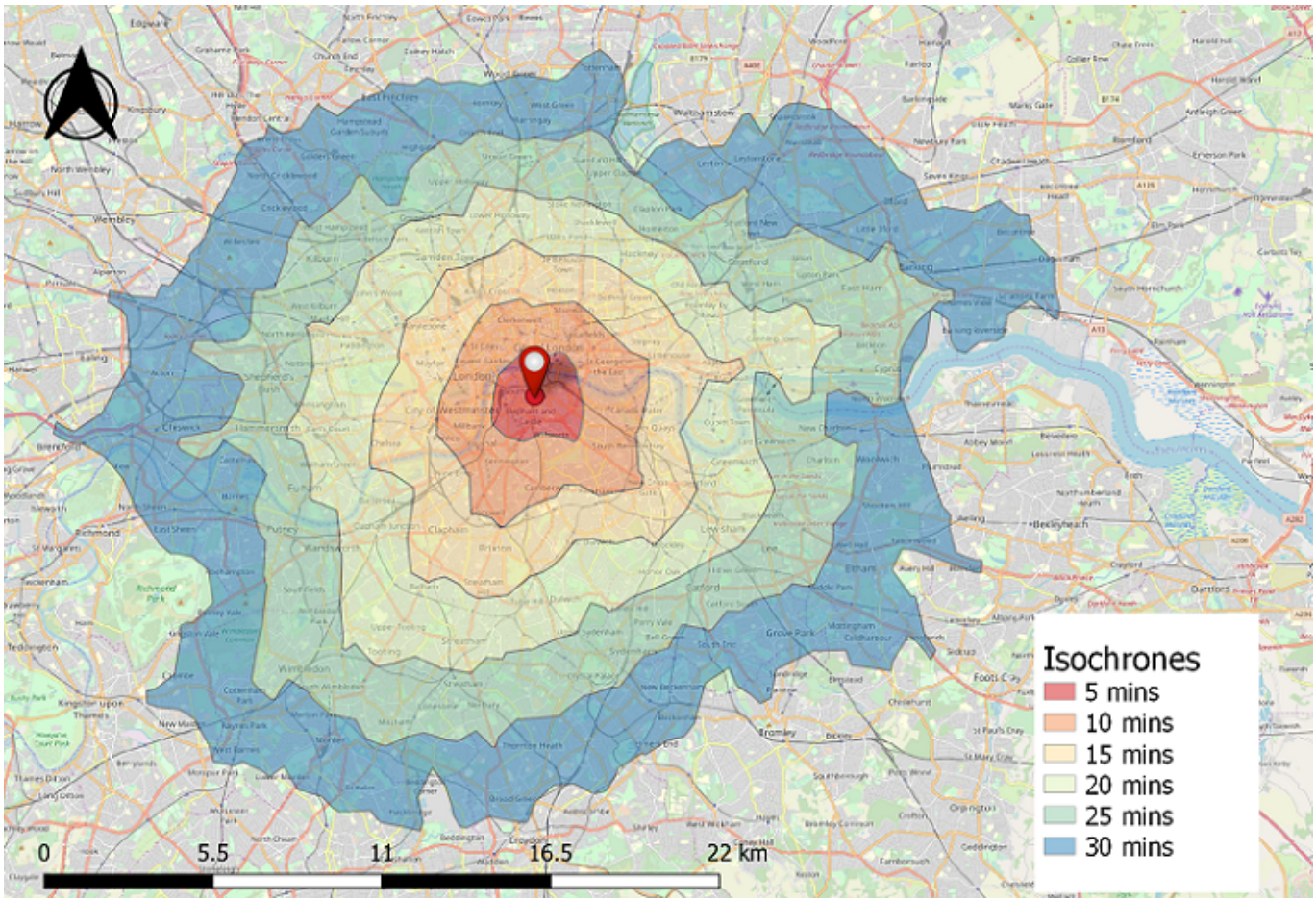


Figure 3. Isochrones showing maximum travel distance possible from the central red point at every 5-minutes interval with the maximum travel speed limited to 100 km/hr.

3.3 Distance Matrix

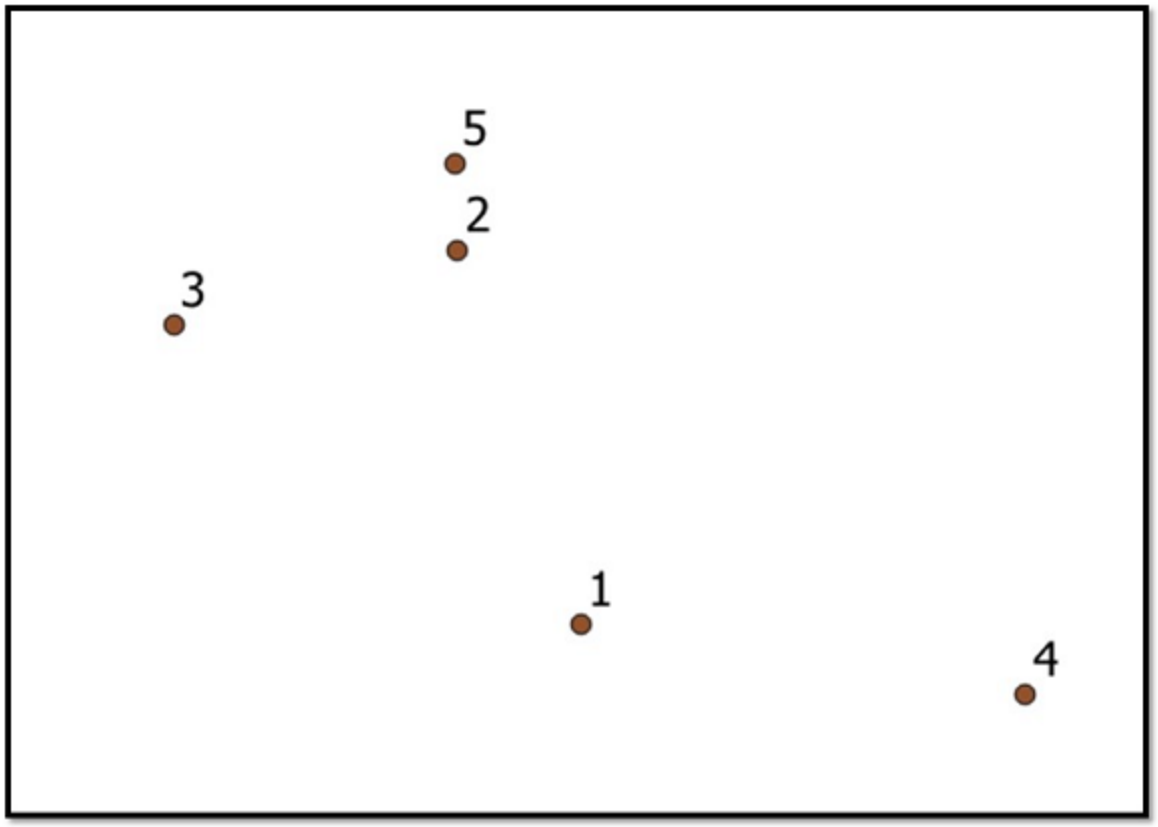
Given a set of features, the distance between each pair of features can be stored in a matrix called distance matrix (Figure 4). The distance matrix may be calculated once and used as a look-up table any time the distance between features is required. The inverse distance matrix is a modification of the distance matrix and stores the inverse of the distance between the features instead of the actual distances (Anselin, 1992). This is to

save computational time in later steps of spatial analysis which may need to account for autocorrelation or clustering among observations. Thus, when distance is used as a surrogate for similarity measure, closer features should have higher weights. Several spatial statistics, such as Moran's I and Getis Ord G*, can require the distance matrix as an input. The entries in the distance matrix can embody different conceptualizations of spatial relationship between the features (Getis, 2009).

Topographic features on the landscape make a landscape harder to traverse. According to the distribution of features, the ease of traveling between points P1 and P2 may be different from P2 to P3 (even P2 and P1, when there is a difference in elevation) even though they are the same Euclidean distance away. A cost matrix looks like a distance matrix but stores a measure of how hard it is to get from one place to the next. Thus, when using statistics that use distance matrix, depending on the application scenario, it may be important to multiply the cost matrix with the distance matrix and produce the cost-distance matrix to provide as input. Depending on the application scenario, this ensures that the on-ground realities of traversing the landscape are factored in. Also, depending on the application scenario and data availability, there will be different ways in which the cost-distance matrix or surface is created. As mentioned earlier, the cost may also be provided as a raster image.

In case of network datasets like roads, a distance matrix called Origin-Destination (O-D) Matrix is often created which stores the shortest distance between a set of locations.

In case of all the types of matrices discussed above, the matrix may be symmetric or asymmetric depending on whether distance/cost of getting from point P1 to point P2 is the same as going from P2 to P1. An example scenario when asymmetric matrices might be required is when restrictions of road networks prevent taking the same route as the forward journey on the return journey. Thus, distance matrices allow encoding anisometric distances between points.



A

	1	2	3	4	5
1	0	80.95	97.49	72.29	97.71
2	80.95	0	39.95	140.25	17.01
3	97.49	39.95	0	166.16	43.03
4	72.29	140.25	166.16	0	154.79
5	97.71	17.01	43.03	154.79	0

B

Figure 4. Random points in Euclidean Space (A) and its corresponding distance matrix (B). The first row and first column of the matrix denote the IDs of the point, each entry in the matrix denotes the distance in meters between the two points.

3.4 K-Nearest Neighbors

Adjacency, as expressed by nearness of features, is also an important consideration in

many applications. The closest neighbors of a feature sometimes have a significant effect on the feature of interest. For example, when setting up a new retail location, an analyst might be interested in finding out where their closest competitors are located. Such operations are known as k-Nearest Neighbor (k-NN for short) where k represents the number of neighbors of interest. The k-NNs can be found out from the distance matrix if it is pre-calculated, or by gradually increasing the search radius if the distance matrix is not present. Note that, while both buffering and k-NNs are used to find proximal features, unlike buffers, the k-NN algorithm does not require prior knowledge of a search radius. Thus, when the features are dispersed in space, the k-NN method is more efficient and can be more meaningful depending on the application scenario.

3.5 Distance Decay

Nearby features tend to be similar, but, as the distance increases, the similarity between features reduce. The underlying geographic processes that produce this spatial pattern tend to lose potency as distance from the source increases. Thus, distance decay is a method of modeling the influence of geographic processes as a function of distance to determine how strongly different features will interact with each other. Spatial interaction between features is generally assumed to have a smooth decay function with respect to distance (Anselin, 1998). Depending on the application scenario, the rate of decay is modeled as a linear, quadratic, or exponential function (Krivoruchko, 2011, p. 302). Modeling the distance decay is an important step for geostatistical techniques like kriging. A simple example of distance decay is, when trying to model vegetation abundance that is dependent on distance to water sources. It is important to model how the abundance changes with distance to a water source. Using a quadratic model of distance decay in this case will result in a much sharper decline of abundance with distance than using a linear model.

Distance decay can also be used to find the area of influence and determine interacting features. For example, gravity models are commonly used to model migration of people between cities (Foot & Milne, 1984; Karemera, Oguledo, & Davis, 2000; Plane, 1984). In such a model, the flow of people is dependent on the mass (e.g. population, economic opportunity) of the city and the distance between the cities. The distance decay may be modeled as the inverse of the square of distance to account for the fact that, the further away the cities are, the less likely people are to move. Thus, the probability of two cities interacting decays with increasing distance.

3.6 Autocorrelation

Features may not be uniformly distributed in space. In fact, given the first law of geography, similar features tend to be located close to each other. Autocorrelation provides a measure of how similar nearby features are. In order to measure the autocorrelation using statistics such as Global Moran’s I, one must compare the average distance between the features in the observed process to the corresponding distance if the features were produced by a random process (Li, Calder, & Cressie, 2007). Thus, knowing distance between features is a key indicator of the whether the observation produced by the underlying process are clustered in space.

3.7 Clustering

Clusters are defined by a group of features that are located closer to each other than a

random process would produce. While global autocorrelation measures quantify the tendency of features in a dataset to cluster, it does not give the location of the clusters. To find the location of individual clusters, spatially contiguous subsets of features needs to be evaluated rather than looking at the entire dataset at once. The simplest way of finding out clusters is by looking at a group of features to determine whether the average distance between the features (density) is greater than that of its neighborhood. This idea is captured by the metric Local Moran's I.

3.8 Traveling Salesperson

The Traveling Salesperson problem builds on the shortest path analysis on network datasets to find the most efficient path to visit a series of locations and return to the origin. Efficiency here can be measured in terms of distance traveled, time, or any other suitable method. An exact solution to this problem is difficult to obtain computationally, especially when the number of cities is large. But, heuristic algorithms can be used to find reasonably accurate solutions to the traveling salesman problem (Orponen, 1990; Sharma, Mioc, & Dharmaraj, 2005). A common variation of the traveling salesperson problem in GIS is termed as a vehicle routing problem. An example application of the traveling salesperson problem is a salesperson trying to most efficiently travel across a planned set of locations to sell their products. Thus, provided the list of places the salesperson is supposed to visit, the solution involves listing the order in which the locations are to be visited so as to optimize the overall journey.

3.9 Closest Facility

Given a road network and information about facilities located along the network, closest facility analysis determines the closest facility to any location of interest. This analysis uses shortest path algorithms and is useful for emergency situations. For example, in case of a medical emergency, a user might need to know where the nearest hospital facility is located.

3.10 Service Area

Service Area is another network-based operation and is similar to isochrones. In this case, given a point location (usually referred to as a facility, such as, fire station, police station) on a road network, the service area finds the geographic region which can be reached from the facility in a certain period of time and vice versa. Thus, in this operation, time of travel determines distance. Service Area analysis can help ensure that different neighborhoods are sufficiently served by civic facilities.

References

Anselin, L. (1992). Spatial Data Analysis with GIS: an Introduction to Application in the Social Sciences. Systems Research, 1–54.

Anselin, L. (1998). GIS Research Infrastructure for Spatial Analysis of Real Estate Markets. Journal of Housing Research, 9(1), 113–133.

Douglas, D. H. (1994). Least-cost Path in GIS Using an Accumulated Cost Surface and Slopelines. Cartographica, 31, 37-51.

Foot, D. K., & Milne, W. J. (1984). Net migration estimation in an extended, multiregional gravity model. Journal of Regional Science, 24(1), 119–133.

Getis, A. (2009). Spatial Weights Matrices. Geographical Analysis, 41(4), 404–410.

Karemera, D., Oguledo, V. I., & Davis, B. (2000). A gravity model analysis of international migration to North America. Applied Economics, 32(13), 1745–1755.

Karney, C. F. F. (2013). Algorithms for geodesics. Journal of Geodesy, 87(1), 43–55.

Krivoruchko, K. (2011). Spatial Statistical Data Analysis for GIS Users. Redlands, CA: Esri Press.

Li, H., Calder, C. A., & Cressie, N. (2007). Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. Geographical Analysis, 39(4), 357–375.

Li, X., Larson, C. M., & Rex, A. B. (2005). Creating Buffers on Surfaces. Cartography and Geographic Information Science, 32(3), 195–212.

Miller, H. J. & Bridwell, S. A. (2009). A Field-Based Theory for Time Geography. Annals of the Association of American Geographers, 99(1), 49-75.

O'Sullivan, D., Morrison, A., & Shearer, J. (2000). Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach. International Journal of Geographical Information Science, 14(1), 85–104.

Orponen, P. and Mannila, H. (1990). On Approximation Preserving Reductions: Complete Problems and Robust Measures (Revised Version).

Plane, D. A. (1984). Migration Space: Doubly Constrained Gravity Model Mapping of Relative Interstate Separation. Annals of the Association of American Geographers, 74(2), 244–256.

Sarkar, D., Sieber, R., & Sengupta, R. (2016). GIScience Considerations in Spatial Social Networks. In J. A. Miller, D. O. Sullivan, & N. Wiegand (Eds.), Lecture Notes in Computer Science (Vol. 1, pp. 85–98). Springer International Publishing.

Sharma, O., Mioc, D., Anton, F., & Dharmaraj, G. (2006). Traveling salesperson approximation algorithm for real road networks. In ISPRS WG II/1,2,7, VII/6 International Symposium on Spatio-temporal Modeling, Spatial Reasoning, Spatial Analysis, Data Mining and Data Fusion (STM 2005), Beijing, China. Taylor & Francis.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46: 234–240.

Tomlin, C. D. (2010). Propagating radial waves of travel cost in a grid. International Journal of Geographical Information Science, 24(9), 1391–1413.