CARTOGRAPHIC GENERALIZATION IN A DIGITAL ENVIRONMENT: WHEN AND HOW TO GENERALIZE

K. Stuart Shea
The Analytic Sciences Corporation (TASC)
12100 Sunset Hills Road
Reston, Virginia 22090

Robert B. McMaster
Department of Geography
Syracuse University
Syracuse, New York 13244-1160

ABSTRACT

A key aspect of the mapping process—cartographic generalization—plays a vital role in assessing the overall utility of both computer-assisted map production systems and geographic information systems. Within the digital environment, a significant, if not the dominant, control on the graphic output is the role and effect of cartographic generalization. Unfortunately, there exists a paucity of research that addresses digital generalization in a holistic manner, looking at the interrelationships between the conditions that indicate a need for its application, the objectives or goals of the process, as well as the specific spatial and attribute transformations required to effect the changes. Given the necessary conditions for generalization in the digital domain, the display of both vector and raster data is, in part, a direct result of the application of such transformations, of their interactions between one another, and of the specific tolerances required.

How then should cartographic generalization be embodied in a digital environment? This paper will address that question by presenting a logical framework of the digital generalization process which includes: a consideration of the intrinsic objectives of why we generalize; an assessment of the situations which indicate when to generalize; and an understanding of how to generalize using spatial and attribute transformations. In a recent publication, the authors examined the first of these three components. This paper focuses on the latter two areas: to examine the underlying conditions or situations when we need to generalize, and the spatial and attribute transformations that are employed to effect the changes.

INTRODUCTION

To fully understand the role that cartographic generalization plays in the digital environment, a comprehensive understanding of the generalization process first becomes necessary. As illustrated in Figure 1, this process includes a consideration of the intrinsic objectives of why we generalize, an assessment of the situations which indicate when to generalize, and an understanding of how to generalize using spatial and attribute transformations. In a recent publication, the authors presented the why component of generalization by formulating objectives of the digital generalization process (McMaster and Shea, 1988). The discussion that
follows will focus exclusively on the latter two considerations—an assessment of the degree and type of generalization and an understanding of the primary types of spatial and attribute operations.

**Figure 1.** Decomposition of the digital generalization process into three components: why, when, and how we generalize. The why component was discussed in a previous paper and will not be covered here.

**SITUATION ASSESSMENT IN GENERALIZATION:**

**WHEN TO GENERALIZE**

The situations in which generalization would be required ideally arise due to the success or failure of the map product to meet its stated goals; that is, during the cartographic abstraction process, the map fails "...to maintain clarity, with appropriate content, at a given scale, for a chosen map purpose and intended audience" (McMaster and Shea, 1988, p.242). As indicated in Figure 2, the when of generalization can be viewed from three vantage points: (1) **conditions** under which generalization procedures would be invoked; (2) **measures** by which that determination was made; and (3) **controls** of the generalization techniques employed to accomplish the change.

**Conditions for Generalization**

Six conditions that will occur under scale reduction may be used to determine a need for generalization.

*Congestion:* refers to the problem where too many features have been positioned in a limited geographical space; that is, feature density is too high.

*Coalescence:* a condition where features will touch as a result of either of two factors: (1) the separating distance is smaller than the resolution of the output device (e.g. pen
width, CRT resolution); or (2) the features will touch as a result of the symbolization process.

**Conflict:** a situation in which the spatial representation of a feature is in conflict with its background. An example here could be illustrated when a road bisects two portions of an urban park. A conflict could arise during the generalization process if it is necessary to combine the two park segments across the existing road. A situation exists that must be resolved either through symbol alteration, displacement, or deletion.

**Complication:** relates to an ambiguity in performance of generalization techniques; that is, the results of the generalization are dependent on many factors, for example: complexity of spatial data, selection of iteration technique, and selection of tolerance levels.

**Inconsistency:** refers to a set of generalization decisions applied non-uniformly across a given map. Here, there would be a bias in the generalization between the mapped elements. Inconsistency is not always an undesirable condition.

**Imperceptibility:** a situation results when a feature falls below a minimal portrayal size for the map. At this point, the feature must either be deleted, enlarged or exaggerated, or converted in appearance from its present state to that of another—for example, the combination of a set of many point features into a single area feature (Leberl, 1986).

It is the presence of the above stated conditions which requires that some type of generalization process occur to counteract, or eliminate, the undesirable consequences of scale change. The conditions noted, however, are highly subjective in nature and, at best, difficult to quantify. Consider, for example, the problem of congestion. Simply stated, this refers to a condition where the density of features is greater than the available space on the graphic. One might question how this determination is made. Is it something that is computed by an algorithm, or must the we rely upon operator intervention? Is it made in the absence or presence of the symbology? Is symbology's influence on perceived density—that is, the percent blackness covered by the symbology—the real factor that requires evaluation? What is the unit area that is used in the density calculation? Is this unit area dynamic or fixed? As one can see, even a relatively straightforward term such as density is an enigma. Assessment of the other remaining conditions—coalescence, conflict, complication, inconsistency, and imperceptibility—can also be highly subjective.

How, then, can we begin to assess the state of the condition if the quantification of those conditions is ill-defined? It appears as though such conditions, as expressed above, may be detected by extracting a series of measurements from the original and/or generalized data to determine the presence or absence of a conditional state. These measurements may indeed be quite complicated and inconsistent between various maps or even across scales within a single map type. To eliminate these differences, the assessment of conditions must be based entirely from outside a map product viewpoint. That is, to view the map as a graphic entity in its most elemental form—points, lines, and areas—and to judge the conditions based upon an analysis of those entities. This is accomplished through the evaluation of **measures** which act as indicators into the geometry of individual features, and assess the spatial relationships between combined features. Significant examples of these measures can be found in the cartographic literature (Catlow and Du, 1984; Christ, 1976; Dutton, 1981; McMaster, 1986; Robinson, et al., 1978).
Measures Which Indicate a Need for Generalization

Conditional measures can be assessed by examining some very basic geometric properties of the inter- and intra-feature relationships. Some of these assessments are evaluated in a singular feature sense, others between two independent features, while still others are computed by viewing the interactions of multiple features. Many of these measures are summarized below. Although this list is by no means complete, it does provide a beginning from which to evaluate conditions within the map which do require, or might require, generalization.

Density Measures. These measures are evaluated by using multi-features and can include such benchmarks as the number of point, line, or area features per unit area; average density of point, line, or area features; or the number and location of cluster nuclei of point, line, or area features.

Distribution Measures. These measures assess the overall distribution of the map features. For example, point features may be examined to measure the dispersion, randomness, and clustering (Davis, 1973). Linear features may be assessed by their complexity. An example here could be the calculation of the overall complexity of a stream network (based on say average angular change per inch) to aid in selecting a representative depiction of the network at a reduced scale. Areal features can be compared in terms of their association with a common, but dissimilar area feature.

Length and Sinuosity Measures. These operate on singular linear or areal boundary features. An example here could be the calculation of stream network lengths. Some sample length measures include: total number of coordinates; total length; and the average number of coordinates or standard deviation of coordinates per inch. Sinuosity measures can include: total angular change; average angular change per inch; average angular change per angle; sum of positive or negative angles; total number of positive or negative angles; total number of positive or negative runs; total number of runs; and mean length of runs (McMaster, 1986).

Shape Measures. Shape assessments are useful in the determination of whether an area feature can be represented at its new scale (Christ, 1976). Shape mensuration can be determined against both symbolized and unsymbolized features. Examples include: geometry of point, line, or area features; perimeter of area features; centroid of line or area features; X and Y variances of area features; covariance of X and Y of area features, and the standard deviation of X and Y of area features (Bachi, 1973).

Distance Measures. Between the basic geometric forms—points, lines, and areas—distance calculations can also be evaluated. Distances between each of these forms can be assessed by examining the appropriate shortest perpendicular distance or shortest euclidean distance between each form. In the case of two geometric points, only three different distance calculations exist: (1) point-to-point; (2) point buffer-to-point buffer; and (3) point-to-point buffer. Here, point buffer delineates the region around a point that accounts for the symbology. A similar buffer exists for both line and area features (Dangermond, 1982). These determinations can indicate if any generalization problems exist if, for instance under scale reduction, the features or their respective buffers are in conflict.

Gestalt Measures. The use of Gestalt theory helps to indicate perceptual characteristics of the feature distributions through isomorphism—that is, the structural kinship between the stimulus pattern and the expression it conveys (Arnheim, 1974). Common examples of this includes closure, continuation, proximity, similarity, common fate, and figure ground (Wertheimer, 1958).

Abstract Measures. The more conceptual evaluations of the spatial distributions can be examined with abstract measures. Possible abstract measures include: homogeneity, neighborliness, symmetry, repetition, recurrence, and complexity.
Many of the above classes of measures can be easily developed for examination in a digital domain, however the Gestalt and Abstract Measures aren't as easily computed. Measurement of the spatial and/or attribute conditions that need to exist before a generalization action is taken depends on scale, purpose of the map, and many other factors. In the end, it appears as though many prototype algorithms need first be developed and then tested and fit into the overall framework of a comprehensive generalization processing system. Ultimately, the exact guidelines on how to apply the measures designed above can not be determined without precise knowledge of the algorithms.

**Controls on How to Apply Generalization Functionality.**

In order to obtain unbiased generalizations, three things need to be determined: (1) the order in which to apply the generalization operators; (2) which algorithms are employed by those operators; and (3) the input parameters required to obtain a given result at a given scale.

An important constituent of the decision-making process is the availability and sophistication of the generalization operators, as well as the algorithms employed by those operators. The generalization process is accomplished through a variety of these operators—each attacking specific problems—each of which can employ a variety of algorithms. To illustrate, the linear simplification operator would access algorithms such as those developed by Douglas—as reported by Douglas and Peucker (1973)—and Lang (1969). Concomitantly, there may be permutations, combinations, and iterations of operators, each employing permutations, combinations, and iterations of algorithms. The algorithms may, in turn, be controlled by multiple, maybe even interacting, parameters.

**Generalization Operator Selection.** The control of generalization operators is probably the most difficult process in the entire concept of automating the digital generalization process. These control decisions must be based upon: (1) the importance of the individual features (this is, of course, related to the map purpose and intended audience); (2) the complexity of feature relationships both in an inter- and intra-feature sense; (3) the presence and resulting influence of map clutter on the communicative efficiency of the map; (4) the need to vary generalization amount, type, or order on different features; and (5) the availability and robustness of generalization operators and computer algorithms.

**Algorithm Selection.** The relative obscurity of complex generalization algorithms, coupled with a limited understanding of the digital generalization process, requires that many of the concepts need to be prototyped, tested, and evaluated against actual requirements. The evaluation process is usually the one that gets ignored or, at best, is only given a cursory review.

**Parameter Selection.** The input parameter (tolerance) selection most probably results in more variation in the final results than either the generalization operator or algorithm selection as discussed above. Other than some very basic guidelines on the selection of weights for smoothing routines, practically no empirical work exists for other generalization routines.

Current trends in sequential data processing require the establishment of a logical sequence of the generalization process. This is done in order to avoid repetitions of processes and frequent corrections (Morrison, 1975). This sequence is determined by how the generalization processes affect the location and representation of features at the reduced scale. Algorithms required to accomplish these changes should be selected based upon cognitive studies, mathematical evaluation, and design and
implementation trade-offs. Once candidate algorithms exist, they should be assessed in terms of their applicability to specific generalization requirements. Finally, specific applications may require different algorithms depending on the data types, and/or scale.

**SPATIAL AND ATTRIBUTE TRANSFORMATIONS IN GENERALIZATION: HOW TO GENERALIZE**

The final area of discussion considers the component of the generalization process that actually performs the actions of generalization in support of scale and data reduction. This how of generalization is most commonly thought of as the operators which perform generalization, and results from an application of generalization techniques that have either arisen out of the emulation of the manual cartographer, or based solely on more mathematical efforts. Twelve categories of generalization operators exist to effect the required data changes (Figure 3).

Since a map is a reduced representation of the Earth's surface, and as all other phenomena are shown in relation to this, the scale of the resultant map largely determines the amount of information which can be shown. As a result, the generalization of cartographic features to support scale reduction must obviously change the way features look in order to fit them within the constraints of the graphic. Data sources for map production and GIS applications are typically of variable scales, resolution, accuracy—and each of these factors contribute to the method in which cartographic information is presented at map scale. The information that is contained within the graphic has two components—location and meaning—and generalization affects both (Keates, 1973). As the amount of space available for portraying the cartographic information decreases with decreasing scale, less locational information can be given about features, both individually and collectively. As a result, the graphic depiction of the features changes to suit the scale-specific needs. Below, each of these
transformation processes or generalization operators are reviewed. Figure 4 provides a concise graphic depicting examples of each in a format employed by Lichtner (1979).

**Simplification.** A digitized representation of a map feature should be accurate in its representation of the feature (shape, location, and character), yet also efficient in terms of retaining the least number of data points necessary to represent the character. A profligate density of coordinates captured in the digitization stage should be reduced by selecting a subset of the original coordinate pairs, while retaining those points considered to be most representative of the line (Jenks, 1981). Glitches should also be removed. Simplification operators will select the characteristic, or shape-describing, points to retain, or will reject the redundant point considered to be unnecessary to display the line's character. Simplification operators produce a reduction in the number of derived data points which are unchanged in their \(x,y\) coordinate positions. Some practical considerations of simplification includes reduced plotting time, increased line crispness due to higher plotting speeds, reduced storage, less problems in attaining plotter resolution due to scale change, and quicker vector to raster conversion (McMaster, 1987).

**Smoothing.** These operators act on a line by relocating or shifting coordinate pairs in an attempt to plane away small perturbations and capture only the most significant trends of the line. A result of the application of this process is to reduce the sharp angularity imposed by digitizers (Töpfer and Pillewizer, 1966). Essentially, these operators produce a derived data set which has had a cosmetic modification in order to produce a line with a more aesthetically pleasing caricature. Here, coordinates are shifted from their digitized locations and the digitized line is moved towards the center of the intended line (Brophy, 1972; Gottschalk, 1973; Rhind, 1973).

**Aggregation.** There are many instances when the number or density of like point features within a region prohibits each from being portrayed and symbolized individually within the graphic. This notwithstanding, from the perspective of the map's purpose, the importance of those features requires that they still be portrayed. To accomplish that goal, the point features must be aggregated into a higher order class feature areas and symbolized as such. For example, if the intervening spaces between houses are smaller than the physical extent of the buildings themselves, the buildings can be aggregated and resymbolized as built-up areas (Keates, 1973).

**Amalgamation.** Through amalgamation of individual features into a larger element, it is often possible to retain the general characteristics of a region despite the scale reduction (Morrison, 1975). To illustrate, an area containing numerous small lakes—each too small to be depicted separately—could with a judicious combination of the areas, retain the original map characteristic. One of the limiting factors of this process is that there is no fixed rule for the degree of detail to be shown at various scales; the end-user must dictate what is of most value. This process is extremely germane to the needs of most mapping applications. Tomlinson and Boyle (1981) term this process dissolving and merging.

**Merging.** If the scale change is substantial, it may be impossible to preserve the character of individual linear features. As such, these linear features must be merged (Nickerson and Freeman, 1986). To illustrate, divided highways are normally represented by two or more adjacent lines, with a separating distance between them. Upon scale reduction, these lines require that they be merged into one positioned approximately halfway between the original two and representative of both.

**Collapse.** As scale is reduced, many areal features must eventually be symbolized as points or lines. The decomposition of line and area features to point features, or area features to line feature, is a common generalization process. Settlements, airports, rivers, lakes, islands, and buildings, often portrayed as area features on large scale maps, can become point or line features at smaller scales and areal tolerances often guide this transformation (Nickerson and Freeman, 1986).
Refinement. In many cases, where like features are either too numerous or too small to show to scale, no attempt should be made to show all the features. Instead, a selective number and pattern of the symbols are depicted. Generally, this is accomplished by leaving out the smallest features, or those which add little to the general impression of the distribution. Though the overall initial features are thinned out, the general pattern of the features is maintained with those features that are chosen by showing them in their correct locations. Excellent examples of this can be found in the Swiss Society of Cartography (1977). This refinement process retains the general characteristics of the features at a greatly reduced complexity.

Typification. In a similar respect to the refinement process when similar features are either too numerous or too small to show to scale, the typification process uses a representative pattern of the symbols, augmented by an appropriate explanatory note (Lichtner, 1979). Here again the features are thinned out, however in this instance, the general pattern of the features is maintained with the features shown in approximate locations.

Exaggeration. The shapes and sizes of features may need to be exaggerated to meet the specific requirements of a map. For example, inlets need to be opened and streams need to be widened if the map must depict important navigational information for shipping. The amplification of environmental features on the map is an important part of the cartographic abstraction process (Muehrcke, 1986). The exaggeration process does tend to lead to features which are in conflict and thereby require displacement (Caldwell, 1984).

Enhancement. The shapes and size of features may need to be exaggerated or emphasized to meet the specific requirements of a map (Leberl, 1986). As compared to the exaggeration operator, enhancement deals primarily with the symbolization component and not with the spatial dimensions of the feature although some spatial enhancements do exist (e.g. fractalization). Proportionate symbols would be unidentifiable at map scale so it is common practice to alter the physical size and shape of these symbols. The delineation of a bridge under an existing road is portrayed as a series of cased lines may represent a feature with a ground distance far greater than actual. This enhancement of the symbology applied is not to exaggerate its meaning, but merely to accommodate the associated symbology.

Displacement. Feature displacement techniques are used to counteract the problems that arise when two or more features are in conflict (either by proximity, overlap, or coincidence). More specifically, the interest here lies in the ability to offset feature locations to allow for the application of symbology (Christ, 1978; Schittenhelm, 1976). The graphic limits of a map make it necessary to move features from what would otherwise be their true planimetric locations. If every feature could realistically be represented at its true scale and location, this displacement would not be necessary. Unfortunately, however, feature boundaries are often an infinitesimal width; when that boundary is represented as a cartographic line, it has a finite width and thereby occupies a finite area on the map surface. These conflicts need to be resolved by: (1) shifting the features from their true locations (displacement); (2) modifying the features (by symbol alteration or interruption); or (3) or deleting them entirely from the graphic.

Classification. One of the principle constituents of the generalization process that is often cited is that of data classification (Muller, 1983; Robinson, et al., 1978). Here, we are concerned with the grouping together of objects into categories of features sharing identical or similar attribution. This process is used for a specific purpose and usually involves the agglomeration of data values placed into groups based upon their numerical proximity to other values along a number array (Dent, 1985). The classification process is often necessary because of the impracticability of symbolizing and mapping each individual value.
<table>
<thead>
<tr>
<th>Spatial and Attribute Transformations (Generalization Operators)</th>
<th>Representation in the Original Map</th>
<th>Representation in the Generalized Map</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simplification</strong></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Smoothing</strong></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Aggregation</strong></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Amalgamation</strong></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Merge</strong></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Collapse</strong></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Typification</strong></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Exaggeration</strong></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Enhancement</strong></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Displacement</strong></td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 4. Sample spatial and attribute transformations of cartographic generalization.
SUMMARY

This paper has observed the digital generalization process through a decomposition of its main components. These include a consideration of the intrinsic objectives of why we generalize; an assessment of the situations which indicate when to generalize, and an understanding of how to generalize using spatial and attribute transformations. This paper specifically addressed the latter two components of the generalization process—that is, the when, and how of generalization—by formulation of a set of assessments which could be developed to indicate a need for, and control the application of, specific generalization operations. A systematic organization of these primitive processes—in the form of operators, algorithms, or tolerances—can help to form a complete approach to digital generalization.

The question of when to generalize was considered in an overall framework that focused on three types of drivers (conditions, measures, and controls). Six conditions (including congestion, coalescence, conflict, complication, inconsistency, and imperceptibility), seven types of measures (density, distribution, length and sinuosity, shape, distance, gestalt, and abstract), and three controls (generalization operator selection, algorithm selection, and parameter selection) were outlined. The application of how to generalize was considered in an overall context that focused on twelve types of operators (simplification, smoothing, aggregation, amalgamation, merging, collapse, refinement, typification, exaggeration, enhancement, displacement, and classification). The ideas presented here, combined with those concepts covered in a previous publication relating to the first of the three components—effectively serves to detail a sizable measure of the digital generalization process.

REFERENCES


CONCEPTUAL BASIS FOR GEOGRAPHIC LINE GENERALIZATION

David M. Mark
National Center for Geographic Information and Analysis
Department of Geography, SUNY at Buffalo
Buffalo NY 14260

BIOGRAPHICAL SKETCH

David M. Mark is a Professor in the Department of Geography, SUNY at Buffalo, where he has taught and conducted research since 1981. He holds a Ph.D. in Geography from Simon Fraser University (1977). Mark is immediate past Chair of the GIS Specialty group of the Association of American Geographers, and is on the editorial boards of The American Cartographer and Geographical Analysis. He also is a member of the NCGIA Scientific Policy Committee. Mark's current research interests include geographic information systems, analytical cartography, cognitive science, navigation and way-finding, artificial intelligence, and expert systems.

ABSTRACT

Line generalization is an important part of any automated map-making effort. Generalization is sometimes performed to reduce data volume while preserving positional accuracy. However, geographic generalization aims to preserve the recognizability of geographic features of the real world, and their interrelations. This essay discusses geographic generalization at a conceptual level.

INTRODUCTION

The digital cartographic line-processing techniques which commonly go under the term "line generalization" have developed primarily to achieve two practical and distinct purposes: to reduce data volume by eliminating or reducing data redundancy, and to modify geometry so that lines obtained from maps of one scale can be plotted clearly at smaller scales. Brassel and Weibel (in press) have termed these statistical and cartographic generalization, respectively. Research has been very successful in providing algorithms to achieve the former, and in evaluating them (cf. McMaster, 1986, 1987a, 1987b); however, very little has been achieved in the latter area.

In this essay, it is claimed that Brassel and Weibel's cartographic generalization should be renamed graphical generalization, and should further be subdivided: visual generalization would refer to generalization procedures based on principles of computational
vision, and its principles would apply equally to generalizing a machine part, a cartoon character, a pollen grain outline, or a shoreline. On the other hand, geographical generalization would take into account knowledge of the geometric structure of the geographic feature or feature-class being generalized, and would be the geographical instance of what might be called phenomenon-based generalization. (If visual and geographic generalization do not need to be separated, then a mechanical draftsman, biological illustrator, and cartographer all should be able to produce equally good reduced-scale drawings of a shoreline, a complicated machine part, or a flower, irrespectively; such an experiment should be conducted!)

This essay assumes the following:

geographical generalization must incorporate information about the geometric structure of geographic phenomena.

It attempts to provide hints and directions for beginning to develop methods for automated geographical generalization by presenting an overview of some geographic phenomena which are commonly represented by lines on maps. The essay focuses on similarities and differences among geographic features and their underlying phenomena, and on geometric properties which must be taken into account in geographical generalization.

OBJECTIVES OF "LINE GENERALIZATION"

Recently, considerable attention has been paid to theoretical and conceptual principles for cartographic generalization, and for the entire process of map design. This is in part due to the recognition that such principles are a prerequisite to fully-automated systems for map design and map-making. Mark and Buttenfield (1988) discussed over-all design criteria for a cartographic expert system. They divided the map design process into three inter-related components: generalization, symbolization, and production. Generalization was characterized as a process which first models geographic phenomena, and then generalizes those models. Generalization was in turn subdivided into: simplification (including reduction, selection, and repositioning); classification (encompassing aggregation, partitioning, and overlay); and enhancement (including smoothing, interpolation, and reconstruction). (For definitions and further discussions, see Mark and Buttenfield, 1988.) Although Mark and Buttenfield's discussion of the modeling phase emphasized a phenomenon-based approach, they did not exclude statistical or other phenomenon-independent approaches. Weibel and Buttenfield
(1988) extended this discussion, providing much detail, and emphasizing the requirements for mapping in a geographic information systems (GIS) environment.

McMaster and Shea (1988) focussed on the generalization process. They organized the top level of their discussion around three questions: Why do we generalize? When do we generalize? How do we generalize? These can be stated more formally as intrinsic objectives, situation assessment, and spatial and attribute transformations, respectively (McMaster and Shea, 1988, p. 241). The rest of their paper concentrated on the first question; this essay will review such issues briefly, but is more concerned with their third objective.

Reduction of Data Volume

Many digital cartographic line-processing procedures have been developed to reduce data volumes. This process has at times been rather aptly termed "line reduction". In many cases, the goal is to eliminate redundant data while changing the geometry of the line as little as possible; this objective is termed "maintaining spatial accuracy" by McMaster and Shea (1988, p. 243). Redundant data commonly occur in cartographic line processing when digital lines are acquired from maps using "stream-mode" digitizing (points sampled at pseudo-constant intervals in x, y, distance, or time); similarly, the initial output from vectorization procedures applied to scan-digitized maps often is even more highly redundant.

One stringent test of a line reduction procedure might be: "can a computer-drafted version of the lines after processing be distinguished visually from the line before processing, or from the line on the original source document?" If the answer to both of these questions is "no", and yet the number of points in the line has been reduced, then the procedure has been successful. A quantitative measure of performance would be to determine the perpendicular distance to the reduced line from each point on the original digital line; for a particular number of points in the reduced line, the lower the root-mean-squared value of these distances, the better is the reduction. Since the performance of the algorithm can often be stated in terms of minimizing some statistical measure of "error", line reduction may be considered to be a kind of "statistical generalization", a term introduced by Brassel and Weibel (in press) to described minimum-change simplifications of digital elevation surfaces.
Preservation of Visual Appearance and Recognizability

As noted above, Brassel and Weibel (in press) distinguish statistical and cartographic generalization. "Cartographic generalization is used only for graphic display and therefore has to aim at visual effectiveness" (Brassel and Weibel, in press). A process with such an aim can only be evaluated through perceptual testing involving subjects representative of intended map users; few such studies have been conducted, and none (to my knowledge) using generalization procedures designed to preserve visual character rather than merely to simplify geometric form.

Preservation of Geographic Features and Relations

Pannekoek (1962) discussed cartographic generalization as an exercise in applied geography. He repeatedly emphasized that individual cartographic features should not be generalized in isolation or in the abstract. Rather, relations among the geographic features they represent must be established, and then should be preserved during scale reduction. A classic example, presented by Pannekoek, is the case of two roads and a railway running along the floor of a narrow mountain valley. At scales smaller than some threshold, the six lines (lowest contours on each wall of the valley; the two roads; the railway; and the river) cannot all be shown in their true positions without overlapping. If the theme of the maps requires all to be shown, then the other lines should be moved away from the river, in order to provide a distinct graphic image while preserving relative spatial relations (for example, the railway is between a particular road and the river). Pannekoek stressed the importance of showing the transportation lines as being on the valley floor. Thus the contours too must be moved, and higher contours as well; the valley floor must be widened to accommodate other map features (an element of cartographic license disturbing to this budding geomorphometer when J. Ross Mackay assigned the article in a graduate course in 1972!). Nickerson and Freeman (1986) discussed a program that included an element of such an adjustment.

A twisting mountain highway provides another kind of example. Recently, when driving north from San Francisco on highway 1, I was startled by the extreme sinuosity of the highway; maps my two major publishing houses gave little hint, showing the road as almost straight as it ran from just north of the Golden Gate bridge westward to the coast. The twists and turns of the road were too small to show at the map scale, and I have little doubt that positional accuracy was maximized by drawing a fairly straight line following the road's "meander axis". The solution used on some Swiss road maps seems better; winding mountain highways are represented by sinuous lines on the map. Again, I have no doubt that, on a 1:600,000 scale map,
the twists and turns in the cartographic line were of a far higher amplitude that the actual bends, and that the winding road symbols had fairly large positional errors. However, the character of the road is clearly communicated to a driver planning a route through the area. In effect, the road is categorized as a "winding mountain highway", and then represented by a "winding mountain highway symbol", namely a highway symbol drafted with a high sinuosity. Positional accuracy probably was sacrificed in order to communicate geographic character.

A necessary prerequisite to geographic line generalization is the identification of the kind of line, or more correctly, the kind of phenomenon that the line represents (see Buttenfield, 1987). Once this is done, the line may in some cases be subdivided into component elements. Individual elements may be generalized, or replaced by prototypical exemplars of their kinds, or whole assemblages of sub-parts may be replaced by examples of their superordinate class. Thus is a rich area for future research.

GEOGRAPHICAL LINE GENERALIZATION

Geographic phenomena which are represented by lines on topographic and road maps are discussed in this section. (Lines on thematic maps, especially "categorical" or "area-class" boundaries, will almost certainly prove more difficult to model than the more concrete features represented by lines on topographic maps, and are not included in the current discussion.) One important principle is:

many geographic phenomena inherit components of their geometry from features of other kinds.

This seems to have been discussed little if at all in the cartographic literature. Because of these tendencies toward inheritance of geometric structure, the sequence of sub-sections here is not arbitrary, but places the more independent (fundamental) phenomena first, and more derived ones later.

Topographic surfaces (contours)

Principles for describing and explaining the form of the earth's surface are addressed in the science of geomorphology. Geomorphologists have identified a variety of terrain types, based on independent variables such as rock structure, climate, geomorphic process, tectonic effects, and stage of development. Although selected properties of topographic surfaces may be mimicked by statistical surfaces such as fractional Brownian models, a kind of fractal (see Goodchild and Mark 1987 for a review), detailed models
of the geometric character of such surfaces will require the application of knowledge of geomorphology. Brassel and Weibel (in press) clearly make the case that contour lines should never be generalized individually, since they are parts of surfaces; rather, digital elevation models must be constructed, generalized, and then re-contoured to achieve satisfactory results, either statistically or cartographically.

Streams

Geomorphologists divide streams into a number of categories. Channel patterns are either straight, meandering, or braided; there are sub-categories for each of these. Generally, streams run orthogonal to the contours, and on an idealized, smooth, single-valued surface, the stream lines and contours for duals of each other. The statistics of stream planform geometry have received much attention in the earth science literature, especially in the case of meandering channels (see O'Neill, 1987). Again, phenomenon-based knowledge should be used in line generalization procedures; in steep terrain, stream/valley generalization is an intimate part of topographic generalization (cf. Brassel and Weibel, in press).

Shorelines

In a geomorphological sense, shorelines might be considered to "originate" as contours, either submarine or terrestrial. A clear example is a reservoir: the shoreline for a fixed water level is just the contour equivalent to that water level. Any statistical difference between the shoreline of a new reservoir, as drawn on a map, and a nearby contour line on that same map is almost certainly due to different construction methods or to different cartographic generalization procedures used for shorelines and contours. Goodchild's (1982) analysis of lake shores and contours on Random Island, Newfoundland, suggests that, cartographically, shorelines tend to be presented in more detail (that is, are relatively less generalized), while contours on the same maps are smoothed to a greater degree. As sea level changes occur over geologic time, due to either oceanographic or tectonic effects, either there is a relative sea-level rise, in which case a terrestrial contour becomes the new shoreline, or a relative sea-level fall, to expose a submarine contour as the shoreline.

Immediately upon the establishment of a water level, coastal geomorphic processes begin to act on the resulting shoreline; the speed of erosion depends on the shore materials, and on the wave, wind, and tidal environment. It is clear that coastal geomorphic processes are scale-dependent, and that the temporal and spatial scales of such processes are functionally linked. Wave refraction
tends to concentrate wave energy at headlands (convexities of the land), whereas energy per unit length of shoreline is below average in bays. Thus, net erosion tends to take place at headlands, whereas net deposition occurs in the bays. On an irregular shoreline, beaches and mudflats (areas of deposition) are found largely in the bays. The net effect of all this is that shorelines tend to straighten out over time. The effect will be evident most quickly at short spatial scales.

Geomorphologists have divided shorelines into a number of types or classes. Each of these types has a particular history and stage, and is composed of members from a discrete set of coastal landforms. Beaches, rocky headlands, and spits are important components. Most headlands which are erosional remnants are rugged, have rough or irregular shorelines, and otherwise have arbitrary shapes determined by initial forms, rock types and structures, wave directions, et cetera. Spits and beaches, however, have forms with a much more controlled (less variable) geometry. For example, the late Robert Packer of the University of Western Ontario found that many spits are closely approximated by logarithmic spirals (Packer, 1980).

Political and Land Survey Boundaries

Most political boundaries follow either physical features or lines of latitude or longitude. Both drainage divides (for example, the France-Spain border in the Pyrenees, or southern part of the British Columbia-Alberta in the Rocky Mountains) and streams (there are a great many many examples) are commonly used as boundaries. The fact that many rivers are dynamic in their planform geometry leads to interesting legal and/or cartographic problems. For example, the boundary between Mississippi and Louisiana is the midline of the Mississippi River when the border was legally established more than a century ago, and does not correspond with the current position of the river.

In areas which were surveyed before they were settled by Europeans, rectangular land survey is common. Then, survey boundaries may also be used as boundaries for minor or major political units. Arbitrary lines of latitude or longitude also often became boundaries as a result of negotiations between distant colonial powers, or between those powers and newly-independent former colonies. An example is the Canada - United States boundary in the west, which approximates the 49th parallel of latitude from Lake-of-the-Woods to the Pacific. Many state boundaries in the western United States are the result of the subdivision of larger territories by officials in Washington. Land survey boundaries are rather "organic" and irregular in the metes-and-bounds systems of most of the original 13 colonies of the United States, and in many other parts of the world. They are often much more rectangular in
the western United States, western Canada, Australia, and other "pre-
surveyed" regions.

Roads

Most roads are constructed according to highway engineering codes, which limit the tightness of curves for roads of certain classes and speeds. These engineering requirements place smoothness constraints on the short-scale geometry of the roads; these constraints are especially evident on freeways and other high-speed roads, and should be determinable from the road type, which is included in the USGS DLG feature codes and other digital cartographic data schemes. However, the longer-scale geometry of these same roads is governed by quite different factors, and often is inherited from other geographic features.

Some roads are "organic", simply wandering across country, or perhaps following older walking, cattle, or game trails. However, many roads follow other types of geographic lines. Some roads "follow" rivers, and others "follow" shorelines. In the United States, Canada, Australia, and perhaps other countries which were surveyed before European settlement, many roads follow the survey lines; in the western United States and Canada, this amounts to a 1 by 1 mile grid (1.6 by 1.6 km) of section boundaries, some or all of which may have actual roads along them. Later, a high-speed, limited access highway may minimize land acquisition costs by following the older, survey-based roadways where practical, with transition segments where needed to provide sufficient smoothness (for example, Highway 401 in south-western Ontario).

A mountain highway also is an example of a road which often follows a geographic line, most of the time. In attempting to climb as quickly as possible, subject to a gradient constraint, the road crosses contours at a slight angle which can be calculated from the ratio of the road slope to the hill slope. [The sine of the angle of intersection (on the map) between the contour and the road is equal to the ratio of the road slope to the hill slope, where both slopes are expressed as tangents (gradients or percentages).] Whenever the steepness of the hill slope is much greater than the maximum allowable road gradient, most parts of the trace of the road will have a very similar longer-scale geometry to a contour line on that slope. Of course, on many mountain highways, such sections are connected by short, tightly-curved connectors of about 180 degrees of arc, when there is a "switch-back", and the hillside switches from the left to the right side of the road (or the opposite).
Railways

Railways have an even more constrained geometry than roads, since tight bends are never constructed, and gradients must be very low. Such smoothness should be preserved during generalization, even if curves must be exaggerated in order to achieve this.

Summary

The purpose of this essay has not been to criticize past and current research on computerized cartographic line generalization. Nor has it been an attempt to define how research in this area should be conducted in the future. Rather, it has been an attempt to move one (small) step toward a truly "geographic" approach to line generalization for mapping. It is a bold assertion on my part to state that, in order to successfully generalize a cartographic line, one must take into account the geometric nature of the real-world phenomenon which that cartographic line represents, but nevertheless I assert just that. My main purpose here is to foster research to achieve that end, or to debate on the validity or utility of my assertions.

Acknowledgements

I wish to thank Bob McMaster for the discussions in Sydney that convinced me that it was time for me to write this essay, Babs Buttenfield for many discussions of this material over recent years, and Rob Weibel and Mark Monmonier for their comments on earlier drafts of the material presented here; the fact that each of them would dispute parts of this essay does not diminish my gratitude to them. The essay was written partly as a contribution to Research Initiative #3 of the National Center for Geographic Information and Analysis, supported by a grant from the National Science Foundation (SES-88-10917); support by NSF is gratefully acknowledged. Parts of the essay were written while Mark was a Visiting Scientist with the CSIRO Centre for Spatial Information Systems, Canberra, Australia.

References


DATA COMPRESSION AND CRITICAL POINTS DETECTION
USING NORMALIZED SYMMETRIC SCATTERED MATRIX

Khagendra Thapa B.Sc. B.Sc(Hons) CNAA, M.Sc.E. M.S. Ph.D.
Department of Surveying and MappingFennis State University
Big Rapids, Michigan 49307

BIOGRAPHICAL SKETCH

Khagendra Thapa is an Associate Professor of Surveying and Mapping at Ferris State University. He received his B.SC. in Mathematics, Physics, and Statistics from Tribhuvan University Kathmandu, Nepal and B.SC.(Hons.) CNAA in Land Surveying from North East London Polytechnic, England, M.S.C.E. in Surveying Engineering from University of New Brunswick, Canada and M.S. and Ph.D. in Geodetic Science from The Ohio State University. He was a lecturer at the Institute of Engineering, Kathmandu Nepal for two years. He also held various teaching and research associate positions both at The Ohio State University and University of New Brunswick.

ABSTRACT

The problems of critical points detection and data compression are very important in computer assisted cartography. In addition, the critical points detection is very useful not only in the field of cartography but in computer vision, image processing, pattern recognition, and artificial intelligence. Consequently, there are many algorithms available to solve this problem but none of them are considered to be satisfactory. In this paper, a new method of finding critical points in digitized curve is explained. This technique, based on the normalized symmetric scattered matrix is good for both critical points detection and data compression. In addition, the critical points detected by this algorithm are compared with those detected by humans.

INTRODUCTION

The advent of computers have had a great impact on mapping sciences in general and cartography in particular. Now-a-days more and more existing maps are being digitized and attempts have been made to make maps automatically using computers. Moreover, once we have the map data in digital form we can make maps for different purposes very quickly and easily. Usually, the digitizers tend to digitize more data than what is required to adequately represent the feature. Therefore, there is a need for data compression without destroying the character of the feature. This can be achieved by the process of critical points detection in the digital data. There are many algorithms available in the literature for the purpose of critical points detection. In this paper, a new method of critical points detection is described which is efficient and has a sound theoretical basis as it uses the eigenvalues of the Normalized Symmetric Scattered (NSS) matrix derived from the digitized data.

DEFINITION OF CRITICAL POINTS

Before defining the critical points, it should be noted that critical points in a digitized curve are of interest not only in the field of cartography but also in other disciplines such as Pattern Recognition, Image Processing, Computer Vision, and Computer Graphics. Marino (1979) defined critical points as "Those points which remain more or less fixed in position, resembling a precis of the written essay, capture the nature or character of the line".

In Cartography one wants to select the critical points along a digitized line so that one can retain the basic character of the line. Researchers both in the field of Computer Vision and Psychology have claimed that the maxima, minima and zeroes of curvature are sufficient
to preserve the character of a line. In the field of Psychology, Attneave (1954) demonstrated with a sketch of a cat that the maxima of curvature points are all one needs to recognize a known object. Hoffman and Richards (1982) suggested that curves should be segmented at points of minimum curvature. In other words, points of minimum curvature are the critical points. They also provided experimental evidence that humans segmented curves at points of curvature minima. Because the minima and maxima of a curve depend on the orientation of the curve, the following points are considered as critical points:

1. curvature maxima
2. curvature minima
3. end points
4. points of intersection.

It should be noted that Freeman (1978) also includes the above points in his definition of critical points.

Hoffman and Richards (1982) state that critical points found by first finding the maxima, minima, and zeroes of curvature are invariant under rotations, translations, and uniform scaling. Marimont (1984) has experimentally proved that critical points remain stable under orthographic projection.

The use of critical points in the fields of Pattern Recognition, and Image Processing has been suggested by Brady (1982), and Duda and Hart (1973). The same was proposed for Cartography by Solovitskiy (1974), Marino (1978), McMaster (1983), and White (1985).

Importance of Critical Point Detection in Line Generalization

Cartographic line generalization has hitherto been a subjective process. When one wants to automate a process which has been vague and subjective, many difficulties are bound to surface. Such is the situation with Cartographic line generalization. One way to tackle this problem would be to determine if one can quantify it (i.e. make it objective) so that it can be solved using a digital computer. Many researchers such as Solovitskiy (1974), Marino (1979), and White (1985) agree that one way to make the process of line generalization more objective is to find out what Cartographers do when they perform line generalization by hand? In addition, find out what in particular makes the map lines more informative to the map readers. Find out if there is anything in common between the map readers and map makers regarding the interpretation of line character.

Marino (1979) carried out an empirical experiment to find if Cartographers and non-cartographers pick up the same critical points from a line. In the experiment, she took different naturally occurring lines representing various features. These lines were given to a group of Cartographers and a group of non-cartographers who were asked to select a set of points which they consider to be important to retain the character of the line. The number of points to be selected was fixed so that the effect of three successive levels or degrees of generalization could be detected. She performed statistical analysis on the data and found that cartographers and non-cartographers were in close agreement as to which points along a line must be retained so as to preserve the character of these lines at different levels of generalization.
When one says one wants to retain the character of a line what he/she really means is that he/she wants to preserve the basic shape of the line as the scale of representation decreases. The purpose behind the retention of the basic shape of the line is that the line is still recognized as a particular feature—river, coastline or boundary despite of the change in scale. The assumption behind this is that the character of different types of line is different. That is to say that the character of a coastline is different from that of a road. Similarly, the character of a river would be different from that of a transmission line and so on.

The fact that during the process of manual generalization one retains the basic shape of the feature has been stated by various veteran Cartographers. For example, Keates (1973) states, "...each individual feature has to be simplified in form by omitting minor irregularities and retaining only the major elements of the shape". Solovitskiy (1974) identified the following quantitative and qualitative criteria for a correct generalization of lines:

1. The quantitative characteristics of a selection of fine details of a line.
2. Preservation of typical tip angles and corners
3. Preservation of the precise location of the basic landform lines.
4. Preservation of certain characteristic points. 5. Preservation of the alternation frequency and specific details.

He further states "The most important qualitative criteria are the preservation of the general character of the curvature of a line, characteristic angles, and corners...". In the above list, what Solovitskiy is basically trying to convey is that he wants to retain the character of a feature by preserving the critical points. Buttenfield (1985) also points out the fact that Cartographers try to retain the basic character of a line during generalization. She states "...Cartographer's attempt to cope objectively with a basically inductive task, namely, retaining the character of a geographic feature as it is represented at various Cartographic reductions".

Boyle (1970) suggested that one should retain the points which are more important (i.e. critical points) during the process of line generalization. He further suggested that these points should be hierarchical and should be assigned weights (1-5) to help Cartographers decide which points to retain.

Campbell (1984) also observes the importance of retaining critical features. He states, "One means of generalization involves simply selecting and retaining the most critical features in a map and eliminating the less critical ones". The fact that retention of shape is important in line generalization is also included in the definition of line generalization. The DMA (Defense Mapping Agency) definition states as "Smoothing the character of features without destroying their visible shape". Tobler as referenced in Steward (1974) also claims that the prime function of generalization is "...to capture the essential characteristics of...a class of objects, and preserve these characteristics during the change of scale".

Advantages of Critical Point Detection

According to Pavlidis and Horowitz (1974), Roberge (1984), and McMaster (1983) the
detection and retention of critical points in a digital curve has the following advantages:

1. Data compaction as a result plotting or display time will be reduced and less storage will be required.
2. Feature extraction.
4. Problems in plotter resolution due to scale change will be avoided.
5. Quicker vector to faster conversion and vice-versa.
6. Faster choropleth shading. This means shading color painting the polygons.

Because of the above advantages, research in this area is going on in various disciplines such as Computer Science, Electrical Engineering, Image Processing, and Cartography.

**Literature Review**

The proliferation of computers not only has had a great impact on existing fields of studies but also created new disciplines such as Computer Graphics, Computer Vision, Pattern Recognition, Image Processing, Robotics etc. Computers play an ever increasing role in modern day automation in many areas. Like many other disciplines, Mapping Sciences in general and Cartography in particular have been greatly changed due to the use of computers. It is known from experience that more than 80% of a map consists of lines. Therefore, when one talks about processing maps, one is essentially referring to processing lines. Fortunately, many other disciplines such as Image Processing, Computer Graphics, and Pattern Recognition are also concerned with line processing. They might be interested in recognizing shapes of various objects, industrial parts recognition, feature extraction or electrocardiogram analysis etc.

Whatever may be the objective of line processing and whichever field it may be, there is one thing in common viz: it is necessary to retain the basic character of the line under consideration. As mentioned above one needs to detect and retain the critical points in order to retain the character of a line. There is a lot of research being carried out in all the above disciplines as to the detection of critical points. Because the problem of critical points detection is common to so many disciplines, it has many nomenclatures. A number of these nomenclatures (Wall and Danielson, 1984), (Dunham, 1986), (Imai and Iri, 1986), (Anderson and Bezdek, 1984), (Herkommer, 1985), (Freeman and Davis, 1977), (Rosenfeld and Johnston, 1973), (Rosenfeld and Thurston, 1971), (Duda and Hart, 1973), (Opheim, 1982), (Williams, 1980), (Roberge, 1984), (Pavlidis and Horowitz, 1974), (Fischler and Bolles, 1983,1986), (Dettori and Falcidieno, 1982), (Reumann and Witkam, 1974), (Sklansky and Gonzlaz, 1980), (Sharma and Shanker, 1978), (Williams, 1978) are listed below:

1. Planer curve segmentation
2. Polygonal Approximation
3. Vertex Detection
4. Piecewise linear approximation
5. Corner finding
6. Angle detection
7. Line description
8. Curve partitioning
9. Data compaction
10. Straight line approximation
11. Selection of main points
12. Detection of dominant points
13. Determination of main points.

Both the amount of literature available for the solution of this problem and its varying nomenclature indicate the intensity of the research being carried out to solve this problem. It is recognized by various researchers (e.g. Fischler and Bolles, 1986) that the problem of critical points detection is in fact a very difficult one and it still is an open problem. Similarly, the problem of line generalization is not very difficult if carried out manually but becomes difficult if one wants to do it by computer. Because of the subjective nature of this problem and due to the lack of any criteria for evaluation of line generalization, it has been very difficult to automate this process. Recently some researchers for example (Marino, 1979) and (White, 1985) have suggested that one should first find critical points and retain them in the process of line generalization.

Algorithms for Finding Critical Points

As noted in the previous section, there are many papers published on critical points detection which is identified by different names by different people. It should, however, be noted that the detection is not generic but, as indicated by Fischler and Bolles (1986) depends on the following factors:

1. purpose
2. vocabulary
3. data representation
4. past experience of the 'partitioning instrument'. In cartography it would mean the past experience of the cartographer.

It is interesting to note that the above four factors are similar to the controls of line generalization that Robinson et al. (1985) have pointed out. However, the fourth factor viz: past experience and mental stability of the Cartographer is missing from the latter list.

THE NATURE OF SCATTER MATRICES AND THEIR EIGENVALUES

Consider the geometry of the quadratic form associated with a sample covariance matrix. Suppose $P = (p_1, p_2, \ldots, p_n)$ be a finite data set in $\mathbb{R}^2$ and $P$ is a sample of $n$ independently and identically distributed observations drawn from real two dimensional population.

Let $(\mu, \sigma^2)$ denote the population mean vector and variance matrix and let $(\mu_P, \sigma_P)$ be the corresponding sample mean vector and sample covariance matrix these are then given by (Uotila, 1986)

\[
\begin{align*}
\mu_P &= \frac{1}{n} \sum p_i; \\
\sigma_P &= \frac{1}{n-1} \sum (p_i - \mu_P)(p_i - \mu_P)
\end{align*}
\]

Multiply both sides of the equation for $\sigma_P$ by $(n - 1)$ and denote the RHS by $S_P$ viz:
The matrices $S$ and $V$ are both 2x2 symmetric and positive semi-definite. Since these matrices are multiples of each other they share identical eigen-spaces. According to Anderson and Bezdek (1983) one can use the eigenvalue and eigenvector structure of $S$ to extract the shape information of the data set it represents. This is because the shape of the data set is supposed to mimic the level shape of the probability density function $f(x)$ of $x$. For example, if the data set is bivariate normal, $S_p$ has two real, non-negative eigenvalues. Let these eigenvalues be $\lambda_1$ and $\lambda_2$. Then the following possibilities exist (Anderson and Bezdek, 1983):

1. If both $\lambda_1$ and $\lambda_2 = 0$, then the data set $P$ is degenerate, and $S_p$ is invertible and there exist with probability 1, constants $a$, $b$, and $c$ such the $ax + by + c = 0$. In this case the sample data in $P$ lie on a straight line.

2. If $\lambda_1 > \lambda_2 > 0$, then the data set represent an elliptical shape.

3. If $\lambda_1 = \lambda_2 > 0$, then the sample data set in $P$ represent a circle.

**EIGENVALUES OF THE NORMALIZED SYMMETRIC SCATTER MATRIX (NSS)**

Supposing that one has the following data:

$$P = (P_1, P_2, \ldots, P_n)$$

where $P_i = (x_i, y_i)$

Then the normalized scattered matrix $A$ is defined as

$$A = \frac{1}{\text{Deno}} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \frac{S_p}{\text{trace}(S_p)} \quad (3)$$

For the above data set $A$ is given by:

$$\text{Deno} = \sum ( (x_i - x_m)^2 + (y_i - y_m)^2 ) \quad (4)$$

$$a_{11} = \frac{1}{\text{Deno}} \sum (x_i - x_m)^2$$

$$a_{12} = \frac{1}{\text{Deno}} \sum (x_i - x_m)(y_i - y_m)$$

$$a_{21} = \frac{1}{\text{Deno}} \sum (x_i - x_m)(y_i - y_m)$$

$$a_{22} = \frac{1}{\text{Deno}} \sum (y_i - y_m)^2 \quad (5)$$

where $\mathbf{v}_x = (x_m, y_m)$ is the mean vector defined as

$$x_m = \sum x / n, \quad y_m = \sum y / n \quad (6)$$

Note that the denominator in (3) will vanish only when all the points under consideration are identical.

The characteristic equation of $A$ is given by:

$$| A - \lambda I | = 0 \quad (7)$$

which may be written as (for 2x2 matrix)
which may be written as (for 2x2 matrix)
\[ \lambda^2 - \text{trace}(A)\lambda + \text{Det}(A) = 0 \]  
where \text{Det}(A) = \text{Determinant of } A.

By design the trace of A is equal to 1. Hence the characteristics equation of A reduces to
\[ \lambda^2 + \text{Det}(A) = 0 \]

The roots of this equation are the eigenvalues and are given by:
\[ \lambda_1 = \frac{1 + \sqrt{1 - 4\text{Det}(A)}}{2} \quad \text{and} \quad \lambda_2 = \frac{1 - \sqrt{1 - 4\text{Det}(A)}}{2} \]

For convenience put \( D_x = \sqrt{1 - 4\text{Det}(A)} \), then
\[ \lambda_1 = \frac{1 + D_x}{2} \]
\[ \lambda_2 = \frac{1 - D_x}{2} \]

Now \( \lambda_1 \) and \( \lambda_2 \) satisfy the following two conditions:
\[ \lambda_1 + \lambda_2 = 1 \]

Since the sum of the roots of an equation of the form
\[ ax^2 + bx + c = 0 \]  are \( \lambda_1 + \lambda_2 = -\frac{b}{a} \)

Subtracting (12) from (11), one obtains
\[ \lambda_1 - \lambda_2 = D_x \]

Since the eigenvalues \( \lambda_1 \) and \( \lambda_2 \) satisfy the equations (13) and (14) the three cases discussed previously reduce to the following from (Anderson and Bezdek, 1983):

1. The data set represent a straight line if and only if \( D_x = 1 \)
2. The data set represent an elliptical shape if and only if \( 0 \leq D_x \leq 1 \)
3. The data set represent a circular shape if \( D_x = 0 \).

ALGORITHM TO DETECT CRITICAL POINTS USING NSS MATRIX

The fact that the analysis of the eigenvalues of the NSS matrix can be used to extract shape of the curve represented by the data set, may be exploited to detect critical points in the digital curve. Assuming that the data is gross error free, and devoid of excessive noise, one can outline the algorithm to detect critical points in the following steps:
3. If $D_x$ is greater than a certain tolerance (e.g. 0.95) add one more point to the data and repeat from step 2.

4. If $D_x$ is less than the tolerance point, point 2 is a critical point. Retain point 2 and repeat the process from step 1 with point two as the new starting point.

5. Repeat the process until the end of the data set is reached.

Results of Critical Points Detection by NSS Matrix

The algorithm discussed in the previous section is useful in detecting the critical points in vector data. The only parameter involved in this technique is $D_x$ which was defined earlier. By varying the value of $D_x$ between say 0.8 to 1.0 one can get a varying amount of detail in a curve. Figure 1 shows the selected critical points for the test figure for $D_x = 0.96$.

![Figure 1: Results of Critical Points Selection by NSS Matrix. There were 50 points selected.](image)

There are 50 points selected in this figure. It is clear from the figure that this method will be very useful for compression of digitized data since it retains the overall shape of the curve without retaining the unnecessary points.

COMPARISON BETWEEN MANUAL AND ALGORITHMIC CRITICAL POINTS DETECTION

In this section, results of critical points detection in the test figure by a group of people are given. These results are then compared with the results obtained from the NSS matrix technique of critical points detection.

MANUAL CRITICAL POINTS DETECTION: THE EXPERIMENT

In order to find if the NSS matrix method critical points detection can mimic humans or not, the test figure was given to a group of 25 people who had at least one course in Cartography.
In addition, they were told about the nature of critical points. In the experiment, they were asked to select not more than 50 points from the test figure. The results of critical points detection by the above group are shown in figure 2. In figure 2 each dot represents 5 respondents. A point was rejected if it was selected by less than four respondents.

A comparison of figures 1 and 2 reveals that the critical points selected by the humans are almost identical with those selected by NSS matrix algorithm. Only five or less people slightly disagreed in the selection of a few points with NSS matrix algorithm. It should be pointed out that the results of critical points selection could have been different if the respondents were asked to select all the critical points in the curve. However, the NSS matrix algorithm also can be made to detect different levels of critical points by simply changing the value of the parameter $D_\times$.

![Figure 2: Points selected by respondents. Each dot represents five respondents](image)

**CONCLUSION**

1. The analysis of the eigenvalues of the normalized symmetric scattered matrix provides an useful way of detecting critical points in digitized curves. Consequently, this technique may be used for data compression for digitized curves.

2. The NSS matrix algorithm for critical points detection can mimic humans in terms of critical points detection in a digitized curve.

**ACKNOWLEDGEMENT**

I express my gratitude to Dr. Toni Schenk for providing partial funding for this research under the seed grant 22214 and NASA project 719035.
REFERENCES


Duda, R.O. and P.E. Hart (1973), Pattern Classification and Scene Analysis, Willey Interscience.


Publication, Winnapeg, Translated from German by W.D. Hoskins and H.W. Sagar.


