Metric Operations on Fuzzy Spatial Objects in Databases

Markus Schneider
FernUniversität Hagen, Praktische Informatik IV
58084 Hagen, Germany
markus.schneider@fernuni-hagen.de

ABSTRACT

Uncertainty management for geometric data is currently an important problem for (extensible) databases in general and for spatial databases, image databases, and GIS in particular. In these systems, spatial data are traditionally kept as determinate and sharply bounded objects; the aspect of spatial vagueness is not and cannot be treated by these systems. However, in many geometric and geographical database applications there is a need to model spatial phenomena rather through vague concepts due to indeterminate and blurred boundaries. Following previous work, we first describe a data model for fuzzy spatial objects including data types for *fuzzy regions* and *fuzzy lines*. We then, in particular, study the important class of *metric operations* on these objects.

Keywords

Fuzzy spatial data type, fuzzy region, fuzzy line, real-valued metric operation, fuzzy-valued metric operation, fuzzy sets

1. INTRODUCTION

Representing, storing, quering, and manipulating spatial information is important for many non-standard database applications. But so far, *spatial data modeling* has implicitly assumed that the extent and hence the borders of spatial objects are precisely determined ("boundary syndrome"). Special data types called *spatial data types* (see [7] for a survey) have been designed for modeling these data in databases. We will denote this kind of entities as *crisp* spatial objects.

In practice, however, there is no apparent reason for the whole boundary of a region to be determined. On the contrary, the feature of *spatial vagueness* is inherent to many geographic data [3]. Many geographical application examples illustrate that the boundaries of spatial objects (like geological, soil, and vegetation units) can be partially or totally indeterminate and blurred; e.g., human concepts like "the Indian Ocean" or "Southern England" are implicitly vague. In this paper we focus on a special kind of spatial vagueness called *fuzziness*. Fuzziness captures the property of many spatial objects in reality which do not have sharp boundaries or whose boundaries cannot be precisely determined. Examples are natural, social, or cultural phenomena like land features with continuously changing properties (such as population density, soil quality, vegetation, pollution, temperature, air pressure), oceans, deserts, English speaking areas, or mountains and valleys. The transition between a valley and a mountain usually cannot be exactly ascertained so that the two spatial objects "valley" and "mountain" cannot be precisely separated and defined in a crisp way. We will designate this kind of entities as *fuzzy* spatial objects.

The goal of this paper is to deal with the important class of *metric operations* on fuzzy spatial objects. Examples are the *area* operation on fuzzy regions or the *length* operation on fuzzy lines. It turns out that their definition is not as trivial as the definition of their crisp counterparts. The underlying formal object model follows the author's previous work in [8] and offers *fuzzy spatial data types* like *fuzzy regions* and *fuzzy lines* in two-dimensional Euclidean space.

Our concept to integrate fuzzy spatial data types into databases is to design them as *abstract data types* whose values can be embedded as complex entities into databases [9] and whose definition is independent of a particular DBMS data model. They can, e.g., be employed as attribute types in a relation. The future design of an SQL-like *fuzzy spatial query language* will profit from the abstract data type approach since it makes the integration of data types, predicates, and operations into SQL easier.

The metric operations described in this paper are part of a so-called *abstract model* for fuzzy spatial objects. This model focuses on the nature of the problem and on its realistic description and solution with mathematical notations; it employs infinite sets and does not worry about finite representations of objects as they are needed in computers. This is done by the so-called *discrete model* which transforms the infinite representations into finite ones and which realizes the abstract operations as algorithms on these finite representations. It is thus closer to implementation.

Section 2 discusses related work. Section 3 introduces fuzzy spatial objects. It gives some basic concepts of fuzzy set theory and then informally presents the design of fuzzy regions and fuzzy lines. Section 4 describes and formalizes metric operations on fuzzy spatial objects and identifies the two classes of real-valued and fuzzy-valued metric operations. Section 5 draws some conclusions and discusses future work.

2. RELATED WORK

Mainly two kinds of *spatial vagueness* can be identified: *uncertainty* is traditionally equated with randomness and chance occurrence and relates either to a lack of knowledge about the position and shape of an object with an existing, real boundary (positional uncertainty) or to the inability of measuring such an object precisely (measurement uncertainty). *Fuzziness*, in which we are only interested in this paper, is an intrinsic feature of an object itself and describes the vagueness of an object which certainly has an extent but which inherently cannot or does not have a precisely definable boundary (e.g., between a mountain and a valley). This kind of vagueness results from the imprecision of the meaning of a concept. Models based on fuzzy sets have, e.g., been proposed in [1, 2, 4, 8].

3. FUZZY SPATIAL OBJECTS

In this section we very briefly and informally present the basic elements of an abstract model for fuzzy spatial objects as it has been formalized in [8]. The model is based on fuzzy set theory (and fuzzy topology) whose main concepts are introduced first, as far as they are needed in this paper. Afterwards the design of spatial data types for fuzzy regions and fuzzy lines is shortly discussed.

3.1 Crisp Versus Fuzzy Sets

Fuzzy set theory [10] is an extension and generalization of Boolean set theory. It replaces the crisp boundary of a classical set by a gradual transition zone and permits partial and multiple set membership. Let X be a classical (crisp) set of objects. Membership in a classical subset A of X can then be described by the *characteristic function* $\chi_A : X \to \{0,1\}$ such that for all $x \in X$ holds $\chi_A(x) = 1$ if and only if $x \in A$ and $\chi_A(x) = 0$ otherwise. This function can be generalized such that all elements of X are mapped to the real interval [0,1] indicating the *degree of membership* of these elements in the set in question. We call $\mu_{\tilde{A}} : X \to [0,1]$ the *membership function* of $\tilde{A}$, and the set $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \,|\, x \in X\}$ is called a *fuzzy set* in X. A [*strict*] α*-cut* of a fuzzy set $\tilde{A}$ for a specified value α is the crisp set $A_\alpha\,[A^*_\alpha] = \{x \in X \mid \mu_{\tilde{A}}(x) \geq [>] \alpha \,\wedge\, 0 \leq \alpha \leq [<] 1\}$. The strict α-cut for $\alpha = 0$ is called *support* of $\tilde{A}$, i.e., $supp(\tilde{A}) = A^*_0$. A fuzzy set is *convex* if and only if each of its α-cuts is a convex set. A fuzzy set $\tilde{A}$ is said to be *connected* if its pertaining collection of α-cuts is connected, i.e., for all points P, Q of $\tilde{A}$, there exists a path lying completely within $\tilde{A}$ such that $\mu_{\tilde{A}}(R) \geq \min(\mu_{\tilde{A}}(P), \mu_{\tilde{A}}(Q))$ holds for any point R on the path.

3.2 Fuzzy Regions

We first describe some desired properties of fuzzy regions and also discuss some differences in comparison with crisp regions. After that, we informally outline a data type for fuzzy regions.

3.2.1 Generalization of Crisp to Fuzzy Regions

A very general model defines a *crisp* region as a regular closed set [5, 6, 7] in the Euclidean space $\mathbb{R}^2$. This model is closed under (appropriately defined) geometric union, intersection, and difference. Similar to the generalization of crisp sets to fuzzy sets, we strive for a generalization of crisp regions to fuzzy regions on the basis of the point set paradigm and fuzzy concepts.

Crisp regions are characterized by sharply determined *boundaries* enclosing and grouping areas with *equal* properties or attributes and separating different regions with different properties from each other; hence *qualitative* concepts play a central role. For fuzzy regions, besides the qualitative aspect, also the *quantitative* aspect becomes important, and boundaries in most cases disappear (between a valley and a mountain there is no boundary!). The distribution of attribute values within a region and transitions between different regions may be *smooth* or *continuous*. This important feature just characterizes fuzzy regions.

A classification of fuzzy regions from an application point of view together with application examples is given in [8].

3.2.2 Definition of Fuzzy Regions

We now briefly give an informal description of a data type for *fuzzy regions*. A detailed formal definition can be found in [8]. A value of type *fregion* for fuzzy regions is a *regular open fuzzy set* whose membership function $\mu_{\tilde{F}} : \mathbb{R}^2 \to [0,1]$ is *predominantly continuous*. $\tilde{F}$ is defined as open due to its vagueness and its lack of boundaries. The property of *regularity* avoids possible "geometric anomalies" (e.g., isolated or dangling line or point features, missing lines and points) of fuzzy regions. The property of $\mu_{\tilde{F}}$ to be predominantly continuous models the intrinsic smoothness of fuzzy regions where a finite number of exceptions ("continuity gaps") are allowed.

3.2.3 Fuzzy Regions As Collection of α*-Level Regions*

A "semantically richer" characterisation of fuzzy regions describes them as collections of crisp α*-level regions* [8]. Given a fuzzy region $\tilde{F}$, we represent a region F_α for an $\alpha \in [0,1]$ as the regular crisp set of points whose membership values in $\tilde{F}$ are greater than or equal to α. F_α can have holes. The α-level regions of $\tilde{F}$ are nested, i.e., if we select membership values $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_n > \alpha_{n+1} = 0$ for some $n \in \mathbb{N}$, then $F_{\alpha_1} \subseteq F_{\alpha_2} \subseteq \cdots \subseteq F_{\alpha_n} \subseteq F_{\alpha_{n+1}}$.

3.3 Fuzzy Lines

In this section we informally describe a data type for *fuzzy lines* whose detailed formal definition can be found in [8]. We start with a *simple* fuzzy line $\tilde{l}$ which is defined as a continuous curve with smooth transitions of membership grades between neighboring points of $\tilde{l}$, i.e., the membership function of $\tilde{l}$ is continuous too. The end points of $\tilde{l}$ may coincide so that loops are allowed. Self-intersections and equality of an interior with an end point, however, are prohibited. If $\tilde{l}$ is closed, the first end point must be the leftmost point to ensure uniqueness of representation.

Let S be the set of fuzzy simple lines. An S*-complex* T is a finite subset of S such that the following conditions are fulfilled. First, the elements of T do not intersect or overlap within their interior. Second, they may not be touched within their interior by an endpoint of another element. Third, isolated fuzzy simple lines are disallowed (connectivity property). Fourth, each endpoint of an element of T must belong to exactly one or more than two incident elements of T to support the requirement of maximal elements and hence to achieve minimality of representation. Fifth, the membership values of more than two elements of T with a common end point must have the same membership value; otherwise we get a contradiction saying that a point of an S-complex has more than one membership value. All conditions together define an S-complex as a *connected planar fuzzy graph* with a unique representation.

A value of the data type *fline* for fuzzy lines is then given as a finite set of disjoint S-complexes.

4. METRIC OPERATIONS ON FUZZY SPATIAL OBJECTS

A very important class of operations on spatial objects are metric operations usually getting one or two spatial objects as arguments and yielding a numerical result. They compute metric (i.e., measurable) properties such as *area* and *perimeter* and are commonly used in the analysis of spatial phenomena. While their definitions are well-known, clear, and relatively easy for crisp spatial objects, it is not always obvious how to measure metric properties of *fuzzy* spatial objects and hence how to define corresponding operations.

A central issue is whether the result of such a metric operation is a crisp number or rather a "fuzzy" number. From an application point of view both kinds of numerical results are acceptable and even desirable. A resulting single crisp number can be interpreted as an appropriately aggregated or weighted real value over all membership values of a fuzzy spatial object. An arising fuzzy number satisfies the expectation that, if the operands of a metric operation are fuzzy, then the numerical result should be fuzzy too. Therefore, we will consider the crisp (Section 4.1) and the fuzzy (Section 4.2) variant of several metric operations. Both variants have in common that they operate on fuzzy spatial objects and that they are reduced to the ordinary definitions in the crisp case.

4.1 Crisp-Valued Numerical Operations

In this section we view operations on fuzzy spatial objects that yield crisp numbers. We first present a special view on membership functions which simplifies an understanding of the metric operations discussed afterwards.

4.1.1 Membership Functions Considered as Functions of Two Variables

Essentially, a membership function for a spatial object s associates with each point $p = (x,y) \in \mathbb{R}^2$ the grade $\mu_s(p) = \mu_s((x,y))$ to which p belongs to s. A slightly modified view considers μ as a function of the two variables x and y. This view has the benefit that we can visualize how μ "works" in terms of its graph. The *graph* of μ is the graph of the equation $z = \mu(x,y)$ and comprises all three-dimensional points $(x,y,z) = (x,y,\mu(x,y))$ satisfying this equation. If s is a fuzzy region, the graph of the corresponding spatial membership function of two variables is a collection of disjoint surfaces (one for each fuzzy face) that lie above their domain s in the Euclidean plane. Each surface determines a solid or volume bounded above by the function graph and bounded below by a fuzzy face of s. If s is a fuzzy line, we obtain a collection of disjoint three-dimensional networks each consisting of a set of three-dimensional curves. As an example, Figure 1 shows the three-dimensional view of the membership function $\mu(x,y) = \exp(-x^2 - y^2)$ of a fuzzy region (showing the expansion of air pollution caused by a power station, for instance). Three-dimensional visualizations of membership functions of fuzzy spatial objects lead to an easier understanding of the metric operations discussed in the following.

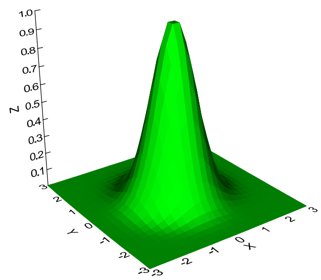

Figure 1: 3D representation of the membership function of a fuzzy region.

4.1.2 Metric Operations on Fuzzy Regions

Metric operations on fuzzy regions usually yield a real value as a result and can be summarized as a collection of functions $\gamma :$ *fregion* $\to$ *real* with different function names for γ and, of course, different semantics. Let $\tilde{F} = \{\tilde{f}_1, \ldots, \tilde{f}_n\} \in$ *fregion* where the $\tilde{f}_i$ are the fuzzy faces of $\tilde{F}$.

The definition of an *area* operator applied to $\tilde{F}$ requires that $\mu_{\tilde{F}}$ is integrable. This condition is always fulfilled since our definition of a fuzzy region requires that $\mu_{\tilde{F}}$ is continuous or at least piecewise continuous. The area of $\tilde{F}$ can be defined as the volume under the membership function $\mu_{\tilde{F}}$:

$$\begin{aligned} area(\tilde{F}) &= \iint_{(x,y)\in\mathbb{R}^2} \mu_{\tilde{F}}(x,y)\,dx\,dy \\ &= \iint_{(x,y)\in supp(\tilde{F})} \mu_{\tilde{F}}(x,y)\,dx\,dy \\ &= \sum_{i=1}^{n} \iint_{(x,y)\in supp(\tilde{f}_i)} \mu_{\tilde{f}_i}(x,y)\,dx\,dy \end{aligned}$$

Thus, the integration can be performed either over the entire Euclidean plane, or equivalently over the support of $\tilde{F}$, which is always bounded, or equivalently over the supports of the faces of $\tilde{F}$. Note that $\mu_{\tilde{f}_i}(x,y) = \mu_{\tilde{F}}(x,y)$ if $(x,y) \in supp(\tilde{f}_i)$, and 0 otherwise. Crisp holes enclosed by fuzzy faces do not cause problems during the integration process; they simply do not contribute to the double integral. Evidently, if $\tilde{F} \subseteq \tilde{G}$, we have $area(\tilde{F}) \leq area(\tilde{G})$. In particular, we obtain

$$\begin{aligned} area(\tilde{F}) &= \iint_{(x,y)\in supp(\tilde{F})} \mu_{\tilde{F}}(x,y)\,dx\,dy \\ &\leq \iint_{(x,y)\in supp(\tilde{F})} 1\,dx\,dy \\ &= area(supp(\tilde{F})) \end{aligned}$$

A special case arises if $\mu_{\tilde{F}}$ is piecewise constant and $\tilde{F}$ consists of a finite collection $\{F_{\alpha_1}, \ldots, F_{\alpha_n}\}$ of crisp α-level regions. Then the area of $\tilde{F}$ is computed as the weighted sum of the areas of all α-level regions F_{α_i}:

$$area(\tilde{F}) = \sum_{i=1}^{n} \iint_{(x,y)\in F_{\alpha_i}} \alpha_i\,dx\,dy = \sum_{i=1}^{n} \alpha_i \cdot area(F_{\alpha_i})$$

Next, we determine the height and the width of a fuzzy region. For computing the *height* (*width*) of a fuzzy region $\tilde{F}$, all maximum membership values along the y-axis (x-axis) and parallel to the x-axis (y-axis) are aggregated and (their square roots are) added up, i.e., $\tilde{F}$ is projected onto the y-axis (x-axis), and the maximum membership value is determined for each y-value (x-value).

$$height(\tilde{F}) = \int_{y\in\mathbb{R}} \max_{x\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\,dy$$

and similarly we obtain

$$width(\tilde{F}) = \int_{x\in\mathbb{R}} \max_{y\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\,dx$$

Both integrals are finite since $\tilde{F}$ has bounded support. Let $\tilde{F} = F$ be crisp. For the *height* operator we obtain $\max_{x\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y_0) = \sqrt{1} = 1$ for a fixed $y_0 \in \mathbb{R}$ if the intersection of the horizontal line $y = y_0$ with F is non-empty. Otherwise, the expression yields $\sqrt{0} = 0$. Hence, $height(F) = \int_{y\in\mathbb{R}} \max_{x\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\,dy$ is the measure of the set of y_0's such that F intersects $y = y_0$. In the case that F consists of several connected components, each component of F gives rise to a y-interval. Then $height(F)$ is the union of these intervals. If F is connected, we only obtain one interval, and $height(F)$ is just its length. Analogous thoughts hold for the *width* operator.

So far, we have not explained the meaning and the necessity of the exponent $\frac{1}{2}$ as part of the integrands. The introduction of this exponent is essential since it ensures our expectation that the area of a fuzzy region is equal or less than its height times its width, i.e.,

LEMMA 1. *area*$(\tilde{F}) \leq$ *height*$(\tilde{F}) \cdot$ *width*$(\tilde{F})$

PROOF. We can show this as follows:

$$\begin{aligned} area(\tilde{F}) &= \iint_{(x,y)\in\mathbb{R}^2} \mu_{\tilde{F}}(x,y)\,dx\,dy \\ &= \int_{y\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\,dy \cdot \int_{x\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\,dx \\ &\leq \int_{y\in\mathbb{R}} \max_{x\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\,dy \cdot \int_{x\in\mathbb{R}} \max_{y\in\mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\,dx \\ &= height(\tilde{F}) \cdot width(\tilde{F}) \end{aligned}$$

□

Without the exponent $\frac{1}{2}$ we would have $area(\tilde{F})^2 \leq height(\tilde{F}) \cdot width(\tilde{F})$ which is a completely different result and which does not correspond to our intuition. Hence, metric operations on fuzzy spatial objects do not only depend on their geometric extent but also on the nature of their membership values so that a kind of "compensation" is necessary. In accordance with the definitions for *height* and *width*, the exponent $\frac{1}{2}$ will also appear in the definitions of the following operators.

If $\tilde{F}$ consists of a finite collection $\{F_{\alpha_1}, \ldots, F_{\alpha_n}\}$ of crisp α-level regions, we obtain

$$\begin{aligned} height(\tilde{F}) &= \int_{y \in \mathbb{R}} \max\{\alpha_i^{\frac{1}{2}} \mid 1 \leq i \leq n, \exists x \in \mathbb{R} : \\ &\qquad \mu_{\tilde{F}}(x,y) = \alpha_i\}\, dy \\ width(\tilde{F}) &= \int_{x \in \mathbb{R}} \max\{\alpha_i^{\frac{1}{2}} \mid 1 \leq i \leq n, \exists y \in \mathbb{R} : \\ &\qquad \mu_{\tilde{F}}(x,y) = \alpha_i\}\, dx \end{aligned}$$

The diameter of a spatial object is defined as the largest distance between any of its points. We will give here two definitions and distinguish between the outer diameter and the inner diameter. For the computation of the outer diameter we may leave the fuzzy region; for the computation of the inner diameter we have to remain within its interior. The *outer diameter* of $\tilde{F}$ is defined as

$$outerDiameter(\tilde{F}) = \max_{u \in \mathbb{R}} \int_{u \in \mathbb{R}} \max_{v \in \mathbb{R}} \mu_{\tilde{F}}^{\frac{1}{2}}(u,v)\, du$$

where u and v are any pair of orthogonal directions and where the maximum is evaluated over all possible directions u (we can also imagine that $\tilde{F}$ is smoothly rotated within the Cartesian coordinate system). If $\tilde{F} = F$ is crisp, the value u yielding the maximum is the direction of the line along which the projection of $\tilde{F}$ has the largest size. Obviously, $height(\tilde{F}) \leq outerDiameter(\tilde{F})$ and $width(\tilde{F}) \leq outerDiameter(\tilde{F})$, and we can further conclude that $area(\tilde{F}) \leq outerDiameter(\tilde{F})^2$.

The *inner diameter* operator only relates to connected fuzzy regions. Let P and Q be any two points of $\tilde{F}$, and let π_{PQ} be a path from P to Q that lies completely in $\tilde{F}$. Such a path must exist since $\tilde{F}$ is assumed to be connected. The *inner diameter* is defined as

$$innerDiameter(\tilde{F}) = \max_{P,Q \in \mathbb{R}^2} \min_{\pi_{PQ}} \iint_{(x,y) \in \pi_{PQ}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\, dx\, dy$$

where the maximum is computed over all points P and Q of the Euclidean plane and where the minimum is determined over all paths between P and Q such that $\mu_{\tilde{F}}(R) \geq \min(\mu_{\tilde{F}}(P), \mu_{\tilde{F}}(Q))$ holds for any point R on π_{PQ}. Since $\tilde{F}$ is connected, such a path always exists.

If $\tilde{F} = F$ is crisp and $P \notin F$ or $Q \notin F$, any path π_{PQ} from P to Q will not yield the maximum. Otherwise, if both P and Q are in F, π_{PQ} must lie completely in F so that $\min_{\pi_{PQ}} \iint_{(x,y) \in \pi_{PQ}} \mu_{\tilde{F}}^{\frac{1}{2}}(x,y)\, dx\, dy = length(\pi_{PQ})$. In this case, the meaning of $innerDiameter(F)$ amounts to its standard definition as the greatest possible distance between any two points in F where only paths lying completely in F are allowed.

The relationship between inner and outer diameter is different for crisp and fuzzy regions. If $\tilde{F} = F$ is crisp, we can derive two propositions:

LEMMA 2. *If F is a connected crisp region, then $outerDiameter(F) \leq innerDiameter(F)$.*

PROOF. Select a line (which is not necessarily unique) upon which the projection of F onto the u-axis is largest. Since F is connected, this projection is an interval, and its length is given by $outerDiameter(F)$. Let us assume that P and Q are those points of F that coincide with the end points of this interval. Then the shortest path p in F between P and Q is at least as long as the straight line segment s joining P and Q. Segment s is at least as long as the interval since the interval is a projection of s. □

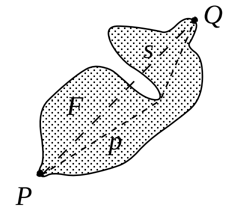

Figure 2: Example of the relationship between inner (p) and outer (s) diameter of a connected crisp region.

An example of this relationship is that s crosses the exterior of F. Then path p is longer than s (Figure 2).

LEMMA 3. *If F is a convex crisp region, then $outerDiameter(F) = innerDiameter(F)$.*

PROOF. Since F is convex, F is connected so that $outerDiameter(F) \leq innerDiameter(F)$ holds according to Lemma 2. Let P and Q be the end points of the shortest path yielding the maximum in the definition of $innerDiameter(F)$. Since F is convex, the shortest path in F between P and Q is the straight line segment joining P and Q. The projection of F on this line segment thus has length at least $innerDiameter(F)$ so that $outerDiameter(F) \geq innerDiameter(F)$. □

For convex fuzzy regions the situation is different in the sense that the outer diameter can even be greater than the inner diameter. An illustration is given in Figure 3. Let $c, d \in \mathbb{R}^{>0}$. We consider a fuzzy region $\tilde{F}$ having the membership function

$$\mu_{\tilde{F}}(x,y) = \begin{cases} \alpha_1 & \text{if } x^2 + y^2 \leq c^2 \\ \alpha_2 & \text{if } c^2 < x^2 + y^2 \leq d^2 \\ 0 & \text{otherwise} \end{cases}$$

and thus consisting of two circular, concentric α-level regions F_{α_1} and F_{α_2}. Moreover, we assume that α_1 is much larger than α_2 and that d is only slightly larger than c. If R and S are two points on the boundary of F_{α_1} which are located on opposite sides so that their straight connection passes the center of F_{α_1}, then for F_{α_1} the inner and the outer diameter are equal according to Lemma 3, and we obtain $innerDiameter(F_{\alpha_1}) = outerDiameter(F_{\alpha_1}) = 2\alpha_1 c$. If P and Q are two points on the boundary of F_{α_2}, they have the largest distance if they are located on opposite sides so that their straight connection passes the center of F_{α_2}. This is just the outer diameter of $\tilde{F}$, and we have $outerDiameter(\tilde{F}) = 2\alpha_1 c + 2\alpha_2(d - c)$. But the inner diameter of $\tilde{F}$ can be smaller if we consider a shortest path π_{PQ} between P and Q within $F_{\alpha_2} \setminus F_{\alpha_1}$. This path avoids F_{α_1} with its high membership value α_1, and we can compute the value for α_1 so that $innerDiameter(\tilde{F}) \leq outerDiameter(\tilde{F})$ holds. We must require that $length(\pi_{PQ}) \cdot \alpha_2 < 2\alpha_1 c + 2\alpha_2(d - c)$. This is the case if $\alpha_1 > \frac{\alpha_2(length(\pi_{PQ}) - 2(d-c))}{2c}$.
The observation that the outer diameter of a fuzzy region $\tilde{F}$ can be larger than its inner diameter is only valid if $\tilde{F}$ is convex:

LEMMA 4. *If $\tilde{F}$ is a convex fuzzy region, then $outerDiameter(\tilde{F}) \geq innerDiameter(\tilde{F})$.*

PROOF. If $\tilde{F}$ is a convex fuzzy region, we know due to the definition of connectedness that $\mu_{\tilde{F}}(R) \geq \min(\mu_{\tilde{F}}(P), \mu_{\tilde{F}}(Q))$ holds for any point R on the straight line segment $\overline{PQ}$ between any two

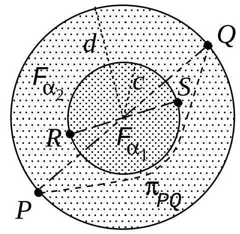

Figure 3: Example of the relationship between inner and outer diameter of a convex fuzzy region.

points P and Q. Therefore, we obtain $\iint_{(x,y)\in\overline{PQ}} \mu_{\tilde F}^{\frac{1}{2}}(x,y)\,dx\,dy \geq \iint_{(x,y)\in\pi_{PQ}} \mu_{\tilde F}^{\frac{1}{2}}(x,y)\,dx\,dy$ for the shortest path π_{PQ} between P and Q. A projection of $\tilde F$ onto the line segment $\overline{PQ}$ in any direction u yields $\int_{u\in\mathbb{R}} \max_{v\in\mathbb{R}} \mu_{\tilde F}^{\frac{1}{2}}(u,v)\,du \geq \iint_{(x,y)\in\overline{PQ}} \mu_{\tilde F}^{\frac{1}{2}}(x,y)\,dx\,dy$. Finally, we obtain

$$\begin{aligned} outerDiameter(\tilde F) &= \max_{u\in\mathbb{R}} \iint_{u\in\mathbb{R}} \max_{v\in\mathbb{R}} \mu_{\tilde F}^{\frac{1}{2}}(u,v)\,du \\ &\geq \iint_{(x,y)\in\overline{PQ}} \mu_{\tilde F}^{\frac{1}{2}}(x,y)\,dx\,dy \\ &\geq \iint_{(x,y)\in\pi_{PQ}} \mu_{\tilde F}^{\frac{1}{2}}(x,y)\,dx\,dy \\ &= innerDiameter(\tilde F) \end{aligned}$$

□

In all other cases where $\tilde F$ is a more general fuzzy region, no general statement can be made about the relationship between inner and outer diameter.

Based on the concept of outer diameter we can specify two other operators which characterize the shape of a fuzzy region. They rate the opposite geometric properties "elongatedness" and "roundness" and are defined in terms of the proportion of the minor outer diameter to the major outer diameter. The first operator is given as

$$elongatedness(\tilde F) = 1 - \frac{\min_{u\in\mathbb{R}} \int_{u\in\mathbb{R}} \max_{v\in\mathbb{R}} \mu_{\tilde F}^{\frac{1}{2}}(u,v)\,du}{outerDiameter(\tilde F)}$$

The geometric property "roundness" can be regarded as the complement of "elongatedness":

$$roundness(\tilde F) = 1 - elongatedness(\tilde F)$$

The next operator of interest computes the perimeter of a fuzzy region $\tilde F$. Assuming that the membership function $\mu_{\tilde F}$ is continuous and that $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ denote the partial derivatives of μ with respect to x and y, respectively, we can define the perimeter of $\tilde F$ as

$$\begin{aligned} perimeter(\tilde F) &= \iint_{(x,y)\in\mathbb{R}^2} \sqrt{\left(\frac{\partial}{\partial x}\mu_{\tilde F}^{\frac{1}{2}}\right)^2 + \left(\frac{\partial}{\partial y}\mu_{\tilde F}^{\frac{1}{2}}\right)^2}\,dx\,dy \\ &= \sum_{i=1}^{n} \iint_{(x,y)\in\mathbb{R}^2} \sqrt{\left(\frac{\partial}{\partial x}\mu_{\tilde f_i}^{\frac{1}{2}}\right)^2 + \left(\frac{\partial}{\partial y}\mu_{\tilde f_i}^{\frac{1}{2}}\right)^2}\,dx\,dy \end{aligned}$$

One can prove that, if $\tilde F \subseteq \tilde G$, we obtain $perimeter(\tilde F) \leq perimeter(\tilde G)$. If the membership function $\mu_{\tilde F}$ is piecewise constant so that $\tilde F$ consists of a finite collection $\{F_{\alpha_1},\ldots,F_{\alpha_n}\}$ of crisp α-level regions, the perimeter of $\tilde F$ is defined as

$$perimeter(\tilde F) = \sum_{1\leq i<j\leq n} \sum_{k=1}^{n_{i,j}} |\alpha_i^{\frac{1}{2}} - \alpha_j^{\frac{1}{2}}| \cdot length(A_{ijk})$$

where $length(A_{ijk})$ calculates the length of the kth arc along which the discontinuity between the α-level regions with membership degrees α_i and α_j occurs and where this length is weighted by the absolute difference of α_i and α_j.

4.1.3 Metric Operations on Fuzzy Lines

For fuzzy lines we consider operations γ with the signature $\gamma : fline \to real$. Let $\tilde L = \{\tilde l_1,\ldots,\tilde l_n\}$ be a fuzzy line. We start with the operation *length* measuring the size of $\tilde L$ and first determine the length of a simple fuzzy line $\tilde l$. We know that the membership function of $\tilde l$ is given as $\mu_{\tilde l} : f_{\tilde l} \to [0,1]$ with $f_{\tilde l} : [0,1] \to \mathbb{R}^2$. Function $f_{\tilde l}$ just yields the support of $\tilde l$, i.e., $supp(\tilde l) = f_{\tilde l}([0,1])$. Hence

$$length(\tilde l) = \iint_{(x,y)\in supp(\tilde l)} \sqrt{\left(\frac{\partial}{\partial x}\mu_{\tilde F}^{\frac{1}{2}}\right)^2 + \left(\frac{\partial}{\partial y}\mu_{\tilde F}^{\frac{1}{2}}\right)^2}\,dx\,dy$$

Consequently, we obtain

$$length(\tilde L) = \sum_{i=1}^{n} length(\tilde l_i)$$

Another operation is *strength* which follows the principle that "a line is as strong as its weakest link". It computes the minimum membership value of a fuzzy line and is thus defined as

$$strength(\tilde L) = \min_{(x,y)\in supp(\tilde L)} \mu_{\tilde L}((x,y))$$

4.2 Fuzzy-Valued Metric Operations

Another interpretation of metric operations on fuzzy spatial objects is that they yield a fuzzy numerical value as a result. This accords with the fuzzy character of the operand objects. We first explain the concept of fuzzy numbers needed for a description of the metric operations afterwards.

4.2.1 Fuzzy Numbers

The concept of a *fuzzy number* arises from the fact that many quantifiable phenomena do not lend themselves to a characterisation in terms of absolutely precise numbers. For instance, frequently our watches are somewhat inaccurate, so we might say that the time is now "around five o'clock". Or we might estimate the age of an elder man at "nearly seventy-five years". Hence, a fuzzy number is described in terms of a central value and a linguistic modifier like *nearly*, *around*, or *approximately*. Intuitively, a concept captured by such a linguistic expression is fuzzy, because it includes some number values on either side of its central value. Whereas the central value is fully compatible with this concept, the numbers around the central value are compatible with it to lesser degrees. Such a concept can be captured by a fuzzy number defined on $\mathbb{R}$. Its membership function should assign the degree of 1 to the central value and lower degrees to other numbers reflecting their proximity to the central value according to some rule. The membership function should thus decrease from 1 to 0 on both sides of the central value.

Formally, a *fuzzy number* $\tilde A$ is a convex normalized fuzzy set of the real line $\mathbb{R}$ such that (i) $\exists! x_0 \in \mathbb{R} : \mu_{\tilde A}(x_0) = 1$ (x_0 is called the *central* of $\tilde A$) and (ii) $\mu_{\tilde A}$ is piecewise continuous. The membership function of $\tilde A$ can also be expressed in a more explicit form. Let $a,b,c \in \mathbb{R}$. Then

$$\mu_{\tilde A} = \begin{cases} f(x) & \text{if } x \in [a,b] \\ 1 & \text{if } x = b \\ g(x) & \text{if } x \in [b,c] \\ 0 & \text{if } x < a \text{ and } x > c \end{cases}$$

where $a < b < c$, f is a piecewise continuous function increasing to 1 at point b, and g is a piecewise continuous function decreasing from 1 at point b. We introduce *freal* as the type of all fuzzy (real) numbers.

For the representation of the fuzzy-valued result of metric operations we introduce two restricted kinds of fuzzy numbers. The first kind contains numbers that are characterized by the property that function f is lacking, i.e., these numbers only have a *right-sided* membership function. The second kind comprises numbers that are characterized by the property that function g is lacking, i.e., these numbers only have a *left-sided* membership function.

4.2.2 Metric Operations on Fuzzy Regions

For fuzzy regions we now consider fuzzy-valued operations γ with the signature $\gamma : fregion \rightarrow freal$. We first confine ourselves to a subset of operations $\gamma \in \{area, perimeter, height, width, diameter\}$, because their result can be expressed in a generic way by a restricted fuzzy number with a right-sided membership function, as we will see.

Let $\tilde{F} \in fregion$. To determine the result of γ we switch to the view of $\tilde{F}$ as a collection of *crisp* α-level regions $\{F_{\alpha_1}, \dots, F_{\alpha_{n+1}}\}$ where n is possibly infinite. Since the regions F_{α_i} are crisp, we can apply the corresponding known *crisp* operations γ_c to them. The relationship between γ_c and the membership values α_i is given in the following lemma:

LEMMA 5. $\alpha > \beta \Leftrightarrow \gamma_c(F_\alpha) \leq \gamma_c(F_\beta)$.

PROOF. From the definition of a fuzzy region as a collection of α-level regions we know that $\alpha > \beta \Leftrightarrow F_\alpha \subseteq F_\beta$. Since γ_c is a monotonically increasing function, we obtain $F_\alpha \subseteq F_\beta \Leftrightarrow \gamma_c(F_\alpha) \leq \gamma_c(F_\beta)$. $\square$

We now define $\gamma(\tilde{F})$ as the following fuzzy number:

$$\gamma(\tilde{F}) = \{(\gamma_c(F_{\alpha_i}), \alpha_i) \mid \alpha_i \in \Lambda_{\tilde{F}}\}$$

This fuzzy number has a right-sided membership function, because for the smallest α-level region F_{α_1} the membership value is $\alpha_1 = 1$ and for all other α-level regions $F_{\alpha_i}, i > 1$, with increasing i and thus increasing $\gamma_c(F_{\alpha_i})$ the membership value α_i decreases from 1 to 0. In particular, the support of $\gamma(\tilde{F})$ does not contain any smaller values than $\gamma_c(F_{\alpha_1})$. If $\Lambda_{\tilde{F}}$ is finite, we obtain a stepwise constant and hence piecewise continuous membership function for $\gamma(\tilde{F})$. Otherwise, if $\mu(\tilde{F})$ is continuous, $\Lambda_{\tilde{F}}$ is infinite, and we get a continuous membership function for $\gamma(\tilde{F})$.

Intuitively, this result documents the vagueness of the operator $\gamma(\tilde{F})$, because with the increase of $\gamma_c(F_{\alpha_i})$, the certainty and knowledge about its correctness decreases. We can confirm with the membership value 1 that $\gamma_c(F_{\alpha_1})$ is the value of $\gamma(\tilde{F})$. Thus, $\gamma_c(F_{\alpha_1})$ is the lower bound (or the core). But the value could be higher, and if so, we can only confirm it with a lower membership value. The membership value here indicates the degree of imprecision of the operation γ.

Let now $\gamma \in \{elongatedness, roundness\}$. These two operations cannot be treated as fuzzy numbers with right-sided membership functions since they are not monotonically increasing, i.e., in general $F_\alpha \subseteq F_\beta \not\Rightarrow \gamma_c(F_\alpha) \leq \gamma_c(F_\beta)$ and $\gamma_c(F_\alpha) \leq \gamma_c(F_\beta) \not\Rightarrow F_\alpha \subseteq F_\beta$. How to measure these two operations as fuzzy-valued numbers is currently an open issue. It is even doubtful whether they can be represented as general fuzzy numbers at all.

4.2.3 Metric Operations on Fuzzy Lines

Each operation $\gamma \in \{length, strength\}$ on fuzzy lines has the signature $\gamma : fline \rightarrow freal$. Let $\tilde{L}$ be a fuzzy line, and let $\{L_{\alpha_1}, \dots, L_{\alpha_{n+1}}\}$ be a collection of crisp α-cuts where n is possibly infinite. For both operations we pursue a similar strategy as for the operations on fuzzy regions. The main difference in the definition of both operations is that *length* is an increasing function whereas *strength* is a decreasing function. Hence, analogously to Lemma 5 we can conclude that

$$\alpha > \beta \Leftrightarrow length_c(L_\alpha) \leq length_c(L_\beta)$$

But since $L_\alpha \subseteq L_\beta \Leftrightarrow strength_c(L_\alpha) \geq strength_c(L_\beta)$, we obtain

$$\alpha > \beta \Leftrightarrow strength_c(L_\alpha) \geq strength_c(L_\beta)$$

Nevertheless, we can define the value of $\gamma(\tilde{L})$ in the same manner for both operations as the fuzzy number

$$\gamma(\tilde{L}) = \{(\gamma_c(L_{\alpha_i}), \alpha_i) \mid \alpha_i \in \Lambda_{\tilde{L}}\}$$

For *length* this leads to a fuzzy number with a right-sided membership function, and for *strength* we obtain a fuzzy number with a left-sided membership function.

5. CONCLUSIONS AND FUTURE WORK

As part of an abstract model for fuzzy spatial objects in the Euclidean space, we have defined metric operations that can have either real-valued or fuzzy-valued results. All operations considered have been unary functions. For future work we will also have to consider binary metric operations, and here we will have, in particular, to deal with fuzzy distance and fuzzy direction operations. Having achieved a formal and rather complete data model for fuzzy spatial objects, we can transform it to a discrete model where we have to think about finite representations for the objects and algorithms for the operations. These are efforts that should later lead to an efficient implementation.

6. REFERENCES

[1] D. Altman. Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis. *Int. Journal of Geographical Information Systems*, 8(3):271–289, 1994.

[2] P. A. Burrough. *Natural Objects with Indeterminate Boundaries*, pages 3–28. In Burrough and Frank [3], 1996.

[3] P. A. Burrough and A. U. Frank, editors. *Geographic Objects with Indeterminate Boundaries*. GISDATA Series, vol. 2. Taylor & Francis, 1996.

[4] S. Dutta. Qualitative Spatial Reasoning: A Semi-Quantitative Approach Using Fuzzy Logic. In *1st Int. Symp. on the Design and Implementation of Large Spatial Databases (SSD)*, LNCS 409, pages 345–364. Springer-Verlag, 1989.

[5] M. Erwig and M. Schneider. Vague Regions. In *5th Int. Symp. on Advances in Spatial Databases*, LNCS 1262, pages 298–320. Springer-Verlag, 1997.

[6] R. H. Güting and M. Schneider. Realm-Based Spatial Data Types: The ROSE Algebra. *VLDB Journal*, 4:100–143, 1995.

[7] M. Schneider. *Spatial Data Types for Database Systems - Finite Resolution Geometry for Geographic Information Systems*, volume LNCS 1288. Springer-Verlag, Berlin Heidelberg, 1997.

[8] M. Schneider. Uncertainty Management for Spatial Data in Databases: Fuzzy Spatial Data Types. In *6th Int. Symp. on Advances in Spatial Databases*, LNCS 1651, pages 330–351. Springer-Verlag, 1999.

[9] M. Stonebraker. Inclusion of New Types in Relational Database Systems. In *Int. Conf. on Data Engineering*, pages 262–269, 1986.

[10] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.