Computational Geometry 43 (2010) 721–748

Contents lists available at ScienceDirect

Computational Geometry: Theory and Applications

www.elsevier.com/locate/comgeo

Extension of one-dimensional proximity regions to higher dimensions

Elvan Ceyhan *

Department of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 11 March 2009
Accepted 25 May 2010
Available online 27 May 2010
Communicated by R. Fleischer

Keywords:
Class cover catch digraph (CCCD)
Delaunay triangulation
Domination number
Proximity map
Proximity catch digraph
Random graph
Relative density
Triangle center

ABSTRACT

Proximity regions (and maps) are defined based on the relative allocation of points from two or more classes in an area of interest and are used to construct random graphs called proximity catch digraphs (PCDs) which have applications in various fields. The simplest of such maps is the spherical proximity map which gave rise to class cover catch digraph (CCCD) and was applied to pattern classification. In this article, we note some appealing properties of the spherical proximity map in compact intervals on the real line, thereby introduce the mechanism and guidelines for defining new proximity maps in higher dimensions. For non-spherical PCDs, Delaunay tessellation (triangulation in the real plane) is used to partition the region of interest in higher dimensions. We also introduce the auxiliary tools used for the construction of the new proximity maps, as well as some related concepts that will be used in the investigation and comparison of these maps and the resulting PCDs. We provide the distribution of graph invariants, namely, domination number and relative density, of the PCDs and characterize the geometry invariance of the distribution of these graph invariants for uniform data and provide some newly defined proximity maps in higher dimensions as illustrative examples.

1. Introduction

Classification and clustering have received considerable attention in the statistics and probability literature. In recent years, a new classification approach which is based on the proximity maps that incorporate the relative positions of data points from various classes has been developed. Proximity maps and the associated (di)graphs are used in disciplines where shape and structure are crucial. Examples include computer vision (dot patterns), image analysis, pattern recognition (prototype selection), geography and cartography, visual perception, biology, etc. *Proximity graphs* were first introduced by Toussaint [25], who called them *relative neighborhood graphs*. From a mathematical and algorithmic point of view, proximity graphs fall under the category of *computational geometry*.

A general definition of proximity graphs is as follows: Let V be any finite or infinite set of points in $\mathbb{R}^d$ where d is a positive integer. Each (unordered) pair of points $p, q \in V$ is associated with a neighborhood $\mathfrak{N}(p, q) \subseteq \mathbb{R}^d$. Let $\mathfrak{P}$ be a property defined on $\mathfrak{N} = \{\mathfrak{N}(p, q): p, q \in V\}$. A *proximity* (or *neighborhood*) *graph* $G_{\mathfrak{N},\mathfrak{P}}(V, E)$ defined by the property $\mathfrak{P}$ is a graph with the vertex set V and the edge set E such that $pq \in E$ iff $\mathfrak{N}(p, q)$ satisfies property $\mathfrak{P}$. Examples of the most commonly used proximity graphs are the Delaunay tessellation, the boundary of the convex hull, the Gabriel graph, the relative neighborhood graph, Euclidean minimum spanning tree, and the sphere of influence graph of a finite data set. For example, the *relative neighborhood graph* of V, denoted $RNG(V)$, is a prominent representative of the family of graphs which are defined by some sort of neighborliness. The graph $RNG(V)$ has vertex set V and edge set which are exactly the pairs pq of points for which $p \neq q$ and $d(p, q) \leqslant \min_{v \in V} \max(d(p, v), d(q, v))$. That is, pq is an edge of $RNG(V)$ iff $\text{Lune}(p, q)$ does

* Tel.: +90 (212) 338 1845, fax: +90 (212) 338 1559.
E-mail address: elceyhan@ku.edu.tr.

0925-7721/$ – see front matter
doi:10.1016/j.comgeo.2010.05.002

not contain any other points of V, where $\text{Lune}(p,q)$ is defined as the intersection of two balls centered at p and q each with radius $d(p,q)$. See, e.g., [14] and [2] for more detail.

A *digraph* is a directed graph, i.e., a graph with directed edges from one vertex to another based on a binary relation. Then the arc from vertex p to vertex q is denoted by the ordered pair $(p,q) \in V \times V$. For example, the nearest neighbor (di)graph in [20] is a proximity digraph. The *nearest neighbor digraph*, denoted as $NND(V)$, has the vertex set V and (p,q) is an arc iff q is a nearest neighbor of p. Note that if (p,q) is an arc in $NND(V)$, then pq is an edge in $RNG(V)$.

The *proximity catch digraphs* (PCDs) are based on the property $\mathfrak{P}$ that is determined by the following mapping which is defined in a more general space than $\mathbb{R}^d$. Let $(\Omega, \mathcal{M})$ be a measurable space. The *proximity map* $N(\cdot)$ is given by $N : \Omega \to \wp(\Omega)$, where $\wp(\Omega)$ is the power set of Ω, and the *proximity region* of $x \in \Omega$, denoted as $N(x)$, is the image of $x \in \Omega$ under $N(\cdot)$. The points in $N(x)$ are thought of as being "closer" to $x \in \Omega$ than are the points in $\Omega \setminus N(x)$. The PCD D has the vertex set $\mathcal{V} = \{p_1, p_2, \ldots, p_n\}$ and the arc set $\mathcal{A}$ is defined as $(p_i, p_j) \in \mathcal{A}$ iff $p_j \in N(p_i)$ for $i \neq j$. Notice that D depends on the *proximity* map $N(\cdot)$; and if $p_j \in N(p_i)$, then $N(p_i)$ is said to *catch* p_j. Hence the name *proximity catch digraph*. If arcs of the form (p_i, p_i) (i.e., loops) were allowed, D would have been called a *pseudodigraph* according to some authors (see, e.g., [9]).

In a digraph $D = (\mathcal{V}, \mathcal{A})$, a vertex $v \in \mathcal{V}$ *dominates* itself and all vertices of the form $\{u\colon\ (v,u) \in \mathcal{A}\}$. A *dominating set* S_D for the digraph D is a subset of $\mathcal{V}$ such that each vertex $v \in \mathcal{V}$ is dominated by a vertex in S_D. A *minimum dominating set* S_D^* is a dominating set of minimum cardinality and the *domination number* $\gamma(D)$ is defined as $\gamma(D) := |S_D^*|$ where $|\cdot|$ denotes the set cardinality functional. Priebe et al. [21] introduced the class cover catch digraphs (CCCDs) and gave the exact and the asymptotic distribution of the domination number of the CCCD based on two data sets $\mathcal{X}_n$ and $\mathcal{Y}_m$ both of which are random samples from uniform distribution on a compact interval in $\mathbb{R}$. DeVinney et al. [12], Marchette and Priebe [17], Priebe et al. [22,23], and DeVinney and Priebe [11] applied the concept in higher dimensions and demonstrated relatively good performance of CCCDs in classification. The methods employed involve *data reduction* (*condensing*) by using approximate minimum dominating sets as *prototype sets* since finding the exact minimum dominating set in general is an NP-hard problem — in particular, for CCCDs — (see [10]).

Although intuitively appealing and easy to extend to higher dimensions, the exact and the asymptotic distribution of the domination number of the CCCDs are not analytically tractable in multiple dimensions. As alternatives to CCCD, Ceyhan and Priebe [4] introduced an (unparametrized) type of PCDs called *central similarity PCDs*; Ceyhan and Priebe [5] also introduced a parametrized family of PCDs called *proportional-edge PCDs* and used the domination number of this PCD with a fixed parameter for testing spatial interaction patterns of segregation and association between two classes of points. *Segregation* is the pattern in which points of one class tend to cluster together, i.e., form one-class clumps. On the other hand, *association* is the pattern in which the points of one class tend to occur more frequently around points from the other class. The relative (arc) density of the proportional-edge PCDs is also used for testing the spatial patterns in [8]. The parametrized version of the central similarity PCDs is introduced and used for the same purpose in [7]. Relative density of a digraph is the ratio of the number of arcs in the given digraph to the number of arcs in a complete digraph with the same order as the given digraph. These new families (other than CCCDs) are applicable to pattern classification also and are designed to have better distributional and mathematical properties. Ceyhan and Priebe [6] derived the asymptotic distribution of the domination number of proportional-edge PCDs for uniform data for the entire range of the parameter. Ceyhan [3] discusses the use of the domination number of proportional-edge PCDs, whose asymptotic distribution was computed in [6] for testing spatial patterns. Domination number (relative density) approach is applicable for testing spatial patterns, since under segregation the domination number tends to be small (large), and under association it tends to be large (small). An extensive treatment of the PCDs based on Delaunay tessellations is available in [1].

In this article, we determine some appealing properties of the spherical proximity region (i.e., the proximity region associated with CCCD) for uniform data in $\mathbb{R}$ and use them in introducing the mechanism and guidelines for defining new proximity maps in higher dimensions. As CCCD behaves "nicely" for uniform data in $\mathbb{R}$ (in the sense that the exact and asymptotic distributions of the domination number are available [21]), by emulating its properties in higher dimensions, we expect the new PCDs will behave in a similar fashion. Furthermore, we introduce some auxiliary tools used for the construction of the new proximity maps, as well as some related concepts that will be used in the investigation and comparison of the proximity maps. Additionally, we discuss the conditions for the geometry invariance for uniform data in triangles. Although the proportional-edge and central similarity PCDs were introduced before, their comparison and assessment in terms of the appealing properties were not done. We also introduce two new PCD families.

Throughout the article, $d(x,y)$, can be any distance in $\mathbb{R}^n$. Furthermore, the distance between a point x and a set A is defined as $d(x,A) := \inf_{y \in A} d(x,y)$; and the distance between two sets A and B is defined as $d(A,B) := \inf_{(x,y) \in A \times B} d(x,y)$. We describe the data-random PCDs and the related concepts such as Voronoi diagram and Delaunay tessellation, provide the appealing properties of spherical proximity maps in $\mathbb{R}$, transformations preserving uniformity on triangles in $\mathbb{R}^2$, and vertex and edge regions in Section 2. We present the proximity regions in Delaunay tessellations in Section 3, the results on relative arc density and the domination number of the PCDs in Section 4, introduce two new proximity maps in Section 5, and discussion and conclusions in Section 6. We also provide a list of the new concepts and their notation in Appendix A for quick reference.

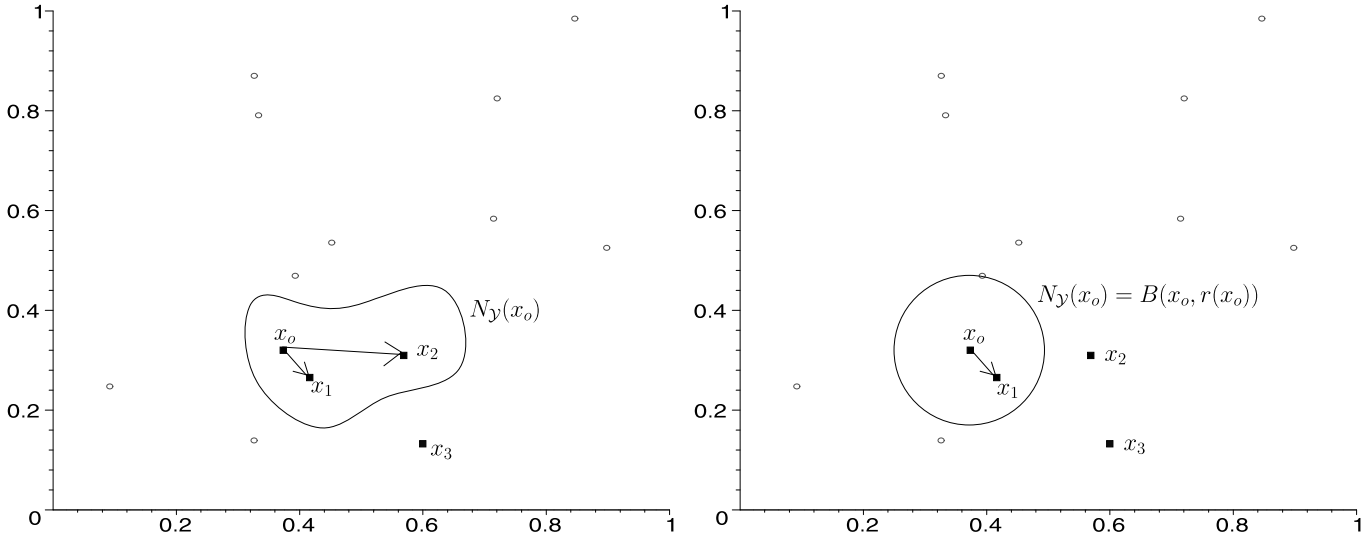

Fig. 1. Plotted are the class $\mathcal{Y}$ points (circles) and class $\mathcal{X}$ points $\{x_o, x_1, x_2, x_3\}$ (solid squares). The proximity region $N_{\mathcal{Y}}(x_o)$ in general form for an x_o (left) and spherical proximity region (right) and the corresponding arcs from x_o to other points in $\{x_o, x_1, x_2, x_3\}$.

2. Data-random proximity catch digraphs and related concepts

In particular, the proximity regions are constructed using data sets from two classes, $\mathcal{X}$ and $\mathcal{Y}$. Given $\mathcal{Y}_m \subseteq \Omega$ from class $\mathcal{Y}$, the *proximity map* $N_{\mathcal{Y}}(\cdot) : \Omega \to \wp(\Omega)$ associates a *proximity region* $N_{\mathcal{Y}}(x) \subseteq \Omega$ with each point $x \in \Omega$. The region $N_{\mathcal{Y}}(x)$ is defined in terms of the relative position of x with respect to points from $\mathcal{Y}_m$. Two examples of $N_{\mathcal{Y}}(x)$ are presented in Fig. 1, where the corresponding proximity region in the left has a general shape, while in the right it is spherical. If $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ is a set of Ω-valued random variables from class $\mathcal{X}$, then $N_{\mathcal{Y}}(X_i)$ are random sets. If X_i are independent identically distributed (iid) and if different X_i induce different $N_{\mathcal{Y}}(X_i)$ a.s., then $N_{\mathcal{Y}}(X_i)$ are iid as well. The data-random PCD D — associated with $N_{\mathcal{Y}}(\cdot)$ — is defined with vertex set $\mathcal{X}_n$ and arc set $\mathcal{A}$ by $(X_i, X_j) \in \mathcal{A}$ iff $X_j \in N_{\mathcal{Y}}(X_i)$. See Fig. 1. Since this relationship is not symmetric, a digraph is used rather than a graph. The random digraph D depends on the (joint) distribution of the X_i and on the map $N_{\mathcal{Y}}(\cdot)$. Let $p_a(N_{\mathcal{Y}}) := P((X_i, X_j) \in \mathcal{A}) = P(X_j \in N_{\mathcal{Y}}(X_i))$; so $p_a(N_{\mathcal{Y}})$ is the probability of having an arc from X_i to X_j, hence is called *arc probability* for the PCD based on $N_{\mathcal{Y}}$.

The PCDs are closely related to the *proximity graphs* of [14] and might be considered as a special case of *covering sets* of [26] and *intersection digraphs* of [24]. This data-random proximity digraph is a *vertex-random proximity digraph* which is not of standard type (see, e.g., [13]). The randomness of the PCDs lies in the fact that the vertices are random with joint probability density function (pdf) $f_{X,Y}$, but arcs (X_i, X_j) are deterministic functions of the random variable X_j and the set $N_{\mathcal{Y}}(X_i)$. For example, the CCCD of [21] can be viewed as an example of PCD with $N_{\mathcal{Y}}(x)$ being the ball $B(x, r(x))$ centered at x with radius $r(x)$, where $r(x) := \min_{\mathsf{y} \in \mathcal{Y}_m} d(x, \mathsf{y})$. The CCCD is the digraph of order n with vertex set $\mathcal{X}_n$ and an arc from X_i to X_j iff $X_j \in B(X_i, r(X_i))$. See Fig. 1 (right). That is, there is an arc from X_i to X_j iff there exists an open ball centered at X_i which is "pure" (or contains no elements) of $\mathcal{Y}_m$, and simultaneously contains (or "catches") point X_j.

Notice that the CCCDs are defined with (open) balls only, whereas PCDs are not based on a particular geometric shape or a functional form; that is, PCDs admit $N_{\mathcal{Y}}(\cdot)$ to be any type of region, e.g., circle (ball), arc slice, triangle, a convex or nonconvex polygon, etc. In this sense, the PCDs are defined in a more general setting compared to CCCDs.

2.1. The appealing properties of spherical proximity regions in $\mathbb{R}$

Let $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_m\} \subset \mathbb{R}$ and $Y_{i:m}$ be the ith order statistic (i.e., ith smallest data point in $\mathcal{Y}_m$). Then the proximity map associated with CCCD is defined as the open ball $N_S(x) := B(x, r(x))$ for all $x \in \mathbb{R} \setminus \mathcal{Y}_m$, where $r(x) = \min_{\mathsf{y} \in \mathcal{Y}_m} d(x, \mathsf{y})$ with $d(x, y)$ being the Euclidean distance between x and y [21]. See, e.g., Fig. 2. For $x \in \mathcal{Y}_m$, define $N_S(x) = \{x\}$. Notice that $N_S(x)$ is a sphere for $x \notin \mathcal{Y}_m$ in higher dimensions, hence the name *spherical proximity map* and the notation N_S. Furthermore, dependence on $\mathcal{Y}_m$ is through $r(x)$. Note that, this proximity map is based on the intervals $I_i = (\mathsf{y}_{(i-1):m}, \mathsf{y}_{i:m})$ for $i = 1, 2, \ldots, (m+1)$ with the additional notation that $\mathsf{y}_{0:m} = -\infty$ and $\mathsf{y}_{(m+1):m} = \infty$.

The CCCDs in $\mathbb{R}$ have desirable properties such as the finite sample and asymptotic distributions of the domination number being available. In this section, we determine some appealing properties of the proximity map associated with CCCD for uniform data in a compact interval in $\mathbb{R}$ and use these properties as guidelines for defining new proximity maps in higher dimensions. Potentially these properties make the CCCD to behave so "nicely" in $\mathbb{R}$ and the more they are satisfied by the new PCDs in higher dimensions, the more likely the new PCDs to have similar behavior.

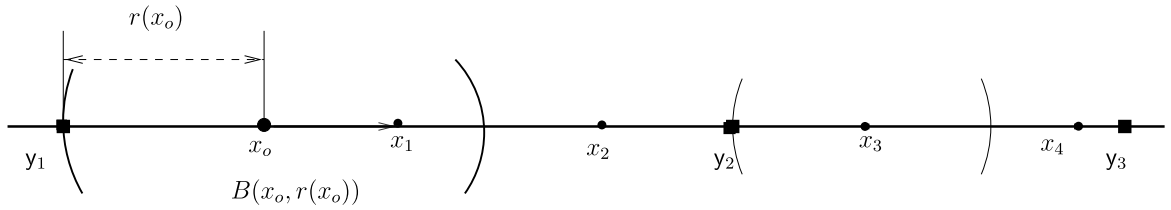

Fig. 2. Depicted are class $\mathcal{Y}$ points $\{y_1, y_2, y_3\}$ (solid squares) and class $\mathcal{X}$ points $\{x_o, x_1, x_2, x_3, x_4\}$ (solid circles) on the real line. The spherical proximity regions for x_o and x_3 and the corresponding arcs are also presented.

For $X_i \overset{iid}{\sim} \mathcal{U}(I_i)$, the uniform distribution on I_i, without loss of generality, we can assume $I_i = (a, b)$ with $a, b \in \mathbb{R}$ and $a < b$. Then the arc probability $p_a(N_S) = P(X_2 \in N_S(X_1)) = 1/2$, since

$$P\big(X_2 \in N_S(X_1)\big) = \int_a^{(a+b)/2} \int_a^{2x_1-a} (b-a)^{-2}\, dx_2\, dx_1 + \int_{(a+b)/2}^{b} \int_{2x_1-b}^{b} (b-a)^{-2}\, dx_2\, dx_1 = 1/2.$$

Observe that the arc probability is independent of a and b, hence the interval I_i.

Below are some appealing properties of the proximity map $N_S(x)$ in $\mathbb{R}$ (for $m > 1$):

P1: The region $N_S(x)$ is well-defined for all $x \in \mathcal{C}_H(\mathcal{Y}_m) = [y_{1:m}, y_{m:m}]$.
P2: For all $x \in \mathcal{C}_H(\mathcal{Y}_m)$, we have $x \in N_S(x)$.
P3: The point x is at the *center* of $N_S(x)$ for all $x \in \mathcal{C}_H(\mathcal{Y}_m)$.
P4: For $x \in I_i \subseteq \mathcal{C}_H(\mathcal{Y}_m)$, $N_S(x)$ and I_i are of the *same type*; i.e., they are both intervals.
P5: For $x \in I_i \subseteq \mathcal{C}_H(\mathcal{Y}_m)$, $N_S(x)$ mimics the shape of I_i; i.e., it is (geometrically) *similar* to I_i.
P6: For $x \in I_i$, $N_S(x)$ is a proper subset of I_i for all $x \in I_i \setminus \{(y_{(i-1):m} + y_{i:m})/2\}$ (or almost everywhere in I_i).
P7: For $x \in I_i$ and $y \in I_j$ with $i \neq j$, $N_S(x)$ and $N_S(y)$ are disjoint regions.
P8: The size (i.e., measure) of $N_S(x)$ is continuous in x; that is, for each $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that $||N_S(y)| - |N_S(x)|| < \varepsilon$ whenever $|d(x, y)| < \delta_\varepsilon$.
P9: The arc probability $p_a(N_S)$ does not depend on the support interval for uniform data in $\mathbb{R}$.

Notice that N_S satisfies the properties **P1**, **P2**, and **P3** for all $x \in \mathbb{R}$ provided that $m \geqslant 1$. Property **P9** implies that not only the arc probability but also the distribution of the relative arc density and domination number do not depend on the support interval either. This independence of the support set is called *geometry invariance* in higher dimensions (see Section 2.3). For N_S in $\mathbb{R}$, it suffices to work with $\mathcal{U}(0, 1)$ data, and in higher dimensions (in $\mathbb{R}^2$) we will be able to consider only uniform data in a standard d-dimensional polytope (an equilateral triangle) for PCDs based on proximity maps that satisfy **P9**.

Suppose we partition the convex hull of $\mathcal{Y}_m$, $\mathcal{C}_H(\mathcal{Y}_m)$ by Delaunay tessellation. Let $\mathcal{T}_i$ be the ith Delaunay cell in the Delaunay tessellation of $\mathcal{Y}_m$ for $i = 1, 2, \ldots, J$. See Fig. 3. Note that **P4** and **P5** are equivalent when $d = 1$ for $x \in \mathcal{C}_H(\mathcal{Y}_m)$, since any two (compact) intervals in $\mathbb{R}$ are (geometrically) similar. For $d > 1$, **P5** implies **P4** only, since, for example, for $d = 2$ a Delaunay tessellation is a triangulation and any two triangles are not necessarily similar, but similar triangles are always of the same type as they are triangles.

Notice that $N_S(\cdot)$ satisfies only **P1**, **P2**, **P3**, and **P8** in $\mathbb{R}^d$ with $d > 1$. Properties **P4** and **P5** fail, since $N_S(x)$ is a sphere for $x \notin \mathcal{Y}_m$, but $\mathcal{T}_i$ is a $(d+1)$-simplex. For any $x \in \mathcal{T}_i \subset \mathbb{R}^d$, $B(x, r(x)) \not\subset \mathcal{T}_i$, so **P6** also fails which also implies that $N_S(x)$ and $N_S(y)$ might overlap for x, y from two distinct cells (see Fig. 5), hence **P7** is violated. The arc probability $p_a(N_S)$ depends on the support set $\mathcal{T}_i$ for $d > 1$, so **P9** is violated.

Let Ω_i, $i \in \{1, 2, \ldots, J\}$ be a partition on Ω and μ be the associated measure on Ω. The appealing properties mentioned above can be extended to more general measurable spaces by replacing the support set with Ω and "size" with measure μ [2].

Property **P6** suggests a new concept. If $N_{\mathcal{Y}}(x)$ is a superset of Ω (i.e., $N_{\mathcal{Y}}(x) \supseteq \Omega$), then the corresponding digraph has domination number equal to 1. So the measure of set of such points is of interest for the domination number of the PCDs.

Definition 2.1. The *superset region* for any proximity map $N(\cdot)$ in Ω is defined to be $\mathcal{R}_S(N) := \{x \in \Omega\colon\ N(x) \supseteq \Omega\}$.

For example, for $\Omega = I_i \subsetneq \mathbb{R}$, $\mathcal{R}_S(N_S) := \{x \in I_i\colon\ N_S(x) = I_i\} = \{(y_{(i-1):m} + y_{i:m})/2\}$, and for $\Omega = \mathcal{T}_i \subsetneq \mathbb{R}^d$, $\mathcal{R}_S(N_S) := \{x \in \mathcal{T}_i\colon\ N_S(x) = \mathcal{T}_i\}$. Note that for $x \in I_i$, $\lambda(N_S(x)) \leqslant \lambda(I_i)$ and $\lambda(N_S(x)) = \lambda(I_i)$ iff $x \in \mathcal{R}_S(N_S)$ where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}$ (also called $\mathbb{R}$-Lebesgue measure). So the proximity region of a point in $\mathcal{R}_S(N_S)$ has the largest $\mathbb{R}$-Lebesgue measure. Note also that $\mathcal{R}_S(N_S)$ is not a random set, but $\mathbf{I}(X \in \mathcal{R}_S(N_S))$ is a random variable. The property **P6** is equivalent to $\mathcal{R}_S(N_S)$ having zero $\mathbb{R}$-Lebesgue measure. Moreover, the larger the superset region, the more likely the relative density to be larger. On the other hand, for $x \in \partial(I_i) = \{y_{(i-1):m}, y_i\}$, the proximity region $N_S(x) = \{x\}$ which has zero $\mathbb{R}$-Lebesgue measure. When $N_{\mathcal{Y}}(x)$ has zero measure, there is no arc from x to other points a.s. This suggests the following concept.

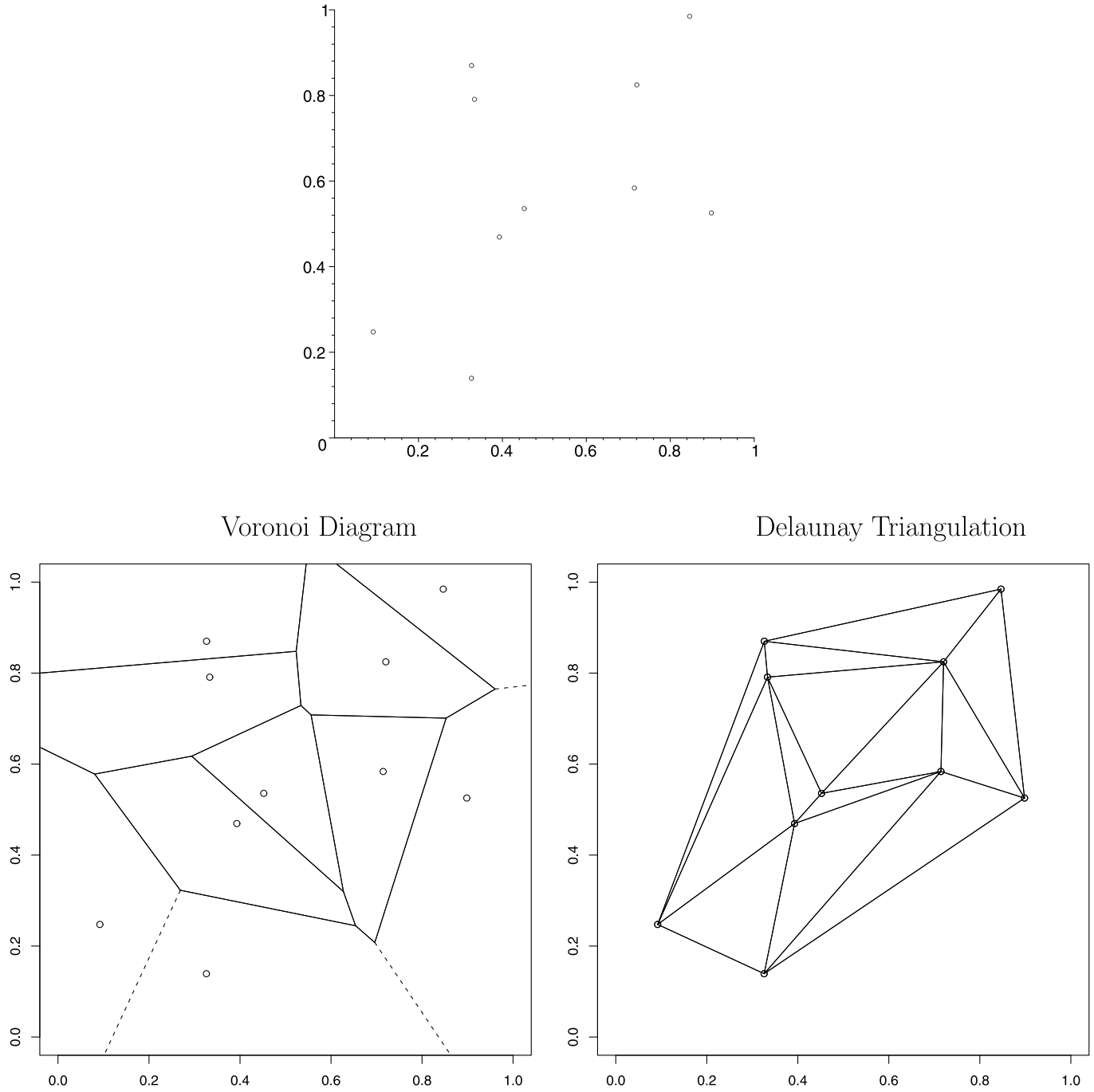

Fig. 3. Depicted are 10 points generated iid $\mathcal{U}(0,1)\times(0,1)$ from class $\mathcal{Y}$ (top), the corresponding Voronoi diagram (bottom left) and the Delaunay triangulation (bottom right).

Definition 2.2. Let (Ω,μ) be a measurable space. The Λ_0-*region* for any proximity map $N(\cdot)$ is defined to be $\Lambda_0(N):=\{x\in\Omega:\ \mu(N(x))=0\}$.

For $\Omega=\mathbb{R}^d$, $\Lambda_0(N_S):=\{x\in\mathbb{R}^d:\ \lambda(N_S(x))=0\}=\mathcal{Y}_m$. The notation Λ_0-region is suggested by the fact that $N_S(x)$ has zero Lebesgue measure for $x\in\Lambda_0(N_S)$.

Given a set $\mathcal{X}_n$ of size n in $[y_{1:m},y_{m:m}]\setminus\mathcal{Y}_m$, **P7** implies that the number of disconnected components in the PCD based on $N_S(\cdot)$ is at least the cardinality of $\{i\in[m+1]:\ \mathcal{X}_n\cap I_i\neq\emptyset\}$, which is the set of indices of the intervals that contain some point(s) from $\mathcal{X}_n$ where $[m]:=\{0,1,2,\ldots,m-1\}$.

The proximity region $N_S(x)$ can easily be extended to multiple dimensions and is well-defined for all $x\in\mathbb{R}^d$ provided that $m\geqslant 1$. However, for $d>1$, finding the minimum dominating set of the PCD associated with $N_S(\cdot)$ is an NP-hard problem [10] and the distribution of the domination number is not analytically tractable [1]. This drawback has motivated the definition of new families of proximity maps in higher dimensions. Note that for $d=1$, such problems do not exist.

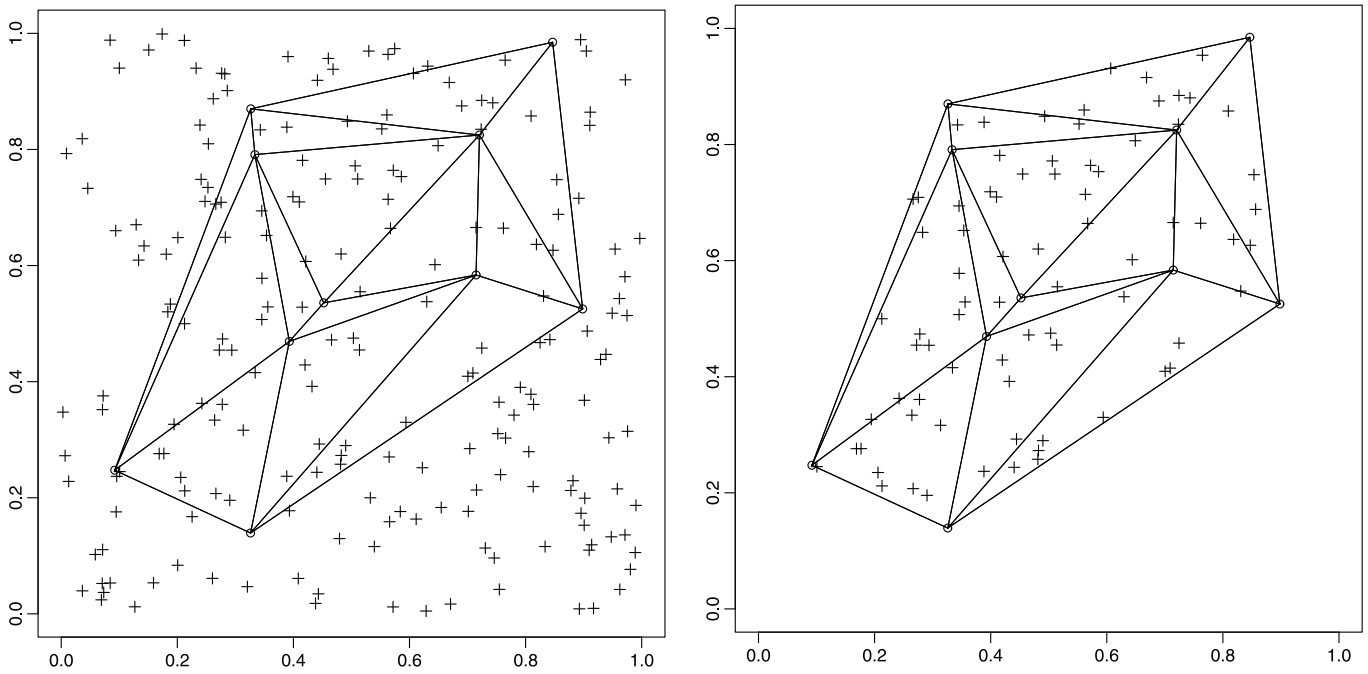

Fig. 4. A realization of 200 points from class $\mathcal{X}$ (pluses) and the Delaunay triangulation based on 10 points from class $\mathcal{Y}$ (circles) in Fig. 3 (left). Out of these 200 class $\mathcal{X}$ points, only 77 are in the convex hull of class $\mathcal{Y}$ points (right).

2.2. Voronoi diagrams and Delaunay tessellations

Our next goal is to extend the PCD concept to higher dimensions and investigate the properties of the associated PCDs. The spherical proximity map in $\mathbb{R}$ is based on the intervals $I_i = (Y_{(i-1):m}, Y_{i:m})$ for $i = 1, \ldots, (m+1)$. This *intervalization* can be viewed as a tessellation, since it partitions $\mathcal{C}_H(\mathcal{Y}_m)$. For $d > 1$, a natural tessellation that partitions $\mathcal{C}_H(\mathcal{Y}_m)$ is the Delaunay tessellation, where each Delaunay cell is a $(d+1)$-simplex. In $\mathbb{R}$, the cell that contains x is implicitly used to define the spherical proximity map. More specifically, our proximity maps will be based on the relative position of points from class $\mathcal{X}$ with respect to the Delaunay tessellation of the points from class $\mathcal{Y}$. See [19] for more on Delaunay tessellations.

By definition a Delaunay tessellation of a finite set of points, P, is the dual of the *Voronoi diagram* based on the same set. The tessellation yields a (unique) polytopization provided that no more than $(d+1)$ points in $\mathbb{R}^d$ are cospherical (i.e., no more than $(d+1)$ points lie on the boundary of a (hyper)sphere in $\mathbb{R}^d$). Moreover, the circumsphere of each Delaunay polytope (i.e., the sphere that contains the vertices of the Delaunay polytope on its boundary) is pure from the set P; i.e., the interior of the circumsphere of the Delaunay polytope does not contain any points from P. The Delaunay tessellation partitions $\mathcal{C}_H(P)$. In particular, in $\mathbb{R}^2$, the tessellation is a *triangulation* that yields triangles T_i, $i = 1, 2, \ldots, J$ (see, e.g., [19]) provided that no more than three points are cocircular (i.e., no more than three points lie on the boundary of some circle in $\mathbb{R}^2$). See Fig. 3 for an example with $m = 10$ class $\mathcal{Y}$ points iid from $\mathcal{U}((0,1)\times(0,1))$. In this article we adopt the convention that a triangle refers to the closed region bounded by its edges.

The Delaunay triangles are based on a given set of points $\mathcal{Y}_m$. The set $\mathcal{Y}_m$ can be assumed to come from a Poisson point process on a finite region to remove the conditioning on $\mathcal{Y}_m$ in the application of PCDs [2]. We call the Delaunay tessellation based on a finite data set from a Poisson point process *Poisson Delaunay tessellation* and denote it $\mathcal{D}_P$. The associated Voronoi diagram is called the *Poisson Voronoi diagram* and is denoted by $\mathcal{V}_P$. For more detail on the properties of $\mathcal{V}_P$, see [19].

Let c and r be the circumcenter and circumradius, respectively, of a $(d+1)$-dimensional Poisson Delaunay cell in $\mathbb{R}^d$. Then the $(d+1)$ vertices of the cell are the points $\{c + ru_i\}$ where $\{u_i\}$ are the unit vectors for $i = 0, 1, 2, \ldots, d$. The ergodic joint pdf of $\mathcal{D}_P$, the pdf of r, and kth moment of the area of a typical Poisson Delaunay cell are provided in [19]. The pdf of the minimum angle and the pdf of the maximum angle, and the distribution of the length of an arbitrary edge of an arbitrary triangle from $\mathcal{D}_P$ are also provided in [19] with relevant references.

The non-spherical PCDs (i.e., PCDs other than CCCDs) we will consider in this article, will be defined only for $\mathcal{X}$ points inside the convex hull of $\mathcal{Y}$ points, $\mathcal{C}_H(\mathcal{Y}_m)$, while the CCCDs are well-defined for all $\mathcal{X}$ points provided $m \geqslant 1$. See Fig. 4 for an example with $n = 200$ class $\mathcal{X}$ points $\overset{iid}{\sim} \mathcal{U}((0,1)\times(0,1))$, and the Delaunay triangulation based on the 10 class $\mathcal{Y}$ points in Fig. 3. This is actually the main advantage of the CCCDs for multi-dimensional data; but CCCDs suffer from the mathematical intractability of the calculation of the graph invariants of interest. On the other hand, the non-spherical PCDs, although restricted to $\mathcal{C}_H(\mathcal{Y}_m)$, have more tractable graph invariants, which is their main advantage. Moreover, these non-spherical PCDs might allow one to work in one Delaunay cell (triangle in $\mathbb{R}^2$) only, if the corresponding proximity regions satisfy properties **P7** and **P9**. The distribution of the graph invariants for the points outside the convex hull should be somehow corrected. For example, Ceyhan [3] suggests a correction coefficient to adjust the domination number of the proportional-

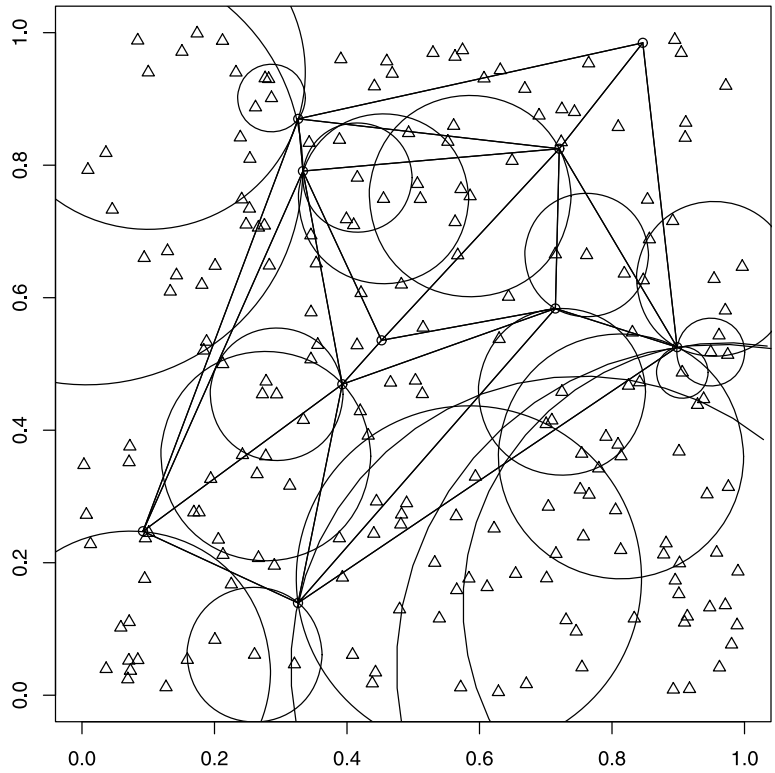

Fig. 5. The spherical proximity regions for 20 of the 200 class $\mathcal{X}$ points (small triangles) depicted in Fig. 4 (left).

edge PCDs. Such a correction can be derived for the relative density as well. The CCCDs were used for classification purposes in [22], and the non-spherical PCDs can be used for the same purpose in a similar fashion. However, since non-spherical PCDs we will consider are only defined for $\mathcal{X}$ points inside the convex hull, a different strategy is required for points outside the convex hull. For example, one can use the PCD approach for the points inside, and employ some other method such as nearest neighbor classification for the points outside, hence the resultant classifier will be kind of a hybrid classifier.

The calculations in the convex hull may also be used to approximate the results for more general proximity regions. Given any proximity region $N_{\mathcal{Y}}(x)$ which is based on the relative position of x points with respect to class $\mathcal{Y}$ points, if we restrict the proximity region $N_{\mathcal{Y}}(x)$ to the Delaunay cell that contains x, denoted $N'_{\mathcal{Y}}(x)$, then we will have a region satisfying **P7** and so will need to do the calculations for Delaunay cells only. Moreover, the PCD based on $N'_{\mathcal{Y}}(x)$ will be a subdigraph of the PCD based on $N_{\mathcal{Y}}(x)$. Hence the domination number (relative density) of the PCD based on $N_{\mathcal{Y}}(x)$ will be stochastically smaller (larger) than the PCD based on $N'_{\mathcal{Y}}(x)$. So the calculations for the PCD based on $N'_{\mathcal{Y}}(x)$ will be informative about the general PCD we start with.

2.3. Transformations preserving uniformity on triangles in $\mathbb{R}^2$

The property **P9**, when satisfied by a proximity region, suggests that in higher dimensions the arc probability of the corresponding PCDs based on uniform data would be geometry invariant, i.e., would not depend on the geometry of the support set. The set $\mathcal{X}_n$ is assumed to be a set of iid uniform random variables on the convex hull of $\mathcal{Y}_m$; i.e., a random sample from $\mathcal{U}(\mathcal{C}_H(\mathcal{Y}_m))$. In particular, in $\mathbb{R}^2$, conditional on $|\mathcal{X}_n \cap T_i| > 0$ being fixed, $\mathcal{X}_n \cap T_i$ will also be a set of iid uniform random variables on T_i for $i \in \{1, 2, \ldots, J\}$, where T_i is the ith Delaunay triangle and J is the total number of Delaunay triangles. The geometry invariance property will reduce the triangle T_i as much as possible while preserving uniformity and the probabilities related to PCDs will simplify in notation and calculations. Below, we present such a transformation that reduces a single triangle to the standard equilateral triangle $T_e = T((0,0),(1,0),(1/2,\sqrt{3}/2))$.

Let $\mathcal{Y}_3 = \{\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3\} \subset \mathbb{R}^2$ be three non-collinear points and $T(\mathcal{Y}_3)$ be the triangle with vertices $\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3$. Let $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ for $i = 1, 2, \ldots, n$. The pdf of $\mathcal{U}(T(\mathcal{Y}_3))$ is $f(u) = \frac{1}{A(T(\mathcal{Y}_3))}\mathbf{I}(u \in T(\mathcal{Y}_3))$, where $A(\cdot)$ is the area functional.

The triangle $T(\mathcal{Y}_3)$ can be carried into the first quadrant by a composition of transformations in such a way that the largest edge has unit length and lies on the x-axis, and the x-coordinate of the vertex nonadjacent to largest edge is less than $1/2$. We call the resultant triangle the *basic triangle* and denote it as T_b where $T_b = T((0,0),(1,0),(c_1,c_2))$ with $0 < c_1 \leqslant 1/2$, and $c_2 > 0$ and $(1-c_1)^2 + c_2^2 \leqslant 1$. See Fig. 7 (left). We will describe such transformations below: Let e_i be the edge opposite to the vertex y_i for $i \in \{1, 2, 3\}$. Find the lengths of the edges; say e_3 is of maximum length. Then scale the triangle so that e_3 is of unit length. Next translate y_1 to $(0,0)$, and if necessary rotate the triangle so that $\mathrm{y}_2 = (1,0)$. If the y-coordinate of y_3 is negative reflect the triangle around the x-axis, then if the x-coordinate of y_3 is greater than $1/2$, reflect it around $x = 1/2$, then the associated basic triangle T_b is obtained by a transformation denoted by ϕ_b which is a

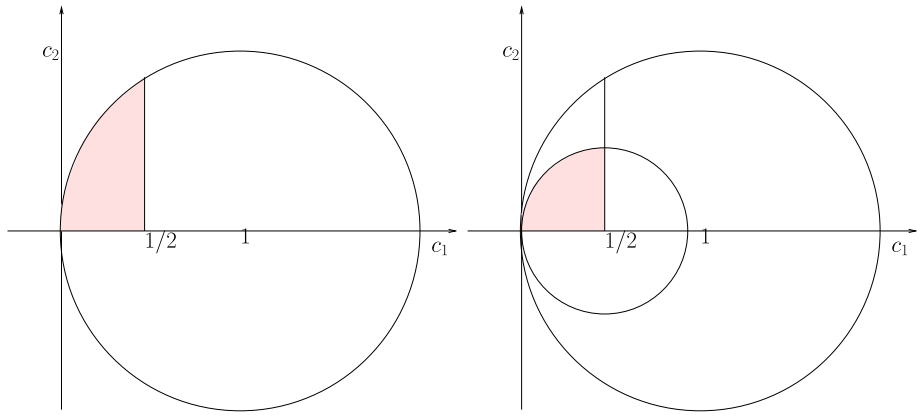

Fig. 6. The shaded regions are the domain of (c_1, c_2) values for the basic triangle T_b (left) and part of the domain where T_b is a non-acute triangle (right).

composition of some of the rigid motion transformations (namely translation, rotation, and reflection) and scaling. Hence if $T(\mathcal{Y}_3)$ is transformed into T_b, then $T(\mathcal{Y}_3)$ is similar to T_b and $\phi_b(T(\mathcal{Y}_3)) = T_b$. Thus the random variables $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ transformed along with $T(\mathcal{Y}_3)$ in the described fashion by ϕ_b satisfy $\phi_b(X_i) \overset{iid}{\sim} \mathcal{U}(T_b)$. So, without loss of generality, we can assume $T(\mathcal{Y}_3)$ to be the basic triangle. If $c_1 = 1/2$ and $c_2 = \sqrt{3}/2$, then T_b is an equilateral triangle; if $c_2 < \sqrt{c_1 - c_1^2}$, then T_b is an obtuse triangle; if $c_2 = \sqrt{c_1 - c_1^2}$, then T_b is a right triangle; and if $c_2 > \sqrt{c_1 - c_1^2}$, then T_b is an acute triangle. If $c_2 = 0$, then the T_b reduces to the unit interval $(0, 1)$. See Fig. 6 for the domain of (c_1, c_2) for T_b and the part of the domain on which T_b is a non-acute triangle.

Lemma 2.3. *The arc probability $p_a(N_{\mathcal{Y}})$ of the PCD based on $N_{\mathcal{Y}}$ for uniform data on $T(\mathcal{Y}_3)$ is rigid-motion and scale invariant; i.e., $p_a(N_{\mathcal{Y}})$ does not change under rigid motion transformations and does not depend on the scale of the support triangle $T(\mathcal{Y}_3)$.*

Proof. We have shown that for $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$, it follows that $\phi_b(X_i) \overset{iid}{\sim} \mathcal{U}(T_b)$, since $T(\mathcal{Y}_3)$ is similar to T_b. For uniform data, the set probabilities are calculated as the ratio of the area of the set to the total area. So $P(X \in S \subseteq T(\mathcal{Y}_3)) = A(S)/A(T(\mathcal{Y}_3))$ and $P(\phi_b(X) \in \phi_b(S) \subseteq \phi_b(T(\mathcal{Y}_3))) = P(\phi_b(X) \in \phi_b(S) \subseteq T_b) = A(\phi_b(S))/A(T_b) = [kA(S)]/[kA(T(\mathcal{Y}_3))] = A(S)/A(T(\mathcal{Y}_3))$ where k is the scaling factor. Letting $X = X_j$ and $S = N_{\mathcal{Y}}(X_i)$, the desired result follows. □

Based on Lemma 2.3, without loss of generality, we can assume $T(\mathcal{Y}_3)$ to be the basic triangle T_b for uniform data.

2.3.1. Transformation of T_b to T_e

There are also transformations that preserve uniformity of the random variable, but not similarity of the triangles. We only describe the transformation that maps $T(\mathcal{Y}_3)$ to the standard equilateral triangle, $T_e = T((0,0),(1,0),(1/2,\sqrt{3}/2))$ for exploiting the symmetry in calculations using T_e.

Let $\phi_e : (x, y) \to (u, v)$, where $u(x, y) = x + \frac{1-2c_1}{\sqrt{3}} y$ and $v(x, y) = \frac{\sqrt{3}}{2c_2} y$. Then y_1 is mapped to $(0, 0)$, y_2 is mapped to $(1, 0)$, and y_3 is mapped to $(1/2, \sqrt{3}/2)$. See also Fig. 7. Note that the inverse transformation is $\phi_e^{-1}(u, v) = (x(u, v), y(u, v))$ where $x(u, v) = u - \frac{(1-2c_1)}{\sqrt{3}} v$ and $y(u, v) = \frac{2c_2}{\sqrt{3}} u$. Then the Jacobian is given by

$$J(x, y) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & \frac{2c_1-1}{\sqrt{3}} \\ 0 & \frac{2c_2}{\sqrt{3}} \end{vmatrix} = \frac{2c_2}{\sqrt{3}}.$$

So $f_{U,V}(u, v) = f_{X,Y}(\phi_e^{-1}(u, v))|J| = \frac{4}{\sqrt{3}} \mathbf{I}((u, v) \in T_e)$. Hence uniformity is preserved.

Theorem 2.4. *The arc probability $p_a(N_{\mathcal{Y}})$ of the PCD based on $N_{\mathcal{Y}}$ for uniform data on T_b is geometry invariant iff $A(\phi_e(N_{\mathcal{Y}}(x))) = A(N_{\phi_e(\mathcal{Y})}(\phi_e(x)))$ for all $x \in T_b$ where $N_{\phi_e(\mathcal{Y})}$ is based on $\phi_e(\mathcal{Y}_3)$.*

Proof. By Lemma 2.3, the PCD based on $N_{\mathcal{Y}}$ for uniform data on $T(\mathcal{Y}_3)$ is rigid-motion and scale invariant. So $T(\mathcal{Y}_3)$ can be transformed to T_b preserving the uniformity of the data and the arc probability for the associated PCD. For uniform data, the set probabilities are calculated as the ratio of the area of the set to the total area. Suppose the arc probability is geometry invariant. Then $p_a(N_{\mathcal{Y}}) = P(X \in N_{\mathcal{Y}}(x)) = P(\phi_e(X) \in N_{\phi_e(\mathcal{Y})}(\phi_e(x)))$. But $P(X \in N_{\mathcal{Y}}(x)) = A(N_{\mathcal{Y}}(x))/A(T_b)$ and $P(\phi_e(X) \in N_{\phi_e(\mathcal{Y})}(\phi_e(x))) = A(N_{\phi_e(\mathcal{Y})}(\phi_e(x)))/A(T_e)$. Moreover $A(N_{\mathcal{Y}}(x))/A(T_b) = A(\phi_e(N_{\mathcal{Y}}(x)))/A(\phi_e(T_b)) = A(\phi_e(N_{\mathcal{Y}}(x)))/A(T_e)$ since the Jacobian cancels out and $\phi_e(T_b) = T_e$. Hence $A(N_{\phi_e(\mathcal{Y})}(\phi_e(x)))/A(T_e) = A(\phi_e(N_{\mathcal{Y}}(x)))/A(T_e)$ implies $A(\phi_e(N_{\mathcal{Y}}(x))) = A(N_{\phi_e(\mathcal{Y})}(\phi_e(x)))$ for all $x \in T_b$. The converse can be proved similarly. □

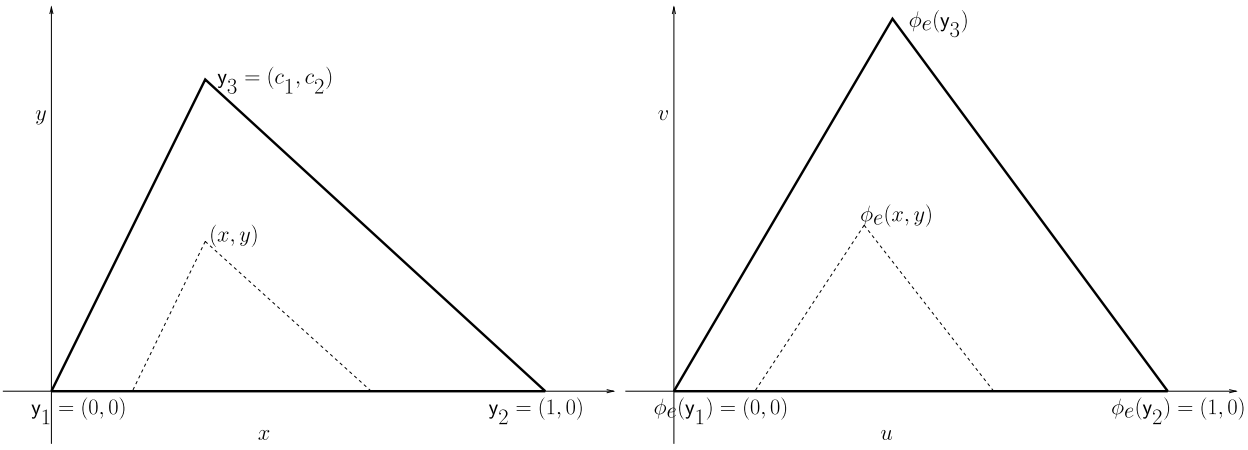

Fig. 7. The description of $\phi_e(x,y)$ for $(x,y)\in T_b$ (left) and the equilateral triangle $\phi_e(T_b)=T_e$ (right).

Corollary 2.5. *If $\phi_e(N_{\mathcal{Y}}(x))=N_{\phi_e(\mathcal{Y})}(\phi_e(x))$ for all $x\in T_b$, then the arc probability $p_a(N_{\mathcal{Y}})$ of the PCD based on $N_{\mathcal{Y}}$ for uniform data on T_b is geometry invariant.*

Proof. Let $x\in T_b$. Then $\phi_e(N_{\mathcal{Y}}(x))=N_{\phi_e(\mathcal{Y})}(\phi_e(x))$ implies $A(\phi_e(N_{\mathcal{Y}}(x)))=A(N_{\phi_e(\mathcal{Y})}(\phi_e(x)))$. Hence the result follows by Theorem 2.4. □

2.4. Vertex and edge regions

The new proximity maps will be based on the Delaunay cell $\mathcal{T}_i$ that contains x. The region $N_{\mathcal{Y}}(x)$ will also depend on the location of x in $\mathcal{T}_i$ with respect to the vertices or faces (edges in $\mathbb{R}^2$) of $\mathcal{T}_i$. Hence for $N_{\mathcal{Y}}(x)$ to be well-defined, the vertex or face of $\mathcal{T}_i$ associated with x should be uniquely determined. This will give rise to two new concepts: *vertex regions* and *face regions* (*edge regions* in $\mathbb{R}^2$).

2.4.1. Triangle centers

The vertex and edge regions will be constructed using a point, preferably, in the interior of the triangle, e.g., a *triangle center*. The *trilinear coordinates* of a point P with respect to $T(\mathcal{Y}_3)$ are an ordered triple of numbers, which are proportional to the distances from P to the edges. Trilinear coordinates are denoted as $(\alpha:\beta:\gamma)$ and also are known as *homogeneous coordinates* or *trilinears*. By convention, the three vertices $\mathsf{y}_1,\mathsf{y}_2$, and y_3 of $T(\mathcal{Y}_3)$ are commonly written as $(1:0:0)$, $(0:1:0)$, and $(0:0:1)$, respectively (see [27]).

Definition 2.6. A *triangle center* is a point whose trilinear coordinates are defined in terms of the edge lengths and (inner) angles of a triangle. The function giving the coordinates $(\alpha:\beta:\gamma)$ is called the *triangle center function*.

Kimberling [15] enumerates 360 triangle centers, among which four have been widely known since the ancient times; namely, *circumcenter* (M_{CC}), *incenter* (M_I), *center of mass* or *centroid* (M_{CM}), and *orthocenter* (M_O).

The trilinear coordinates of the circumcenter M_{CC} are $(\cos\theta_1:\cos\theta_2:\cos\theta_3)$ where θ_i is the inner angle of $T(\mathcal{Y}_3)$ at vertex y_i for $i\in\{1,2,3\}$. The circumcenter of a triangle is in the interior, at the midpoint of the hypotenuse, or in the exterior of the triangle, if the triangle is acute, right, or obtuse, respectively. See Fig. 8 where an acute and an obtuse triangle are depicted. Using the pdf of an arbitrary angle of a triangle T_i from Poisson Delaunay triangulation $\mathcal{D}_P$ [18], we see that, $P(T_i \text{ is a right triangle})=P(\theta=\pi/2)=0$, hence $P(M_{CC}$ is the midpoint of the hypotenuse$)=0$. Furthermore,

$$P(T_i \text{ is an obtuse triangle}) = P(M_{CC}\notin T_i) = P(\theta_{\max}>\pi/2) = \int_{\pi/2}^{\pi} f_3(x)\,dx$$
$$= \frac{3f_S(\sqrt{2\pi}) - f_C(\sqrt{2\pi}) - 3f_S(\sqrt{\frac{\pi}{2}}) + f_C(\sqrt{\frac{\pi}{2}})}{\sqrt{2\pi}} \approx .03726$$

where $f_3(x)$ is the pdf of the maximum angle and is given by

$$f_3(x) = \left[\frac{2}{\pi}\left(3x(\sin 2x) - \cos 2x + \cos 4x - \pi\sin 2x\right)\right]\mathbf{I}(\pi/3<x<\pi/2)$$
$$+\left[\frac{1}{\pi}\left(4\pi(\cos x)(\sin x) + 3\sin x^2 - \cos x^2 - 4x(\cos x)(\sin x) + 1\right)\right]\mathbf{I}(\pi/2<x<\pi),$$

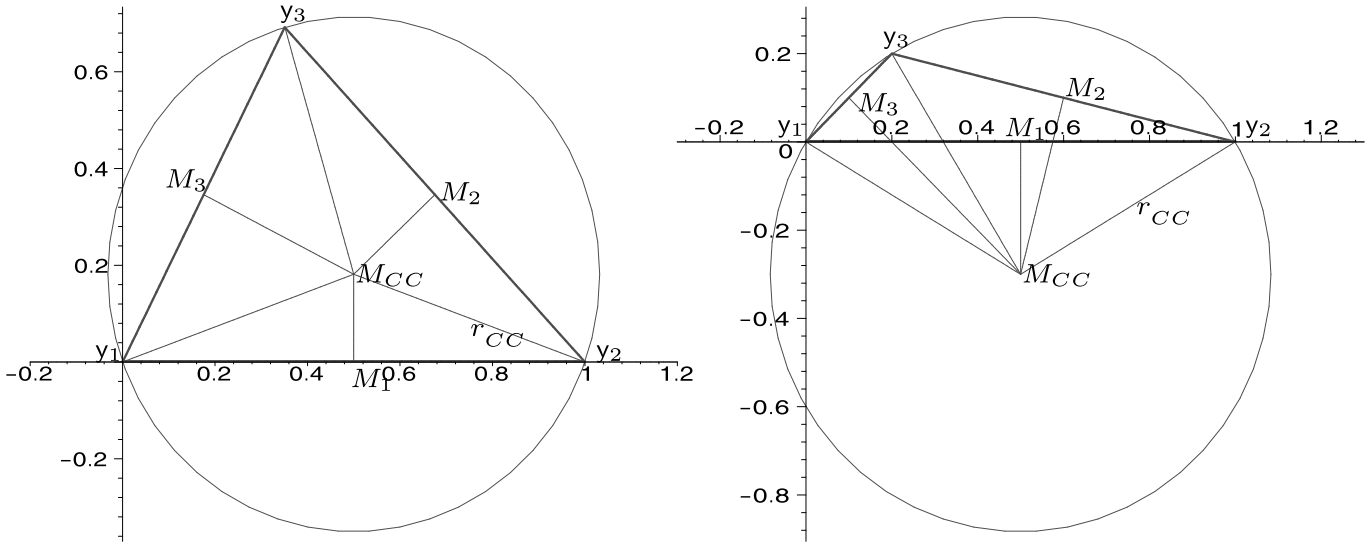

Fig. 8. The circumcircle, circumcenter M_{CC}, and circumradius r_{CC} of an acute triangle (left) and an obtuse triangle (right). M_i is the midpoint of edge e_i, for $i = 1, 2, 3$.

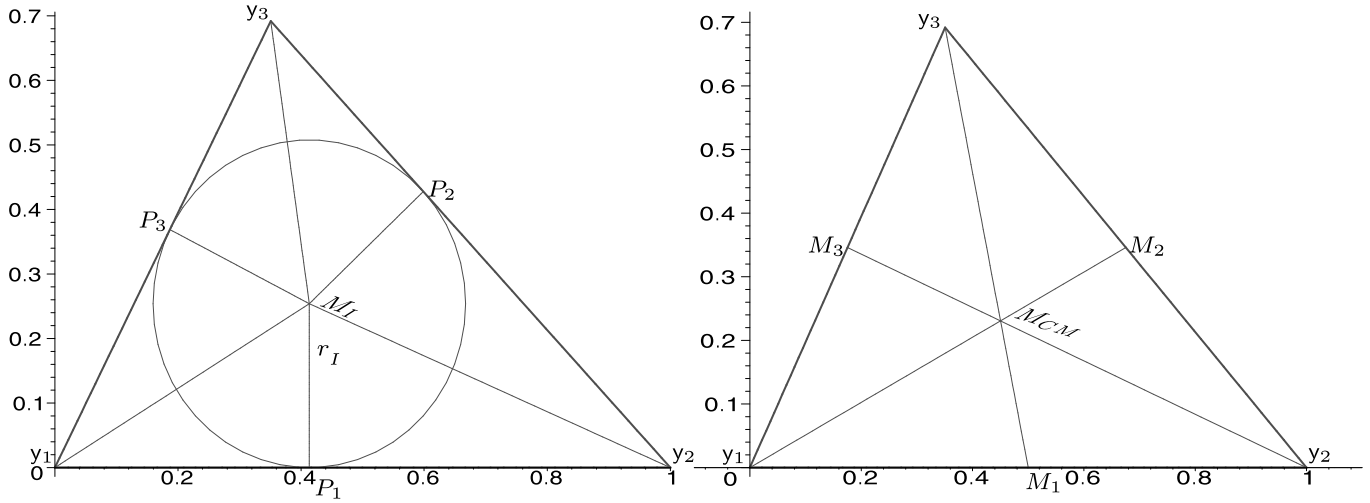

Fig. 9. The incircle, incenter M_I, inradius r_I of a triangle (left) and the centroid or center of mass of a triangle (right). P_i is the point where the incircle is tangent to edge e_i for $i = 1, 2, 3$.

$f_C(x) = \int_0^x \cos(\pi t^2/2)\,dt$, and $f_S(x) = \int_0^x \sin(\pi t^2/2)\,dt$ are the Fresnel cosine and sine functions, respectively. The coordinates of M_{CC} in the basic triangle T_b are $(\frac{1}{2}, \frac{c_1^2 - c_1 + c_2^2}{2c_2})$.

The incenter M_I and has trilinear coordinates $(1 : 1 : 1)$. See Fig. 9 (left). The coordinates of M_I for the basic triangle T_b are (x_I, y_I), where

$$x_I = \frac{c_1 - \sqrt{c_1^2 + c_2^2}}{1 + \sqrt{c_1^2 + c_2^2} + \sqrt{(1-c_1)^2 + c_2^2}}, \qquad y_I = \frac{c_2}{1 + \sqrt{c_1^2 + c_2^2} + \sqrt{(1-c_1)^2 + c_2^2}}.$$

Unlike the circumcenter, the incenter is guaranteed to be inside the triangle.

The *median line* of a triangle is the line from one of its vertices to the midpoint of the opposite edge. The three median lines of any triangle intersect at the triangle's *centroid* (i.e., *center of mass*), denoted as M_{CM}. See Fig. 9 (right). It has trilinear coordinates $(1/|e_1| : 1/|e_2| : 1/|e_3|)$ or $(\csc\theta_1 : \csc\theta_1 : \csc\theta_1)$ where e_i denotes the edge opposite to the vertex y_i for $i \in \{1, 2, 3\}$. The centroid is also guaranteed to be in the interior of the triangle. The coordinates of M_{CM} in the basic triangle are $((1 + c_1)/3, c_2/3)$.

The intersection of the three altitudes of a triangle is called the *orthocenter*, M_O, which has trilinear coordinates $(\cos\theta_2 \cos\theta_3 : \cos\theta_1 \cos\theta_3 : \cos\theta_1 \cos\theta_2)$. The orthocenter of a triangle is in the interior, at vertex y_3, or in the exterior of the basic triangle, T_b, if T_b is acute, right, or obtuse, respectively. The functional form of M_O in the basic triangle is $(c_1, c_1(1 - c_1)/c_2)$.

Note that in an equilateral triangle, $M_I = M_{CC} = M_O = M_{CM}$ (i.e., all the centers we have described coincide).

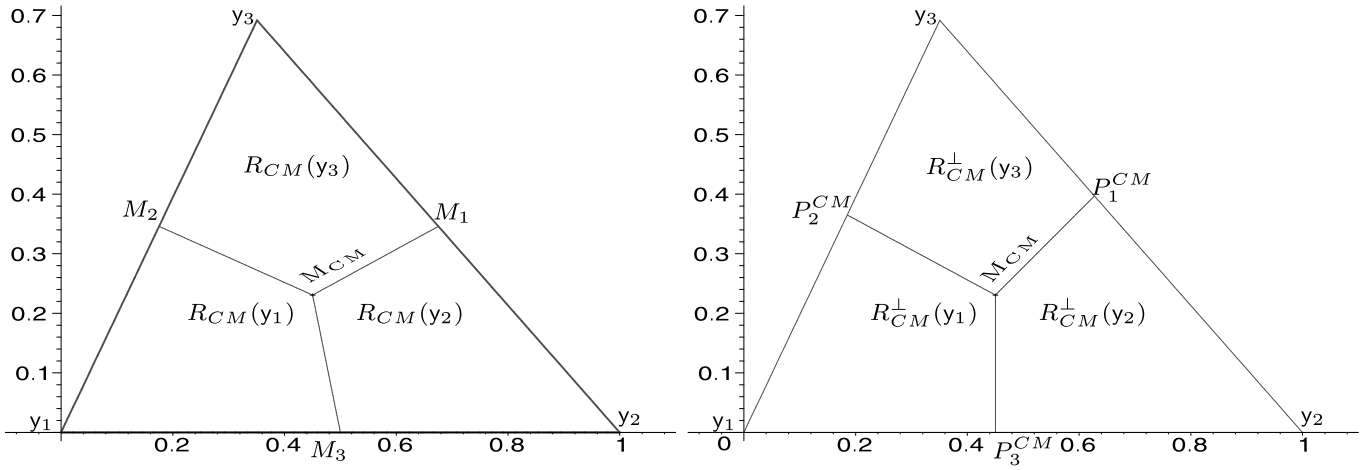

Fig. 10. The M_{CM}-vertex regions with median lines (left) and with orthogonal projections (right). P_i^{CM} is the point where the orthogonal projection from M_{CM} crosses edge e_i for $i = 1, 2, 3$.

2.4.2. Vertex regions

Recall that for $x \in T(\mathcal{Y}_3)$, $N_S(x) = B(x, r(x))$ where $r(x) = \min_{y \in \mathcal{Y}_3} d(x, y)$. That is, $r(x) = d(x, y_i)$ iff $x \in \mathcal{V}_C(y_i) \cap T(\mathcal{Y}_3)$ for $i \in \{1, 2, 3\}$, where $\mathcal{V}_C(y_i)$ is the Voronoi cell generated by y_i in the Voronoi diagram based on $\mathcal{Y}_3$. Notice that these cells partition the triangle $T(\mathcal{Y}_3)$ and each $\mathcal{V}_C(y_i) \cap T(\mathcal{Y}_3)$ is adjacent only to vertex y_i and their intersection is the point M which is equidistant to the vertices. So M is in fact the circumcenter, M_{CC}, of $T(\mathcal{Y}_3)$. To define new proximity regions based on some sort of distance or dissimilarity relative to the vertices $\mathcal{Y}_3$, we associate each point in $T(\mathcal{Y}_3)$ to a vertex of $T(\mathcal{Y}_3)$ as in the spherical case. This gives rise to the concept of *vertex regions*. Note that $N_S(x)$ is constructed using the vertex region based on the closest vertex, $\text{argmin}_{y \in \mathcal{Y}_3} d(x, y)$. If two vertices were equidistant from x (i.e., $\text{argmin}_{y \in \mathcal{Y}_3} d(x, y)$ were not unique), x is arbitrarily assigned to a region of one of them. In fact, for N_S, by construction, it would not matter which vertex to pick when the vertices are equidistant to x, the region $N_S(x)$ will be the same.

Definition 2.7. The connected regions that partition the triangle, $T(\mathcal{Y}_3)$ (in the sense that the intersections of the regions have zero $\mathbb{R}^2$-Lebesgue measure) such that each region has one and only one vertex of $T(\mathcal{Y}_3)$ on its boundary are called *vertex regions*.

This definition implies that there are three vertex regions. The vertex regions can be constructed starting with a point $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$. Join the point M to a point on each edge by a curve such that the resultant regions satisfy the above definition. Such regions are called *M-vertex regions* and we denote the vertex region associated with vertex y as $R_M(y)$ for $y \in \mathcal{Y}_3$. In particular, one can use a *center* of the triangle $T(\mathcal{Y}_3)$ as the starting point M for vertex regions. The points in $R_M(y)$ can be thought as being "closer" to y than to the other vertices.

It is reasonable to require that the area of the region $R_M(y)$ gets larger as $d(M, y)$ increases. Usually the curves will be taken to be lines or even the orthogonal projections to the edges. But these lines do not necessarily yield three vertex regions for M in the exterior of $T(\mathcal{Y}_3)$. Unless stated otherwise, M-vertex regions will refer to regions constructed by joining M to the edges with *straight line segments*, henceforth.

We construct M-vertex regions by straight lines in the following two ways:

Method I: with the extensions of the line segments joining y **to** M**:** Let $T(\mathcal{Y}_3)^o$ denote the interior of the triangle $T(\mathcal{Y}_3)$. M-vertex regions with $M \in T(\mathcal{Y}_3)^o$ can be constructed by using the extensions of the line segments joining y to M for each $y \in \mathcal{Y}_3$. See Fig. 10 (left) with $M = M_{CM}$. The functional forms of $R_M(y_i)$ for $i \in \{1, 2, 3\}$ with $M = (m_1, m_2)$ and $m_1 > c_1$ in the basic triangle, T_b, are provided in [2]. If x falls on the boundary of two M-vertex regions, then x is arbitrarily assigned to one of the M-vertex regions.

Method II: with the orthogonal projections from M **to edges:** In this method, we draw the orthogonal projections from M to the edges to obtain the vertex regions denoted as $R^{\perp}_M(y)$. For instance see Fig. 10 (right) with $M = M_{CM}$. The functional forms of $R^{\perp}_M(y)$ for $M = (m_1, m_2)$ in the basic triangle are provided in [2]. However, the orthogonal projections from M to the edges does not necessarily fall on the boundary of $T(\mathcal{Y}_3)$. For example, letting P_2^M be the orthogonal projection of M to edge e_2, it is easy to see that P_2^M might fall outside $T(\mathcal{Y}_3)$ which contradicts the definition of vertex regions. In fact $P_2^M \in e_2$ iff $\frac{c_2(m_2c_2+c_1m_1)}{c_1^2+c_2^2} \leqslant c_2$ iff $c_2(c_2 - m_2) + c_1(c_1 - m_1) \geqslant 0$. By definition, $R_O(y)$ and $R^{\perp}_O(y)$ are identical. But, for $M \in \{M_{CM}, M_{CC}, M_I\}$, $R_M(y)$ can have both versions.

We define and provide the explicit forms of M_{CC}-vertex regions, M_{CM}-vertex regions, and M_I-vertex regions in [2]. See also Fig. 11 for M_{CC}-vertex regions with Method II (i.e., with orthogonal projections) for acute and obtuse triangles; and Fig. 12 M_I-vertex regions with Methods I and II.

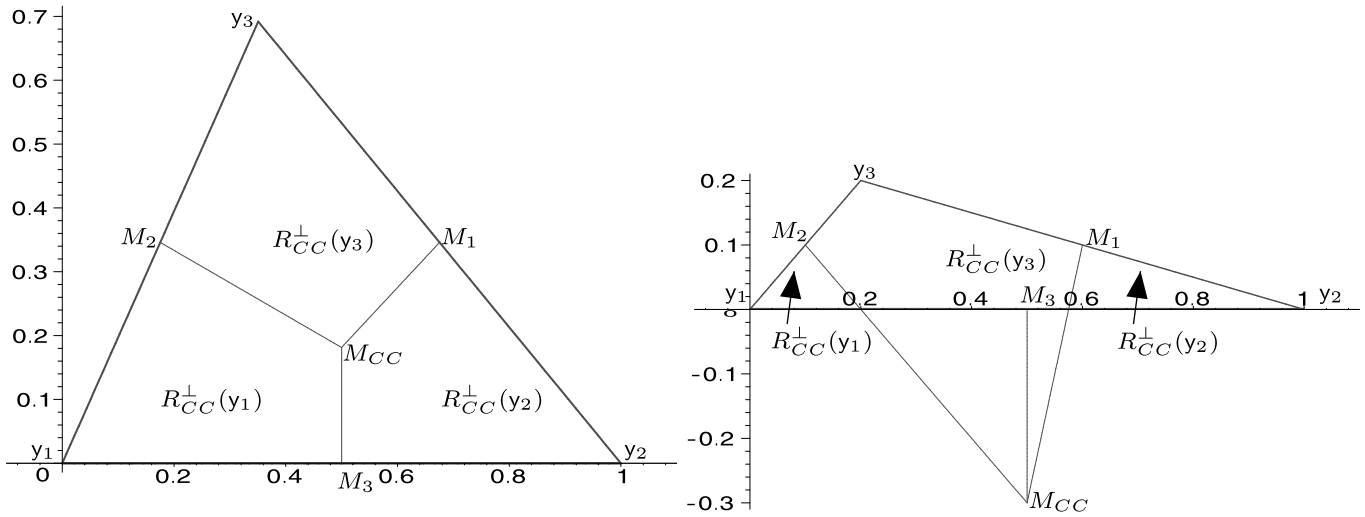

Fig. 11. The M_{CC}-vertex regions with orthogonal projections in an acute triangle (left) and in an obtuse triangle (right).

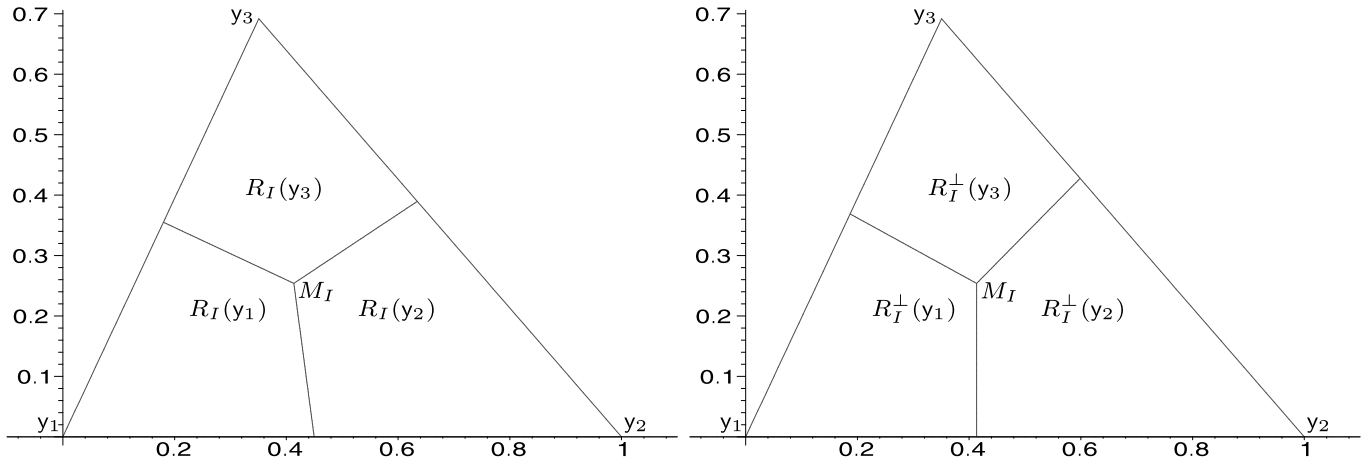

Fig. 12. The M_I-vertex regions with extension of the line segments joining the vertices (i.e., with Method I) to M_I (left) and with orthogonal projections (i.e., with Method II) (right).

2.4.3. Edge regions

The spherical proximity region seen earlier is constructed by using the vertex region based on the closest vertex, $\text{argmin}_{y\in\mathcal{Y}_3} d(x, y)$. One can also use the closest edge, $\text{argmin}_{i\in\{1,2,3\}} d(x, e_i)$, in defining a proximity region, which suggests the concept of *edge regions*. While using the edge $\text{argmin}_{i\in\{1,2,3\}} d(x, e_i)$, the triangle is again partitioned into three regions whose intersection is some point M with Euclidean distance to the edges $d(M, e_1) = d(M, e_2) = d(M, e_3)$, so M is in fact the incenter of $T(\mathcal{Y}_3)$ and $d(M, e) = r_I$ is the inradius.

Definition 2.8. The connected regions that partition the triangle, $T(\mathcal{Y}_3)$, in such a way that each region has one and only one edge of $T(\mathcal{Y}_3)$ on its boundary, are called *edge regions*.

This definition implies that there are exactly three edge regions which intersect at only one point, M in $T(\mathcal{Y}_3)^o$. In fact, one can describe the edge regions starting with M. Join the point M to the vertices by curves such that the resultant regions satisfy the above definition. Such regions are called *M-edge regions* and the edge region for edge e is denoted as $R_M(e)$ for $e \in \{e_1, e_2, e_3\}$. Unless stated otherwise, M-edge regions will refer to the regions constructed by joining M to the vertices by straight lines, henceforth. In particular, one can use a *center* of $T(\mathcal{Y}_3)$ for the starting point M. See Fig. 13 for M-edge regions with $M = M_{CM}$ and $M = M_I$. One can also consider the points in $R_M(e)$ to be "closer" to e than to the other edges. Furthermore, it is reasonable to require that the area of the region $R_M(e)$ get larger as $d(M, e)$ increases. In higher dimensions, the corresponding regions are called "face regions".

The functional forms of $R_M(e_i)$ for $i \in \{1, 2, 3\}$, for $M = (m_1, m_2) \in T(\mathcal{Y}_3)^o$ and $m_1 > c_1$ in the basic triangle are provided in [2]. If x falls on the boundary of two M-edge regions, then it is arbitrarily assigned to one of the M-edge regions. The center of mass edge regions (M_{CM}-edge regions) and other edge regions are described in detail in [2].

Below are the results about the geometry invariance of vertex- and edge-regions.

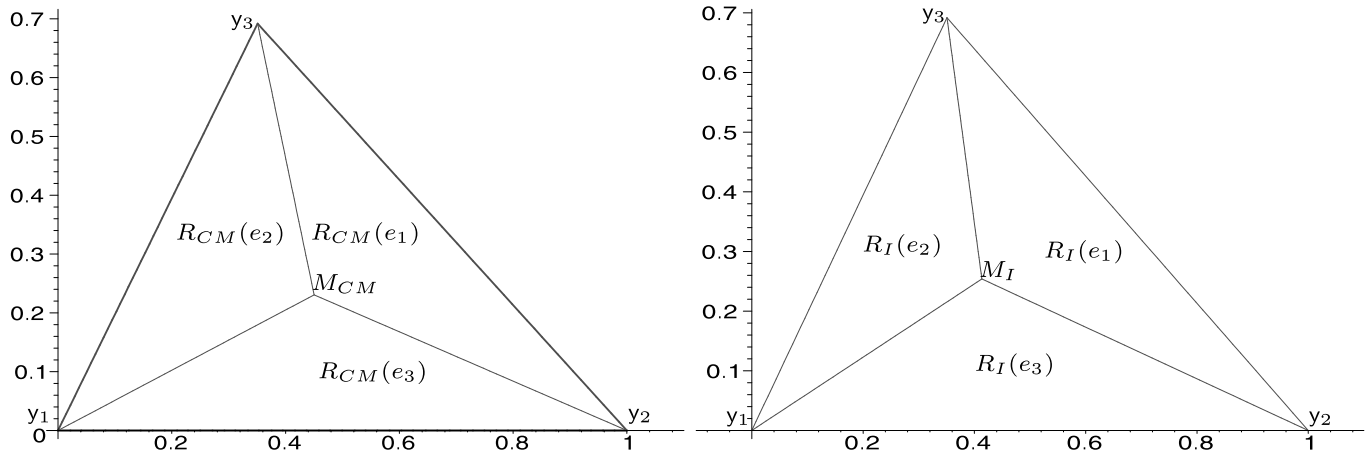

Fig. 13. The M-edge regions with $M = M_{CM}$ (left) and $M = M_I$ (right).

Definition 2.9. The M-edge regions are said to be *geometry invariant* if $\phi_e(R_M(e_i)) = R_{\phi_e(M)}(\phi_e(e_i))$ for $i = 1, 2, 3$ where ϕ_e is the transformation defined in Section 2.3.1. The M-vertex regions are said to be *geometry invariant* if $\phi_e(R_M(\mathrm{y}_i)) = R_{\phi_e(M)}(\phi_e(\mathrm{y}_i))$ for $i = 1, 2, 3$.

As a corollary to Theorem 2.4, we obtain the following.

Corollary 2.10. *Suppose $N_{\mathcal{Y}}$ is based on geometry invariant edge or vertex regions. If the proximity regions are based on boundary of $T(\mathcal{Y}_3)$ and parallel lines to edges, then geometry invariance of the arc probability for uniform data follows.*

Proof. Such proximity maps with geometry invariant edge or vertex regions, satisfy $\phi_e(N_{\mathcal{Y}}(x)) = N_{\phi_e(\mathcal{Y})}(\phi_e(x))$. Hence the desired result holds by Corollary 2.5. □

Corollary 2.11. *If the edge or vertex regions are based on specific angles in T_b in the sense that they have specific (inner) angular values, then these regions are not geometry invariant. Similarly, if the proximity regions are based on specific angles in T_b then they are not geometry invariant either.*

Proof. The transformation ϕ_e clearly does not preserve the angles in T_b. Hence the regions dependent on (inner) angles of T_b fail to be preserved. □

3. Families of proximity regions in Delaunay tessellations

Let $\mathcal{Y}_m$ be m points in general position in $\mathbb{R}^d$. Moreover, let $\mathcal{X}_n$ be a random sample from F with support $\mathcal{S}(F) \subseteq \mathcal{C}_H(\mathcal{Y}_m)$. That is, $\Omega = \mathcal{C}_H(\mathcal{Y}_m)$ and $\Omega_i = \mathcal{T}_i$ with μ being the Lebesgue measure. Then the appealing properties for proximity regions in Section 2.1 can be extended to this special case also [2].

For illustrative purposes, we focus on $\mathbb{R}^2$, where a Delaunay tessellation is a triangulation, provided that no more than three points of $\mathcal{Y}_m$ are cocircular. Let $\mathcal{X}_n$ be a random sample from F with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. The spherical proximity map is the first proximity map defined in the literature (see [12,17,22,23,11]) where M_{CC}-vertex regions with Method II were implicitly used for points in $\mathcal{C}_H(\mathcal{Y}_m)$. In the following sections, we will describe arc-slice proximity maps $N_{AS}(\cdot)$, two families of proximity regions for which **P4** and **P5** will automatically hold, and introduce two new families of proximity regions.

3.1. Arc-slice proximity maps

Recall that for $N_S(\cdot)$, **P7** is violated, since for any $x \in \mathcal{T}_i \subset \mathbb{R}^d$, $B(x, r(x)) \not\subset \mathcal{T}_i$, which implies that two proximity regions $N_S(x)$ and $N_S(y)$ might overlap for x and y in two distinct cells. See, e.g., Fig. 5. Such an overlap of the regions make the distribution of the domination number of the PCD associated with $N_S(\cdot)$, if not impossible, hard to calculate. In order to avoid the overlap of regions $B(x, r(x))$ and $B(y, r(y))$ for x, y in different Delaunay cells, the balls are restricted to the corresponding cells, which leads to *arc-slice proximity regions*, $N_{AS}(x) := \bar{B}(x, r(x)) \cap \mathcal{T}_i$, where $\bar{B}(x, r(x))$ is the closure of the ball $B(x, r(x))$. The closed ball is used in the definition of the arc-slice proximity map for consistency with the other proximity maps that will be defined on Delaunay cells. The arc-slice proximity map $N_{AS}(x)$ is well-defined only for points in $\mathcal{C}_H(\mathcal{Y}_m)$, provided that $\mathcal{Y}_m$ is in general position and $m \geqslant (d+1)$ in $\mathbb{R}^d$.

By construction, the M_{CC}-vertex regions with Method II are implicitly used, since x is in the M_{CC}-vertex region of y iff $\mathrm{y} = \mathrm{argmin}_{u \in \mathcal{Y}_m} d(x, u)$. To make this dependence explicit, we will use the notations $N_{AS}(\cdot, M_{CC}^{\perp})$ and $R_{CC}^{\perp}(\mathrm{y})$ for the

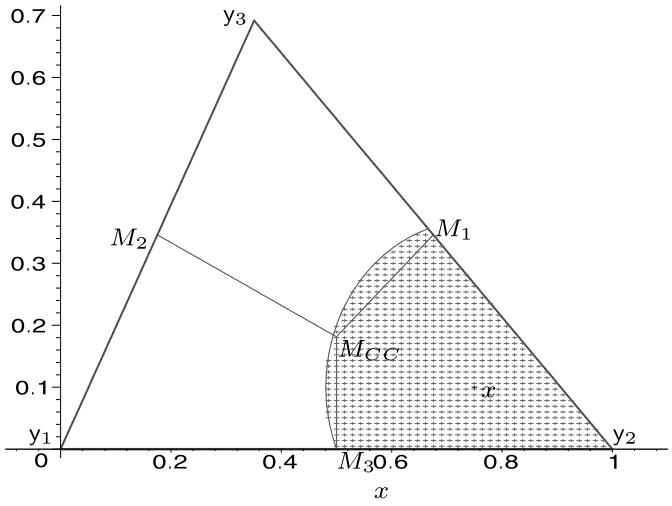

Fig. 14. The region $N_{AS}(x, M_{CC}^{\perp})$ with an $x \in R_{CC}(y_2)$.

proximity region and the vertex region, respectively. See Fig. 14 for $N_{AS}(x, M_{CC}^{\perp})$ for an $x \in R_{CC}^{\perp}(y_2)$. The properties **P1**, **P2**, **P7** hold by definition. Notice that $N_{AS}(x, M_{CC}^{\perp}) \subseteq T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3)$ and $N_{AS}(x, M_{CC}^{\perp}) = T(\mathcal{Y}_3)$ iff $x = M_{CC}$, since $\bar{B}(x, r(x)) \supset T(\mathcal{Y}_3)$ only when $x = M_{CC}$. Hence the superset region for arc-slice proximity maps with M_{CC}-vertex regions is $\mathcal{R}_S^{\perp}(N_{AS}, M_{CC}) = \{M_{CC}\}$. Notice the $^{\perp}$ in the superscript to indicate the construction with Method II. So property **P6** follows. Furthermore, **P8** holds since the area $A(N_{AS}(x, M_{CC}^{\perp}))$ is a continuous function of $r(x) = \min_{y \in \mathcal{Y}_3} d(x, y)$ which is a continuous function of x. Properties **P3**, **P4**, **P5**, and **P9** are violated for $N_{AS}(x, M_{CC}^{\perp})$. See Fig. 15 for the arcs based on $N_{AS}(x, M_{CC}^{\perp})$ for a realization of 7 class $\mathcal{X}$ points in the one triangle case, and for the realization of 77 class $\mathcal{X}$ points in the multi-triangle case in Fig. 4 (right).

One can define arc-slice proximity regions with any type of M-vertex regions with Method II as

$$N_{AS}\big(x, M^{\perp}\big) := \bar{B}\big(x, r(x)\big) \cap T(\mathcal{Y}_3) \quad \text{where } r(x) := d(x, y) \text{ for } x \in R_M(y).$$

But for $M \neq M_{CC}$, $N_{AS}(\cdot, M^{\perp})$ satisfies only **P1**, **P2**, and **P7**, property **P6** fails to hold, since $\mathcal{R}_S^{\perp}(N_{AS}, M)$ has positive area, and **P8** fails, since the size of $N_{AS}(x, M^{\perp})$ is not continuous in x. See [2] for illustrations of $\mathcal{R}_S^{\perp}(N_{AS}, M)$ with $M = M_{CM}$ and $M = M_I$.

The arc-slice proximity regions can also be defined with M-vertex regions constructed as in Method I (i.e., with the extensions of the line segments joining M to the vertices). Such proximity regions are denoted as $N_{AS}(x, M)$ as opposed to $N_{AS}(x, M^{\perp})$. The $N_{AS}(\cdot, M)$ satisfies the same properties as $N_{AS}(\cdot, M^{\perp})$, except property **P8**. That is, $N_{AS}(\cdot, M)$ violates property **P8** for all M.

In terms of the appealing properties in Section 2.1, $N_{AS}(\cdot, M_{CC}^{\perp})$ is the most appealing proximity map in the family $\mathcal{N}_{AS} := \{N_{AS}(\cdot, M)\colon\ M \in \mathbb{R}^2 \setminus \mathcal{Y}_3\} \cup \{N_{AS}(\cdot, M^{\perp})\colon\ M \in \mathbb{R}^2 \setminus \mathcal{Y}_3\}$. Moreover, $\Lambda_0(N_{AS}, M) = \mathcal{Y}_3$ for all $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, since $\lambda(N_{AS}(x, M)) = 0$ iff $x \in \mathcal{Y}_3$.

3.2. Proportional-edge proximity maps

The first type of triangular proximity map we will consider is the proportional-edge proximity map. For this proximity map, the asymptotic distribution of domination number and the relative density of the corresponding PCD has mathematical tractability. See [5,6,8].

For the expansion parameter $r \in [1, \infty]$, define $N_{PE}^r(\cdot, M) := N(\cdot, M; r, \mathcal{Y}_3)$ to be the *proportional-edge proximity map* with M-vertex regions obtained as in Method I as follows (see also Fig. 16 with $M = M_{CM}$ and $r = 2$). For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $v(x) \in \mathcal{Y}_3$ be the vertex whose region contains x; i.e., $x \in R_M(v(x))$. If x falls on the boundary of two M-vertex regions, $v(x)$ arbitrarily assigned. Let $e(x)$ be the edge of $T(\mathcal{Y}_3)$ opposite $v(x)$. Let $\ell(v(x), x)$ be the line parallel to $e(x)$ through x. Let $d(v(x), \ell(v(x), x))$ be the Euclidean distance from $v(x)$ to $\ell(v(x), x)$. For $r \in [1, \infty)$, let $\ell_r(v(x), x)$ be the line parallel to $e(x)$ such that

$$d\big(v(x), \ell_r\big(v(x), x\big)\big) = r d\big(v(x), \ell\big(v(x), x\big)\big) \quad \text{and} \quad d\big(\ell\big(v(x), x\big), \ell_r\big(v(x), x\big)\big) < d\big(v(x), \ell_r\big(v(x), x\big)\big).$$

Let $T_r(x)$ be the triangle similar to and with the same orientation as $T(\mathcal{Y}_3)$ having $v(x)$ as a vertex and $\ell_r(v(x), x)$ as the opposite edge. Then the *proportional-edge proximity region* $N_{PE}^r(x, M)$ is defined to be $T_r(x) \cap T(\mathcal{Y}_3)$. Notice that $\ell(v(x), x)$ divides the edges of $T_r(x)$ (other than $\ell_r(v(x), x)$) proportionally with the factor r. Hence the name *proportional edge proximity map* and the notation $N_{PE}^r(\cdot, M)$.

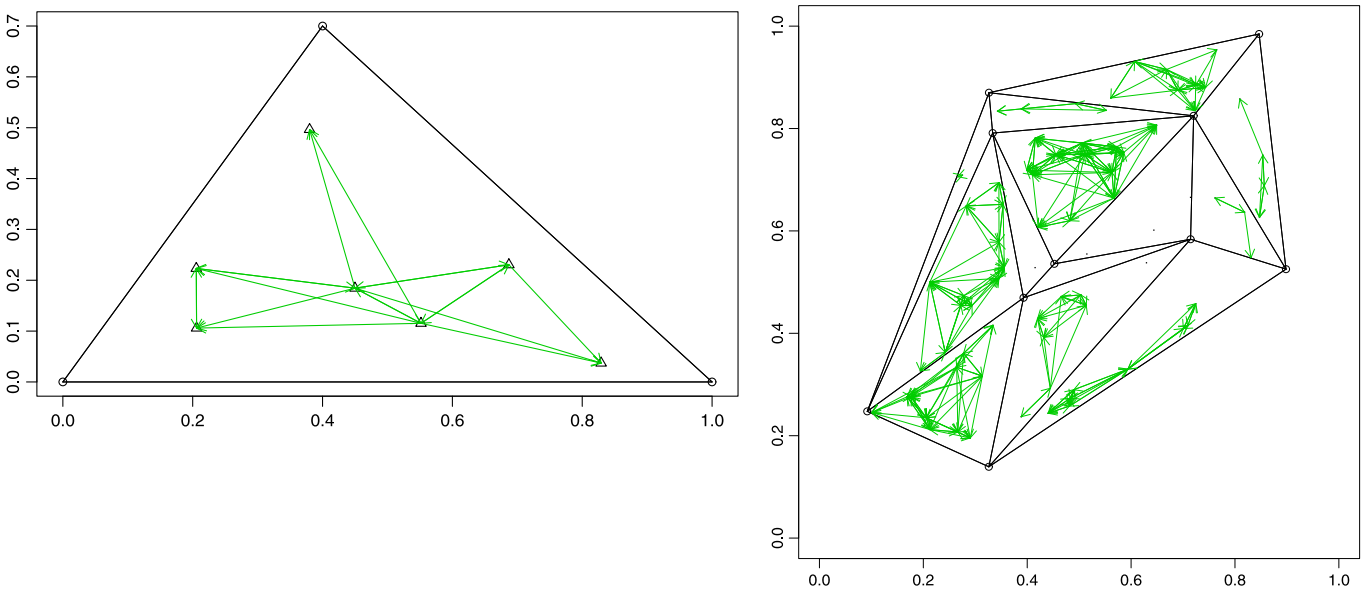

Fig. 15. A realization of 7 class $\mathcal{X}$ points (small triangles) generated iid $\mathcal{U}(T(\mathcal{Y}_3))$ and the corresponding arcs for $N_{AS}(x, M_{CC}^{\perp})$ (left). The arcs for arc-slice PCDs with $N_{AS}(x, M_{CC}^{\perp})$ for the 77 class $\mathcal{X}$ points that lie in the $\mathcal{C}_H(\mathcal{Y}_{10})$ (right) where the same $\mathcal{Y}_{10}$ in Fig. 4 is used.

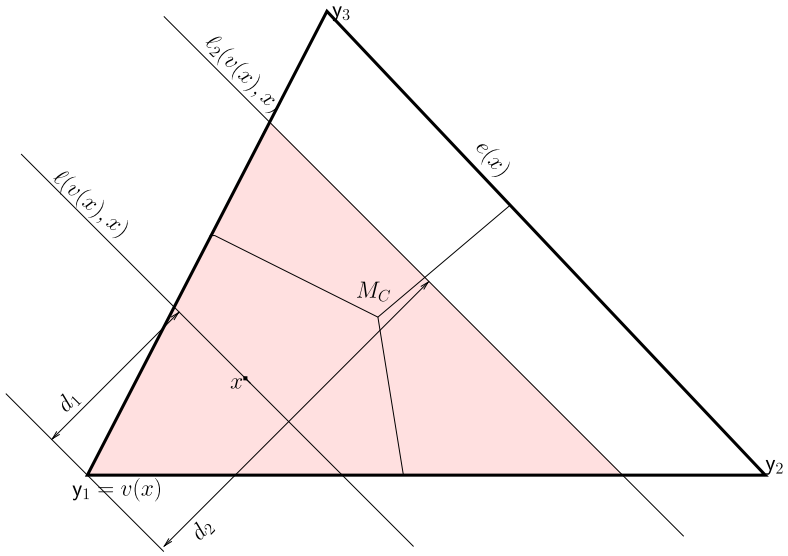

Fig. 16. Construction of proximity region, $N_{PE}^{2}(x)$ (shaded region) for an $x \in R_{CM}(\mathsf{y}_1)$ where $d_1 = d(v(x), \ell(v(x), x))$ and $d_2 = d(v(x), \ell_2(v(x), x)) = 2d(v(x), \ell(v(x), x))$.

Notice that $r \geqslant 1$ implies $x \in N_{PE}^{r}(x, M)$. Furthermore, $\lim_{r\to\infty} N_{PE}^{r}(x, M) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, so $N_{PE}^{\infty}(x, M) := T(\mathcal{Y}_3)$ for all such x. For $x \in \mathcal{Y}_3$, $N_{PE}^{r}(x, M) := \{x\}$ for all $r \in [1, \infty]$. See Fig. 17 for the arcs based on $N_{PE}^{r=2}(x, M_{CM})$ in the one triangle and the multi-triangle cases.

For $X_i \overset{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional pdf f exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$, implies that the special case in the construction of $N_{PE}^{r} - X$ falls on the boundary of two vertex regions – occurs with probability zero. For such an F, $N_{PE}^{r}(X, M)$ is a triangle a.s. The functional form of $N_{PE}^{r}(x, M)$ for $x = (x_0, y_0) \in T_b$ is given in the technical report by [2].

The proportional-edge PCDs based on vertex regions constructed as in Method II are denoted as $N_{PE}^{r}(x, M^{\perp})$ and the corresponding superset region is denoted as $\mathcal{R}_S^{\perp}(N_{PE}^{r}, M)$. See Fig. 18 for the superset region $\mathcal{R}_S^{\perp}(N_{PE}^{2}, M)$ with $M \in \{M_{CC}, M_I\}$. On the other hand, the superset region for $N_{PE}^{r}(x, M)$ is denoted as $\mathcal{R}_S(N_{PE}^{r}, M)$.

Of particular interest is N_{PE}^{r} with any M and $r \in \{\sqrt{2}, 3/2, 2\}$. For $r = \sqrt{2}$, $\ell(v(x), x)$ divides $T_{\sqrt{2}}(x)$ into two regions of equal area, hence $N_{PE}^{\sqrt{2}}$ is also referred to as *double-area proximity map*. See, e.g., Fig. 19 (left) for an illustration of $N_{PE}^{\sqrt{2}}(x, M_I^{\perp})$. For $r = 2$, $\ell(v(x), x)$ divides the edges of $T_2(x)$ – other than $\ell_r(v(x), x)$ – into two segments of equal length, hence N_{PE}^{2} is also referred to as *double-edge proximity map*. See, e.g., Fig. 19 (right) for an illustration of $N_{PE}^{2}(x, M_I^{\perp})$. The superset region is empty for $r < 3/2$, has positive area for $r > 3/2$; and is $\{M_{CM}\}$ for $r = 3/2$. Therefore, $r = 3/2$ is the

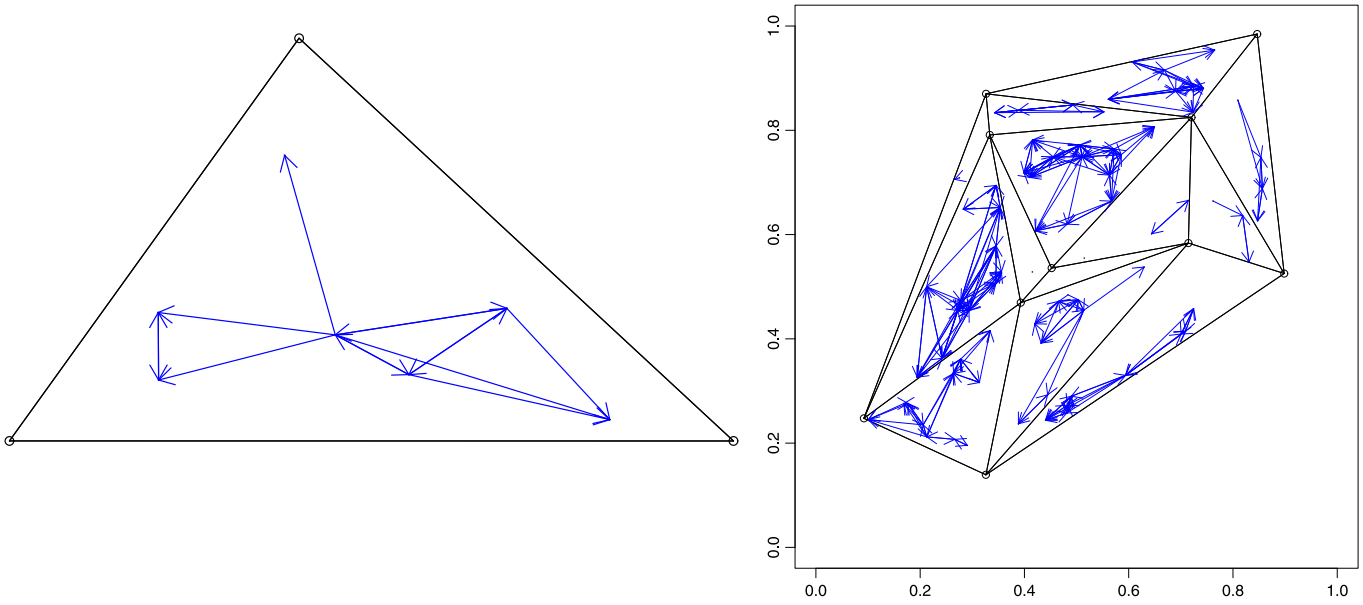

Fig. 17. The arcs for $N_{PE}^{r=2}(x, M_{CM})$ the 7 class $\mathcal{X}$ points in Fig. 15, and the arcs for $N_{PE}^{r=2}(x, M_{CM})$ for the 77 class $\mathcal{X}$ points that lie in the $\mathcal{C}_H(\mathcal{Y}_{10})$ where the same $\mathcal{Y}_{10}$ in Fig. 4 is used.

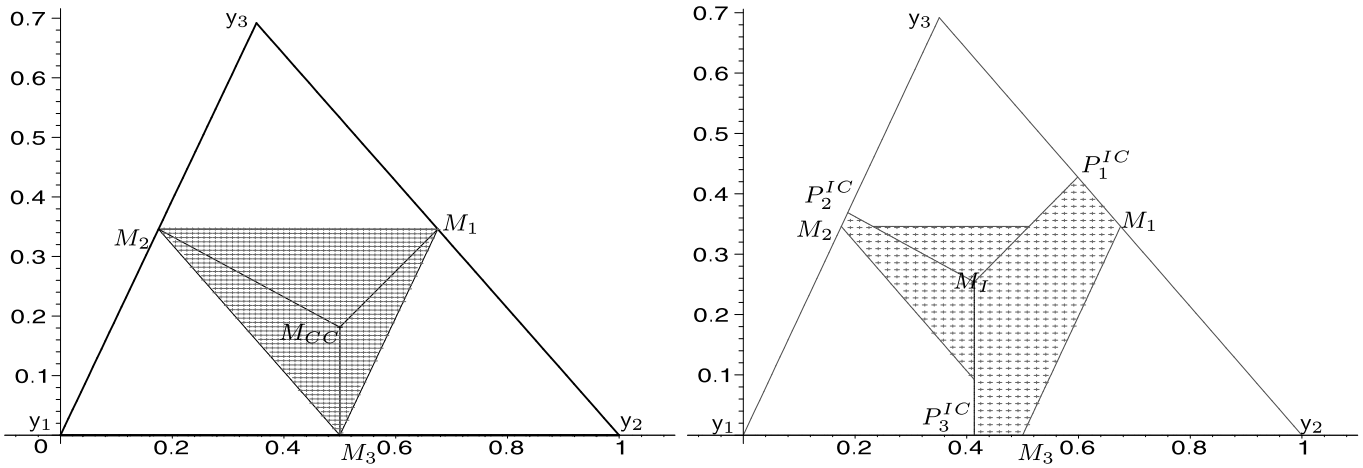

Fig. 18. The superset regions $\mathcal{R}_S^{\perp}(N_{PE}^2, M_{CC})$ in an acute triangle (left), and $\mathcal{R}_S^{\perp}(N_{PE}^2, M_I)$ (right). P_i^{IC} is the point where the orthogonal projection from M_I crosses edge e_i for $i = 1, 2, 3$.

threshold for $N_{PE}^r(\cdot, M_{CM})$ (and $N_{PE}^r(\cdot, M_{CM}^{\perp})$) to satisfy **P6**. Furthermore, $r = 3/2$ is the value at which the asymptotic distribution of the domination number of the PCD based on $N_{PE}^r(\cdot, M_{CM})$ is non-degenerate (see [1] and [5]).

For $N_{PE}^r(x, M)$, the properties **P1**, **P2**, **P4**, **P5**, and **P7** follow by definition for all M and r. Furthermore **P9** holds, since $N_{PE}^r(x, M)$ is geometry invariant for uniform data. Property **P5** holds with similarity ratio of $N_{PE}^r(x, M)$ to $T(\mathcal{Y}_3)$: $[\min(d(v(x), e(x)), rd(v(x), \ell(v(x), x)))]/d(v(x), e(x))$; that is, $N_{PE}^r(x, M)$ is similar to $T(\mathcal{Y}_3)$ with the given ratio. **P6** holds depending on the pair M and r. That is, there exists an $r_o := r_o(M)$ so that $N_{PE}^{r_o}(x, M)$ satisfies **P6** for all $r \leqslant r_o(M)$, and fails to satisfy otherwise. Property **P6** fails for all M when $r = \infty$, and **P8** holds only when $M = M_{CM}$. With M_{CM}-vertex regions, for all $r \in [1, \infty]$, the area $A(N_{PE}^r(x, M_{CM}))$ is a continuous function of $d(\ell_r(v(x), x), v(x))$ which is a continuous function of $d(\ell(v(x), x), v(x))$ which is a continuous function of x. Moreover, $\Lambda_0(N_{PE}^r, M) = \mathcal{Y}_3$ for all $r \in [1, \infty]$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, since the $\mathbb{R}^2$-Lebesgue measure $\lambda(N_{PE}^r(x, M)) = 0$ iff $x \in \mathcal{Y}_3$.

As for **P3**, for $T_2(x) \subseteq T(\mathcal{Y}_3)$ one can loosen the concept of center by treating the line $\ell(v(x), x)$ as the *edge-wise central line*, so **P3** is satisfied in this loose sense for $r = 2$. Notice that x is not the unique center in this sense, but a point on a central line. Let $T(M_1, M_2, M_3)$ be triangle whose vertices are the midpoints of the edges M_i for $i = 1, 2, 3$. Then for any $x \in T(M_1, M_2, M_3)$, $N_{PE}^2(x, M) = T(\mathcal{Y}_3)$, so $T(M_1, M_2, M_3) \subseteq \mathcal{R}_S(N_{PE}^2, M)$ where equality holds for $M = M_{CM}$ for all triangles. For $r = \sqrt{2}$, one can loosen the concept of center by treating the line $\ell(v(x), x)$ as the *area-wise central line* in $N_{PE}^{\sqrt{2}}(x, M)$, so **P3** is satisfied in this loose sense. Note that if x is close enough to M, it is possible to have $N_{PE}^{\sqrt{2}}(x, M) = T(\mathcal{Y}_3)$. We could also use M_I-vertex regions obtained by inner angle bisectors.

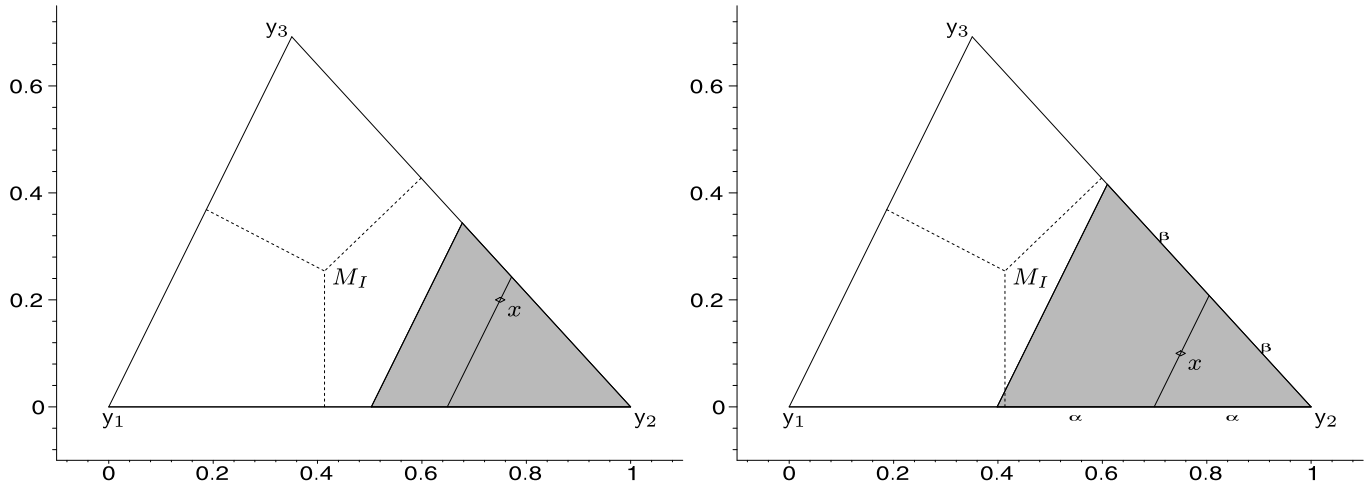

Fig. 19. Shaded regions are double-area proximity region $N_{PE}^{\sqrt{2}}(x, M_I^{\perp})$ (left) and double-edge proximity region $N_{PE}^{2}(x, M_I^{\perp})$ (right) for an $x \in R_I^{\perp}(y_2)$.

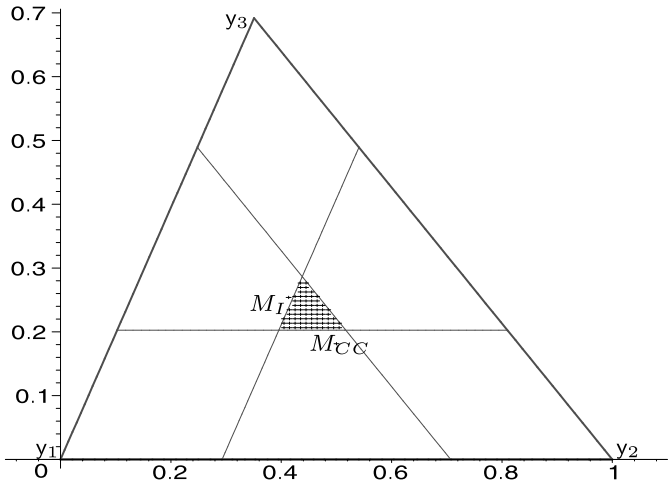

Fig. 20. The hatched region is the triangle $\mathcal{T}^r$ with $r = \sqrt{2}$.

The proximity region $N_{PE}^{r}(x, M^{\perp})$ satisfies all the properties as $N_{PE}^{r}(x, M)$, except for properties **P8** and **P9**. The property **P8** fails as the continuity does not hold with orthogonal projections even for $M = M_{CM}$; and **P9** fails since with orthogonal projections, the geometry invariance is violated.

In non-obtuse triangles, $\mathcal{R}_S^{\perp}(N_{PE}^{2}, M_{CC}) = T(M_1, M_2, M_3)$ (see Fig. 18 (left)). But, in obtuse triangles, $\mathcal{R}_S^{\perp}(N_{PE}^{2}, M_{CC}) \supsetneq T(M_1, M_2, M_3)$ and is a quadrilateral. The functional forms of the superset region, $\mathcal{R}_S^{\perp}(N_{PE}^{r}, M)$, and $T(M_1, M_2, M_3)$ in T_b are given in [2]. Moreover $T(M_1, M_2, M_3) \subseteq \mathcal{R}_S^{\perp}(N_{PE}^{2}, M_I)$ for all $T(\mathcal{Y}_3)$ (see Fig. 18 (right)) with equality holding when $T(\mathcal{Y}_3)$ is an equilateral triangle. For $N_{PE}^{2}(\cdot, M_{CM})$ constructed using the median lines $\mathcal{R}_S(N_{PE}^{2}, M_{CM}) = T(M_1, M_2, M_3)$ and for $N_{PE}^{2}(\cdot, M_{CM})$ constructed by the orthogonal projections, $\mathcal{R}_S^{\perp}(N_{PE}^{2}, M_{CM}) \supseteq T(M_1, M_2, M_3)$ with equality holding when $T(\mathcal{Y}_3)$ is an equilateral triangle.

In $T(\mathcal{Y}_3)$, drawing the lines $q_i(r, x)$ such that $d(y_i, e_i) = rd(q_i(r, x), y_i)$ for $i \in \{1, 2, 3\}$ yields a triangle, $\mathcal{T}^r$, for $r < 3/2$. See Fig. 20 for $\mathcal{T}^r$ with $r = \sqrt{2}$. The functional form of $\mathcal{T}^r$ in T_b is

$$\begin{aligned}\mathcal{T}^r &= T\big(t_1(r), t_2(r), t_3(r)\big)\\ &:= T\left(\left(\frac{(r-1)(1+c_1)}{r}, \frac{c_2(r-1)}{r}\right), \left(\frac{2-r+c_1(r-1)}{r}, \frac{c_2(r-1)}{r}\right), \left(\frac{c_1(2-r)+r-1}{r}, \frac{c_2(r-2)}{r}\right)\right). \qquad (1)\end{aligned}$$

There is a crucial difference between $\mathcal{T}^r$ and $T(M_1, M_2, M_3)$: $T(M_1, M_2, M_3) \subseteq \mathcal{R}_S(N_{PE}^{r}, M)$ for all M and $r \geqslant 2$, but $(\mathcal{T}^r)^o$ and $\mathcal{R}_S(N_{PE}^{r}, M)$ are disjoint regions for all M and r. So if $M \in (\mathcal{T}^r)^o$, then $\mathcal{R}_S(N_{PE}^{r}, M) = \emptyset$; if $M \in \partial(\mathcal{T}^r)$, then $\mathcal{R}_S(N_{PE}^{r}, M) = \{M\}$; and if $M \notin \mathcal{T}^r$, then $\mathcal{R}_S(N_{PE}^{r}, M)$ has positive area. Thus $N_{PE}^{r}(\cdot, M)$ fails to satisfy **P6** if $M \notin \mathcal{T}^r$. The same holds for $N_{PE}^{r}(\cdot, M^{\perp})$ also. The triangle $\mathcal{T}^r$ defined above plays a crucial role in the analysis of the distribution of the domination number of the proportional-edge PCD. In fact, it has been shown that for $M \in \{t_1(r), t_2(r), t_3(r)\}$, there exists a specific value of r for which the asymptotic distribution of the domination number is non-degenerate [6]. The superset region $\mathcal{R}_S(N_{PE}^{r}, M)$ will be important for both the domination number and the relative density of the corresponding PCDs.

In non-acute triangles, $M_{CC} \notin T(\mathcal{Y}_3)^o$ implies $M_{CC} \notin \mathcal{T}^r$ (since $T(\mathcal{Y}_3) \supset \mathcal{T}^r$). Let $M_{CC} = (x_{CC}, y_{CC})$. For an acute basic triangle, if $y_{CC} < \frac{c_2(\sqrt{2}-2x_{CC})}{2(1-c_1)}$ holds, then $M_{CC} \notin \mathcal{T}^{r=\sqrt{2}}$ (see Fig. 20).

Remark 3.1.

- For $r_1 \leqslant r_2$, $N_{PE}^{r_1}(x, M) \subseteq N_{PE}^{r_2}(x, M)$ for all $x \in T(\mathcal{Y}_3)$. For $r_1 < r_2$, $N_{PE}^{r_1}(x, M) \subseteq N_{PE}^{r_2}(x, M)$ with equality holding for only $x \in \mathcal{Y}_3$ or $x \in \mathcal{R}_S(N_{PE}^{\min(r_1,r_2)}, M)$.
- For $3/2 < r_1 < r_2$, we have $\mathcal{R}_S(N_{PE}^{r_1}, M) \subsetneq \mathcal{R}_S(N_{PE}^{r_2}, M)$ and for $r < 3/2$, $\mathcal{R}_S(N_{PE}^{r}, M) = \emptyset$.
- For $r_1 < r_2$, $A(N_{PE}^{r_1}(X, M)) \leqslant^{ST} A(N_{PE}^{r_2}(X, M))$ for X from a continuous distribution on $T(\mathcal{Y}_3)$ where $\leqslant^{ST}$ stands for "stochastically smaller than".

The same results hold for $N_{PE}^{r}(X, M^{\perp})$ also.

Remark 3.2. In terms of the properties stated in Section 2.1, $N_{PE}^{3/2}(\cdot, M_{CM})$ is the most appealing proximity map in the family $\mathcal{N}_{PE}^{r} := \{N_{PE}^{r}(\cdot, M)\colon\ r \in [1, \infty], M \in \mathbb{R}^2 \setminus \mathcal{Y}_3\} \cup \{N_{PE}^{r}(\cdot, M^{\perp})\colon\ r \in [1, \infty], M \in \mathbb{R}^2 \setminus \mathcal{Y}_3\}$. It is also noteworthy that the asymptotic distribution of the domination number of the PCD based on $N_{PE}^{3/2}(\cdot, M_{CM})$ is non-degenerate. Since, in general, $N_{PE}^{r}(\cdot, M)$ satisfies more of the properties compared to $N_{PE}^{r}(\cdot, M^{\perp})$, we will use the former, henceforth.

3.2.1. Extension of N_{PE}^{r} to higher dimensions

The extension to $\mathbb{R}^d$ for $d > 2$ is straightforward. The extension with $M = M_{CM}$ is given here, but the extension for general M is similar. Let $\mathcal{Y}_{d+1} = \{y_1, y_2, \ldots, y_{d+1}\}$ be $d+1$ points that do not lie on the same $(d-1)$-dimensional hyperplane. Denote the simplex formed by these $d+1$ points as $\mathcal{S}(\mathcal{Y}_{d+1})$. A simplex is the simplest polytope in $\mathbb{R}^d$ having $d+1$ vertices, $d(d+1)/2$ edges and $d+1$ faces of dimension $(d-1)$. For $r \in [1, \infty]$, define the proximity map as follows. Given a point x in $\mathcal{S}(\mathcal{Y}_{d+1})$, let $v := \operatorname{argmin}_{y \in \mathcal{Y}_{d+1}} V(Q_y(x))$ where $Q_y(x)$ is the polytope with vertices being the $d(d+1)/2$ midpoints of the edges, the vertex v and x and $V(\cdot)$ is the d-dimensional volume functional. That is, the vertex region for vertex v is the polytope with vertices given by v and the midpoints of the edges. Let $v(x)$ be the vertex in whose region x falls. If x falls on the boundary of two vertex regions, $v(x)$ is assigned arbitrarily. Let $\varphi(x)$ be the face opposite to vertex $v(x)$, and $\Upsilon(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ which contains x. Let $d(v(x), \Upsilon(v(x), x))$ be the Euclidean distance from $v(x)$ to $\Upsilon(v(x), x)$. For $r \in [1, \infty)$, let $\Upsilon_r(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ such that

$$d\big(v(x), \Upsilon_r\big(v(x), x\big)\big) = r d\big(v(x), \Upsilon\big(v(x), x\big)\big) \quad \text{and} \quad d\big(\Upsilon\big(v(x), x\big), \Upsilon_r\big(v(x), x\big)\big) < d\big(v(x), \Upsilon_r\big(v(x), x\big)\big).$$

Let $\mathcal{S}_r(x)$ be the polytope similar to and with the same orientation as $\mathcal{S}(\mathcal{Y}_{d+1})$ having $v(x)$ as a vertex and $\Upsilon_r(v(x), x)$ as the opposite face. Then the proximity region $N_{PE}^{r}(x, M_{CM}) := \mathcal{S}_r(x) \cap \mathcal{S}(\mathcal{Y}_{d+1})$. Notice that $r \geqslant 1$ implies $x \in N_{PE}^{r}(x, M_{CM})$.

3.3. Central similarity proximity maps

For the expansion parameter $\tau \in (0, 1]$, define $N_{CS}^{\tau}(\cdot, M)$ to be the *central similarity proximity map* with M-edge regions as follows; see also Fig. 21 with $M = M_{CM}$. For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $e(x)$ be the edge in whose region x falls; i.e., $x \in R_M(e(x))$. If x falls on the boundary of two edge regions, $e(x)$ is assigned to x arbitrarily. For $\tau \in (0, 1]$, the central similarity proximity region $N_{CS}^{\tau}(x, M)$ is defined to be the triangle $\mathcal{T}_\tau(x)$ with the following properties:

(i) The triangle $\mathcal{T}_\tau(x)$ has edges $e_i^{\tau}(x)$ parallel to e_i for $i \in \{1, 2, 3\}$, and for $x \in R_M(e(x))$, $d(x, e^{\tau}(x)) = \tau d(x, e(x))$ and $d(e^{\tau}(x), e(x)) \leqslant d(x, e(x))$ where $d(x, e(x))$ is the Euclidean distance from x to $e(x)$;
(ii) The triangle $\mathcal{T}_\tau(x)$ has the same orientation as and is similar to $T(\mathcal{Y}_3)$;
(iii) The point x is the same type of center of $\mathcal{T}_\tau(x)$ as M is of $T(\mathcal{Y}_3)$.

Note that (i) implies the parametrization of the PCD, (ii) explains "similarity", and (iii) explains "central" in the name, *central similarity proximity map*. For $\tau = 0$, we let $N_{CS}^{\tau=0}(x, M) := \{x\}$ for all $x \in T(\mathcal{Y}_3)$. For $x \in \partial(T(\mathcal{Y}_3))$, we have $N_{CS}^{\tau}(x, M) := \{x\}$ for all $\tau \in [0, 1]$.

By definition $x \in N_{CS}^{\tau}(x, M)$ for all $x \in T(\mathcal{Y}_3)$. Furthermore, $\tau \leqslant 1$ implies that $N_{CS}^{\tau}(x, M) \subseteq T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3)$ and $M \in T(\mathcal{Y}_3)^o$. For all $x \in T(\mathcal{Y}_3)^o \cap R_M(e(x))$, the edges $e^{\tau}(x)$ and $e(x)$ are coincident iff $\tau = 1$. See Fig. 22 for the arcs based on $N_{CS}^{\tau=1}(x, M_{CM})$ for 20 class $\mathcal{X}$ points in the one triangle case.

Notice that $X_i \overset{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional pdf f exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$, implies that the special case in the construction of $N_{CS}^{\tau}(\cdot) - X$ falls on the boundary of two edge regions – occurs with probability zero. Note that for such an F, $N_{CS}^{\tau}(X, M)$ is a triangle for $\tau > 0$ a.s. The central similarity proximity maps are defined with M-edge regions for $M \in T(\mathcal{Y}_3)^o$. Among the four centers considered in Section 2.4.1, M_{CM} and M_I are always inside the triangle, so they can be used in construction of the central similarity proximity map.

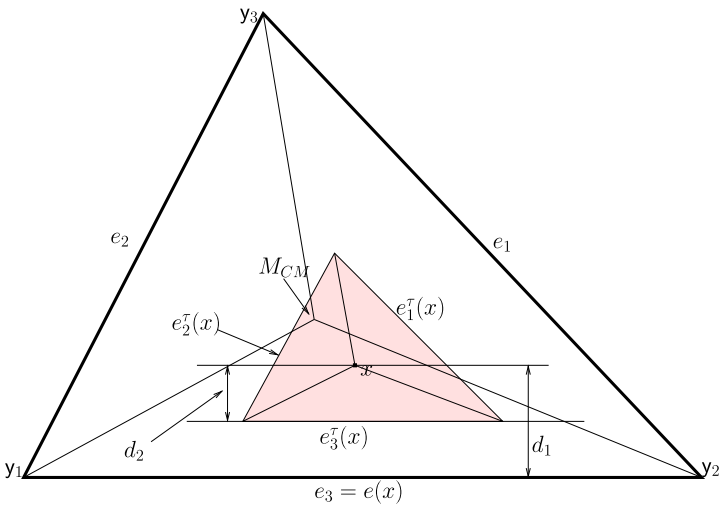

Fig. 21. Construction of central similarity proximity region, $N_{CS}^{\tau=1/2}(x, M_{CM})$ (shaded region) where $d_2 = d(x, e_3^\tau(x)) = \frac{1}{2}d(x, e(x))$ and $d_1 = d(x, e(x))$.

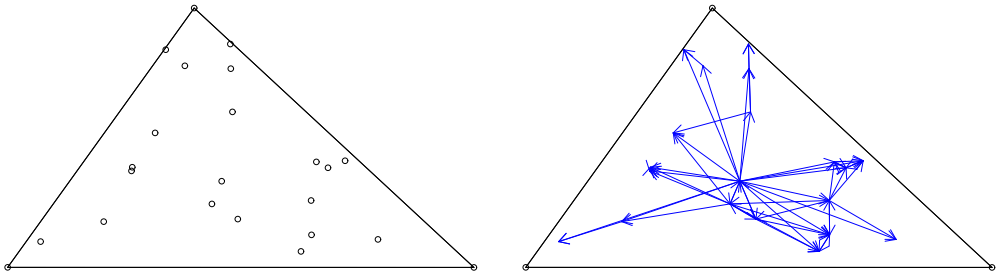

Fig. 22. A realization of 20 class $\mathcal{X}$ points (circles) generated iid $\mathcal{U}(T(\mathcal{Y}_3))$ (left) and the corresponding arcs for $N_{CS}^{\tau=1}(x, M_{CM})$ (right).

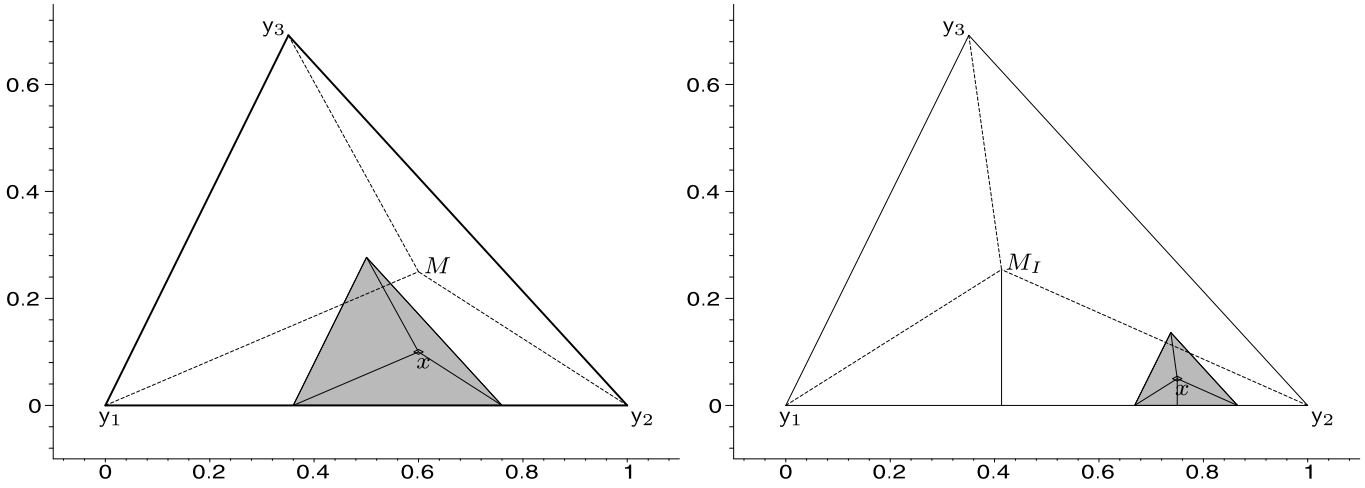

Fig. 23. The regions $N_{CS}^{\tau=1}(x, M)$ for an $x \in R_M(e_3)$ (left); and $N_{CS}^{\tau=1}(x, M_I)$ for an $x \in R_{M_I}(e_3)$ (right).

With $M = M_{CM}$, for $x \in R_{CM}(e)$, the similarity ratio of $N_{CS}^\tau(x, M_{CM})$ to $T(\mathcal{Y}_3)$ is $d(x, e_\tau(x))/d(M_{CM}, e)$. See Fig. 21 for $N_{CS}^{\tau=1/2}(x, M_{CM})$ with $e = e_3$. The functional form of $N_{CS}^\tau(x, M_{CM})$ for an $x = (x_0, y_0) \in R_{CM}(e)$ is provided in [2]. The proximity regions $N_{CS}^\tau(x, M_I)$ are also described in detail in [2]. In general, for central similarity proximity regions with M-edge regions, the similarity ratio of $N_{CS}^\tau(x, M)$ to $T(\mathcal{Y}_3)$ is $d(x, e^\tau(x))/d(M, e(x))$. See Fig. 23 (left) for $N_{CS}^{\tau=1}(x, M)$ and (right) $N_{CS}^{\tau=1}(x, M_I)$ with $e(x) = e_3$. Notice that $N_{CS}^\tau(\cdot, M)$ satisfies properties **P1**–**P9**. Furthermore, $\Lambda_0(N_{CS}^\tau(\cdot, M)) = \partial(T(\mathcal{Y}_3))$ for all $\tau \in (0, 1]$ and $\Lambda_0(N_{CS}^{\tau=0}(\cdot, M)) = T(\mathcal{Y}_3)$, since $\lambda(N_{CS}^\tau(x)) = 0$ iff $x \in e_i$ for $i \in \{1, 2, 3\}$ or $\tau = 0$.

Remark 3.3. For acute triangles we could use M_{CC}- or M_O-edge regions in central similarity proximity regions which will also satisfy properties **P1**–**P9**. But for obtuse triangles, **P2** is not satisfied and edge regions are not defined in a natural manner.

Remark 3.4. In the family $\mathcal{N}_{CS}^{\tau} := \{N_{CS}^{\tau}(\cdot, M):\ \tau \in [0,1], M \in T(\mathcal{Y}_3)^o\}$, every $N_{CS}^{\tau}(\cdot, M)$ with $\tau \in [0,1]$ satisfies all the properties in Section 2.1. Furthermore,

- For $\tau_1 \leqslant \tau_2$, $N_{CS}^{\tau_1}(x, M) \subseteq N_{CS}^{\tau_2}(x, M)$ for all $x \in T(\mathcal{Y}_3)$. For $\tau_1 < \tau_2$, $N_{CS}^{\tau_1}(x, M) \subseteq N_{CS}^{\tau_2}(x, M)$ with equality holding only for $x \in \partial(T(\mathcal{Y}_3))$.
- The superset region $\mathcal{R}_S(N_{CS}^{\tau}, M) = \emptyset$ for $\tau \in [0,1)$ and $\mathcal{R}_S(N_{CS}^{\tau=1}, M) = \{M\}$.
- For $\tau_1 < \tau_2$, $A(N_{CS}^{\tau_1}(X, M)) \leqslant^{ST} A(N_{CS}^{\tau_2}(X, M))$ for X from a continuous distribution on $T(\mathcal{Y}_3)$.

3.3.1. Extension of N_{CS}^{τ} to higher dimensions

The extension of N_{CS}^{τ} to $\mathbb{R}^d$ for $d > 2$ is straightforward. The extension for $M = M_{CM}$ is described, the extension for general M is similar. Let $\mathcal{Y}_{d+1} = \{\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_{d+1}\}$ be $d+1$ points that do not lie on the same $(d-1)$-dimensional hyperplane. For $\tau \in (0,1]$, define the central similarity proximity map as follows. Let φ_i be the face opposite vertex y_i for $i \in \{1, 2, \ldots, (d+1)\}$, and "face regions" $R_{CM}(\varphi_1), R_{CM}(\varphi_2), \ldots, R_{CM}(\varphi_{d+1})$ partition $\mathfrak{S}(\mathcal{Y}_{d+1})$ into $d+1$ regions, namely the $d+1$ polytopes with vertices being the center of mass together with d vertices chosen from $d+1$ vertices. For $x \in \mathfrak{S}(\mathcal{Y}_{d+1}) \setminus \mathcal{Y}_{d+1}$, let $\varphi(x)$ be the face in whose region x falls; $x \in R(\varphi(x))$. If x falls on the boundary of two face regions, $\varphi(x)$ is assigned arbitrarily. For $\tau \in (0,1]$, the central similarity proximity region $N_{CS}^{\tau}(x, M_{CM}) = \mathbb{S}_{\tau}(x)$ is defined to be the simplex $\mathbb{S}_{\tau}(x)$ with the following properties:

(i) The region $\mathbb{S}_{\tau}(x)$ has faces $\varphi_i^{\tau}(x)$ parallel to $\varphi_i(x)$ for $i \in \{1, 2, \ldots, (d+1)\}$, and for $x \in R_{CM}(\varphi(x))$, $\tau d(x, \varphi(x)) = d(\varphi^{\tau}(x), x)$ where $d(x, \varphi(x))$ is the Euclidean distance from x to $\varphi(x)$;
(ii) The region $\mathbb{S}_{\tau}(x)$ has the same orientation as and similar to $\mathfrak{S}(\mathcal{Y}_{d+1})$;
(iii) The point x is the center of mass of $\mathbb{S}_{\tau}(x)$, as M_{CM} is of $\mathfrak{S}(\mathcal{Y}_{d+1})$. Note that $\tau > 1$ implies that $x \in N_{CS}^{\tau}(x)$.

3.4. The behavior of proximity regions

Let $N(\cdot)$ be any proximity map defined on the measurable space Ω with measure μ, and let $\{x_n\}_{n=1}^{\infty}$ be a sequence of points in Ω. We say $N(x_n)$ *gets larger* as n increases if $N(x_n) \subseteq N(x_m)$ for $m \geqslant n$, and $N(x_n)$ *gets strictly larger* if $N(x_n) \subsetneq N(x_m)$ for $m > n$.

In the following theorems we will assume $\Omega = \mathbb{R}^2$ with μ being the $\mathbb{R}^2$-Lebesgue measure λ and M-vertex regions are defined with points$M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$.

Theorem 3.5. *For $N(\cdot) \in \mathcal{N}_{AS}$, as $d(x, \mathrm{y})$ (strictly) increases with x lying on a ray from y in $R_M(\mathrm{y}) \setminus \mathcal{R}_S(N, M)$, $N(x)$ gets (strictly) larger.*

Proof. Let $N(\cdot) \in \mathcal{N}_{AS}$. For x, y lying on a ray from y in $R_M(\mathrm{y}) \setminus \mathcal{R}_S(N, M)$, if $d(x, \mathrm{y}) \leqslant d(y, \mathrm{y})$, then $B(x, r(x)) \subseteq B(y, r(y))$, which implies $N(x) \subseteq N(y)$, hence $N(x)$ gets larger as $d(x, \mathrm{y})$ increases for x lying on a ray from y in $R_M(\mathrm{y}) \setminus \mathcal{R}_S(N, M)$. The strict version follows similarly. If $x, y \in R_M(\mathrm{y}) \cap \mathcal{R}_S(N, M)$, then $N(x) = N(y) = T(\mathcal{Y}_3)$. □

Let $\ell(\mathrm{y}, x)$ be the line at x parallel to $e(x)$ for $x \in R_M(\mathrm{y})$ where $e(x)$ is the edge opposite vertex y.

Theorem 3.6. *For $N(\cdot) \in \mathcal{N}_{PE}^{r}$, as $d(\ell(\mathrm{y}, x), \mathrm{y})$ (strictly) increases with $x \in R_M(\mathrm{y}) \setminus \mathcal{R}_S(N, M)$, $N(x)$ gets (strictly) larger for $r < \infty$.*

Proof. Let $N(\cdot) \in \mathcal{N}_{PE}^{r}$. For $x, y \in R_M(\mathrm{y}) \setminus \mathcal{R}_S(N, M)$, if $d(\ell(\mathrm{y}, x), \mathrm{y}) \leqslant d(\ell(\mathrm{y}, y), \mathrm{y})$, then by definition $N(x) \subseteq N(y)$, hence the result follows. The strict version follows similarly. If $x, y \in R_M(\mathrm{y}) \cap \mathcal{R}_S(N, M)$, then $N(x) = N(y) = T(\mathcal{Y}_3)$, and if $r = \infty$ and $x, y \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, $N(x) = N(y) = T(\mathcal{Y}_3)$. □

Note that as $d(\ell(\mathrm{y}, x), \mathrm{y})$ increases for $x \in R_M(\mathrm{y})$, $d(\ell(\mathrm{y}, x), M)$ decreases, provided that $M \in T(\mathcal{Y}_3)^o$ and M-vertex regions are convex.

Theorem 3.7. *For $N(\cdot) \in \mathcal{N}_{CS}^{\tau}$ with $\tau \in (0,1]$, as $d(x, e)$ (strictly) increases with $x \in R_M(e)$, the area $A(N_{CS}^{\tau}(x, M))$ (strictly) increases.*

Proof. Let $N(\cdot) \in \mathcal{N}_{CS}^{\tau}$ with $\tau \in (0,1]$. For $x, y \in R_M(e)$ and $\tau \in (0,1]$, if $d(x, e) \leqslant d(y, e)$ then the similarity ratio of $N(y)$ to $T(\mathcal{Y}_3)$ is larger than or equal to that of $N(x)$, which in turn implies that $A(N(x)) \leqslant A(N(y))$. The strict version follows similarly. □

Observe that the statement of Theorem 3.7 is about the area $A(N_{CS}^{\tau}(x, M))$. We need further restrictions for $N_{CS}^{\tau}(x, M)$ to get larger.

Theorem 3.8. *Let $\ell_M(\mathrm{y})$ be the line joining M and vertex $\mathrm{y} \in \mathcal{Y}_3$ and let $N(\cdot) \in \mathcal{N}_{CS}^{\tau}$ with $\tau \in (0,1]$. As $d(x, \ell_M(\mathrm{y}_j))$ and $d(x, \ell_M(\mathrm{y}_k))$ both (strictly) decrease for $x \in R_M(e_l)$ where j, k, l are distinct, $N(x)$ (strictly) increases.*

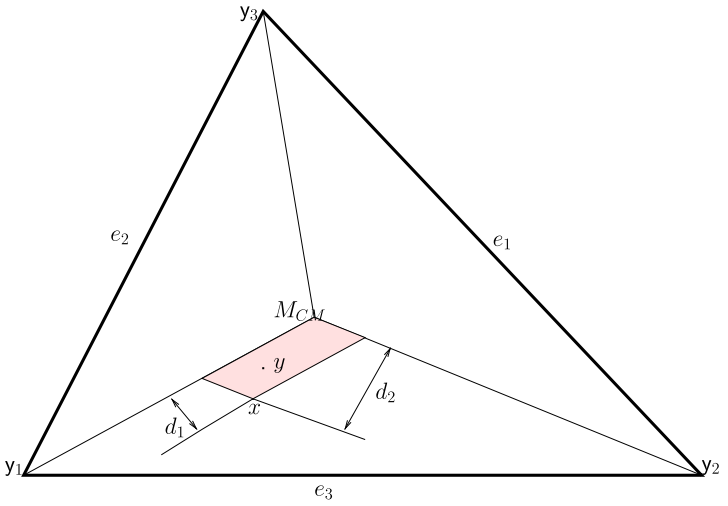

Fig. 24. The figure for $x, y \in R_M(e_3)$ described in Theorem 3.8. Here $d_1 = d(x, \ell_{M_C}(\mathsf{y}_1))$ and $d_2 = d(x, \ell_{M_C}(\mathsf{y}_2))$.

Proof. Suppose, without loss of generality, that $x, y \in R_M(e_3)$. Consider the set

$$S(e_3, x) := \big\{ y \in R_M(e_3) \colon d\big(y, \ell_M(\mathsf{y}_1)\big) \leqslant d\big(x, \ell_M(\mathsf{y}_1)\big) \quad \text{and} \quad d\big(y, \ell_M(\mathsf{y}_2)\big) \leqslant d\big(x, \ell_M(\mathsf{y}_2)\big) \big\},$$

which is a parallelogram. See Fig. 24 for an example of $S(e_3, x)$ with $M = M_{CM}$ and $e = e_3$. Given x, for $y \in S(e_3, x)$, by construction, $N(x) \subseteq N(y)$. Then the desired result follows for $\tau \in (0, 1]$. Observe that if x_{n+1} is in $S(e_3, x_n)$, then $d(x_n, \ell_M(\mathsf{y}_1))$ and $d(x_n, \ell_M(\mathsf{y}_2))$ both decrease. The strict version follows similarly. □

Remark 3.9. For $\mathcal{R}_S(N_{\mathcal{Y}})$ with positive area, by definition, as $x \to y \in \mathcal{R}_S(N_{\mathcal{Y}})$, $N_{\mathcal{Y}}(x) \to T(\mathcal{Y}_3)$ and hence $\operatorname{argsup}_{x \in T(\mathcal{Y}_3)} A(N_{\mathcal{Y}}(x)) \in \mathcal{R}_S(N_{\mathcal{Y}})$ with $\sup_{x \in T(\mathcal{Y}_3)} A(N_{\mathcal{Y}}(x)) = A(T(\mathcal{Y}_3))$. Furthermore, the following also hold.

- Let $N(\cdot) \in \mathcal{N}_{AS}$. Then as $x \to M_{CC}$ in a non-obtuse triangle $T(\mathcal{Y}_3)$, we have $N(x) \to T(\mathcal{Y}_3)$ and
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = M_{CC} \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = A\big(T(\mathcal{Y}_3)\big).$$
- Let $N(\cdot) \in \mathcal{N}_{PE}^r$. Then for $r > 3/2$, as $x \to y \in \mathcal{R}_S(N_{PE}^r, M)$, we have $N(x) \to T(\mathcal{Y}_3)$ hence
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) \in \mathcal{R}_S(N, M) \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = A\big(T(\mathcal{Y}_3)\big).$$
For $r < 3/2$, if $M \notin \mathcal{T}^r$, then as $x \to M$, we have $N(x) \to T(\mathcal{Y}_3)$ and
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = M \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = A\big(T(\mathcal{Y}_3)\big).$$
If $M \in (\mathcal{T}^r)^o$, then as $x \to M$, we have $N(x) \to N(M) \subsetneq T(\mathcal{Y}_3)$, but still
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = M \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = A\big(N(M)\big).$$
If $M \in \partial(\mathcal{T}^r)$, then as $x \to M$, $N(x) \to N(M) \subseteq T(\mathcal{Y}_3)$, and
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = M \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N(x)\big) = A\big(N(M)\big)$$
which might be $T(\mathcal{Y}_3)$ or a proper subset of $T(\mathcal{Y}_3)$.
- As $x \to M_{CM}$, $N_{PE}^{3/2}(x, M_{CM}) \to T(\mathcal{Y}_3)$ and
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N_{PE}^{3/2}(x, M_{CM})\big) = M_{CM} \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N_{PE}^{3/2}(x, M_{CM})\big) = A\big(T(\mathcal{Y}_3)\big).$$
- As $x \to M$, we have $N_{CS}^{\tau=1}(x, M) \to T(\mathcal{Y}_3)$ and
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N_{CS}^{\tau=1}(x, M)\big) = M \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N_{CS}^{\tau=1}(x, M)\big) = A\big(T(\mathcal{Y}_3)\big).$$
For $\tau \in (0, 1)$, as $x \to M$, $N_{CS}^{\tau}(x, M) \to N_{CS}^{\tau}(M, M)$ and
$$\operatorname*{argsup}_{x \in T(\mathcal{Y}_3)} A\big(N_{CS}^{\tau}(x, M)\big) = M \quad \text{with} \quad \sup_{x \in T(\mathcal{Y}_3)} A\big(N_{CS}^{\tau}(x, M)\big) = A\big(N_{CS}^{\tau}(M, M)\big).$$

Although the comments in the above remark mostly follow by definition, they will be indicative of whether the asymptotic distribution of the domination number of the associated PCD is degenerate or not.

4. Relative density and domination number of PCDs

4.1. Relative density

The *relative density* of a digraph $D=(\mathcal{V},\mathcal{A})$ of order $|\mathcal{V}|=n$, denoted as $\rho(D)$, is defined as

$$\rho(D)=\frac{|\mathcal{A}|}{n(n-1)}$$

where $|\cdot|$ denotes the cardinality of sets [13]. Thus $\rho(D)$ represents the ratio of the number of arcs in the digraph D to the number of arcs in the complete symmetric digraph of order n, which is $n(n-1)$.

If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$ the relative density of the associated data-random PCD D, denoted as ρ_n for brevity, is a U-statistic,

$$\rho_n=\frac{1}{n(n-1)}\sum\sum_{i<j} h_{ij} \tag{2}$$

where

$$h_{ij}=\mathbf{I}\big\{(X_i,X_j)\in\mathcal{A}\big\}+\mathbf{I}\big\{(X_j,X_i)\in\mathcal{A}\big\}=\mathbf{I}\big\{X_j\in N_{\mathcal{Y}}(X_i)\big\}+\mathbf{I}\big\{X_i\in N_{\mathcal{Y}}(X_j)\big\}, \tag{3}$$

where $\mathbf{I}(\cdot)$ is the indicator function. Since the digraph is asymmetric, h_{ij} is defined as the number of arcs in D between vertices X_i and X_j, in order to produce a symmetric kernel with finite variance [16].

The random variable ρ_n depends on n and $N_{\mathcal{Y}}$ explicitly and on F implicitly. The expectation $\mathbf{E}[\rho_n]$, however, is independent of n and depends on only F and $N_{\mathcal{Y}}$:

$$0\leqslant \mathbf{E}[\rho_n]=\frac{1}{2}\mathbf{E}[h_{12}]\leqslant 1 \quad \text{for all } n\geqslant 2. \tag{4}$$

The variance $\mathbf{Var}[\rho_n]$ simplifies to

$$0\leqslant \mathbf{Var}[\rho_n]=\frac{1}{2n(n-1)}\mathbf{Var}[h_{12}]+\frac{n-2}{n(n-1)}\mathbf{Cov}[h_{12},h_{13}]\leqslant 1/4. \tag{5}$$

A central limit theorem for U-statistics [16] yields

$$\sqrt{n}\big(\rho_n-\mathbf{E}[\rho_n]\big)\xrightarrow{\mathcal{L}}\mathcal{N}\big(0,\mathbf{Cov}[h_{12},h_{13}]\big) \tag{6}$$

provided $\mathbf{Cov}[h_{12},h_{13}]>0$. The asymptotic variance of ρ_n, $\mathbf{Cov}[h_{12},h_{13}]$, depends on only F and $N_{\mathcal{Y}}$. Thus, we need determine only $\mathbf{E}[h_{12}]$ and $\mathbf{Cov}[h_{12},h_{13}]$ in order to obtain the normal approximation

$$\rho_n \overset{\text{approx}}{\sim}\mathcal{N}\big(\mathbf{E}[\rho_n],\mathbf{Var}[\rho_n]\big)=\mathcal{N}\left(\frac{\mathbf{E}[h_{12}]}{2},\frac{\mathbf{Cov}[h_{12},h_{13}]}{n}\right) \quad \text{for large } n. \tag{7}$$

4.1.1. Asymptotic distribution of relative density of PCDs

By detailed geometric probability calculations, provided in [8] and [7] the mean and the asymptotic variance of the relative density of the proportional-edge and central similarity PCDs can explicitly be computed. The central limit theorem for U-statistics then establishes the asymptotic normality under the uniform null hypothesis. These results are summarized in the following theorems.

Theorem 4.1. *Let $\rho_n(N_{PE}^r)$ be the relative density of the proportional-edge PCD with parameter r and $M=M_{CM}$ based on a random sample of $\mathcal{X}_n$ from $\mathcal{U}(T(\mathcal{Y}_3))$ and $p_a(N_{PE}^r)$ be the corresponding arc probability, and $v(N_{PE}^r)$ be the $\mathbf{Cov}(h_{12},h_{13})$. Then for $r\in[1,\infty)$,*

$$\frac{\sqrt{n}(\rho_n(N_{PE}^r)-p_a(N_{PE}^r))}{\sqrt{v(N_{PE}^r)}}\xrightarrow{\mathcal{L}}\mathcal{N}(0,1) \tag{8}$$

where

$$p_a\big(N_{PE}^r\big)=\begin{cases}\frac{37}{216}r^2 & \text{for } r\in[1,3/2),\\ -\frac{1}{8}r^2+4-8r^{-1}+\frac{9}{2}r^{-2} & \text{for } r\in[3/2,2),\\ 1-\frac{3}{2}r^{-2} & \text{for } r\in[2,\infty),\end{cases} \tag{9}$$

and

$$\nu\big(N_{PE}^r\big) = \nu_1(r)\mathbf{I}\big(r \in [1, 4/3)\big) + \nu_2(r)\mathbf{I}\big(r \in [4/3, 3/2)\big) + \nu_3(r)\mathbf{I}\big(r \in [3/2, 2)\big) + \nu_4(r)\mathbf{I}\big(r \in [2, \infty]\big) \tag{10}$$

with

$$\begin{aligned}
\nu_1(r) &= \frac{3007r^{10} - 13824r^9 + 898r^8 + 77760r^7 - 117953r^6 + 48888r^5 - 24246r^4 + 60480r^3 - 38880r^2 + 3888}{58320r^4},\\
\nu_2(r) &= \frac{5467r^{10} - 37800r^9 + 61912r^8 + 46588r^6 - 191520r^5 + 13608r^4 + 241920r^3 - 155520r^2 + 15552}{233280r^4},\\
\nu_3(r) &= -\big[7r^{12} - 72r^{11} + 312r^{10} - 5332r^8 + 15072r^7 + 13704r^6 - 139264r^5 + 273600r^4 - 242176r^3\\
&\quad + 103232r^2 - 27648r + 8640\big]/\big[960r^6\big],\\
\nu_4(r) &= \frac{15r^4 - 11r^2 - 48r + 25}{15r^6}.
\end{aligned}$$

For $r = \infty$, $\rho_n(N_{PE}^r)$ is degenerate.

Theorem 4.2. *Let $\rho_n(N_{CS}^\tau)$ be the relative density of the central similarity PCD with parameter τ and $M = M_{CM}$ based on a random sample of $\mathcal{X}_n$ from $\mathcal{U}(T(\mathcal{Y}_3))$ and $p_a(N_{CS}^\tau)$ be the corresponding arc probability, and $\nu(N_{CS}^\tau)$ be the* **Cov**(h_{12}, h_{13})*. Then for $\tau \in (0, 1]$, the relative density of the central similarity proximity digraph converges in law to the normal distribution; i.e., as $n \to \infty$,*

$$\frac{\sqrt{n}(\rho_n(N_{CS}^\tau) - p_a(N_{CS}^\tau))}{\sqrt{\nu(N_{CS}^\tau)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \tag{11}$$

where

$$p_a\big(N_{CS}^\tau\big) = \tau^2/6 \quad \textit{and} \quad \nu\big(N_{CS}^\tau\big) = \frac{\tau^4(6\tau^5 - 3\tau^4 - 25\tau^3 + \tau^2 + 49\tau + 14)}{45(\tau + 1)(2\tau + 1)(\tau + 2)}. \tag{12}$$

For $\tau = 0$, $\rho_n(N_{CS}^\tau)$ is degenerate for all $n > 1$.

4.2. Domination number of the PCDs

In a digraph $D = (\mathcal{V}, \mathcal{A})$, recall that the domination number $\gamma(D)$ is the cardinality of the minimum dominating set. If a minimum dominating set is of size one, we call it a *dominating point*. Note that for $|\mathcal{V}| = n > 0$, $1 \leqslant \gamma(D) \leqslant n$, since $\mathcal{V}$ itself is always a dominating set.

4.2.1. Asymptotic distribution of domination number of the PCDs

Recall the triangle $\mathcal{T}^r$ defined in Eq. (1) (see also Fig. 20 for $\mathcal{T}^r$ with $r = \sqrt{2}$). Let $\gamma_n(r, M)$ be the domination number of the PCD based on $N_{PE}^r(\cdot, M)$ with $\mathcal{X}_n$, a set of iid random variables from $\mathcal{U}(T(\mathcal{Y}_3))$, with M-vertex regions.

The domination number $\gamma_n(r, M)$ of the PCD has the following asymptotic distribution [6]. As $n \to \infty$,

$$\gamma_n(r, M) \xrightarrow{\mathcal{L}} \begin{cases} 2 + \text{BER}(1 - \pi_r) & \text{for } r \in [1, 3/2) \text{ and } M \in \{t_1(r), t_2(r), t_3(r)\},\\ 1 & \text{for } r > 3/2 \text{ and } M \in T(\mathcal{Y}_3)^o,\\ 3 & \text{for } r \in [1, 3/2) \text{ and } M \in \mathcal{T}^r \setminus \{t_1(r), t_2(r), t_3(r)\}, \end{cases} \tag{13}$$

where $\xrightarrow{\mathcal{L}}$ stands for "convergence in law" and BER(p) stands for Bernoulli distribution with probability of success p, $\mathcal{T}^r$ and $t_i(r)$ are defined in Eq. (1), and for $r \in [1, 3/2)$ and $M \in \{t_1(r), t_2(r), t_3(r)\}$,

$$\pi_r = \int_0^\infty \int_0^\infty \frac{64r^2}{9(r-1)^2} w_1 w_3 \exp\left(\frac{4r}{3(r-1)}\big(w_1^2 + w_3^2 + 2r(r-1)w_1 w_3\big)\right) dw_3 w_1. \tag{14}$$

For example, for $r = 5/4$ and $M \in \{t_1(r) = (3/10, \sqrt{3}/10), t_2(r) = (7/10, \sqrt{3}/10), t_3(r) = (1/2, 3\sqrt{3}/5)\}$, $\pi_r \approx 0.6514$. See Fig. 25 for the plot of the values computed by numerical integration of π_r as a function of r. Notice that in the non-degenerate case in (13), $\mathbf{E}[\gamma_n(r, M)] = 3 - \pi_r$ and $\mathbf{Var}[\gamma_n(r, M)] = \pi_r(1 - \pi_r)$. For $r = 3/2$ and $M = M_{CM} = (1/2, \sqrt{3}/6)$, we have $\pi_r \approx 0.7413$, which is computed differently from that in Eq. (14); see [5] for its computation.

The distribution of the domination number of N_{CS}^τ is still an open problem.

The results in Theorem 2.4 and Corollaries 2.10 and 2.11 also hold for relative density and the domination number of PCDs based on $N_{\mathcal{Y}}$. That is, we have the following corollary.

Corollary 4.3. *Given any triangle T_o and $\mathcal{X}_n$ a random sample from $\mathcal{U}(T_o)$. Suppose the PCD, D_o is defined in such a way that the ratio of the area of $N(x)$ to the area of the triangle T_o is preserved under the uniformity preserving transformation, then the distributions of the relative density and the domination number of D_o are geometry invariant.*

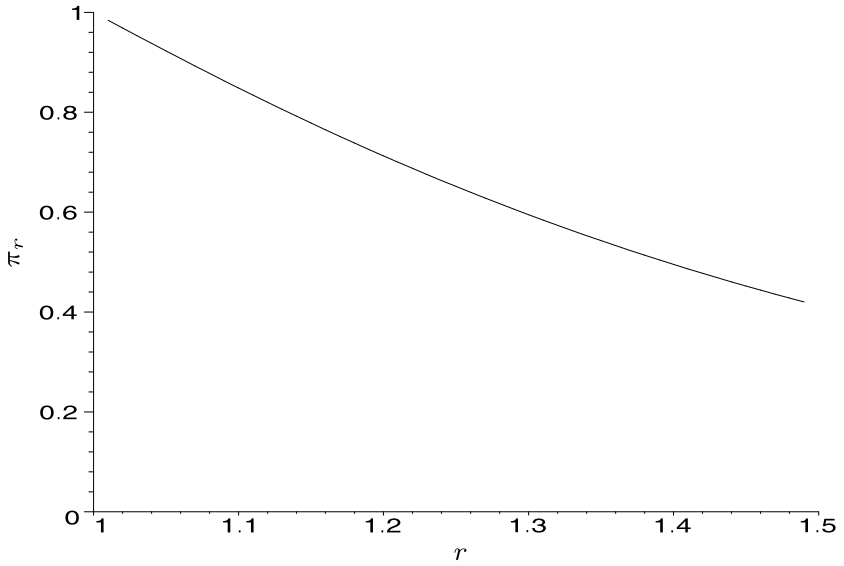

Fig. 25. The probability $\pi_r = \lim_{n\to\infty} P(\gamma_n(r, M) = 2)$ given in Eq. (14) as a function of r for $r \in [1, 3/2)$ and $M \in \{t_1(r), t_2(r), t_3(r)\}$.

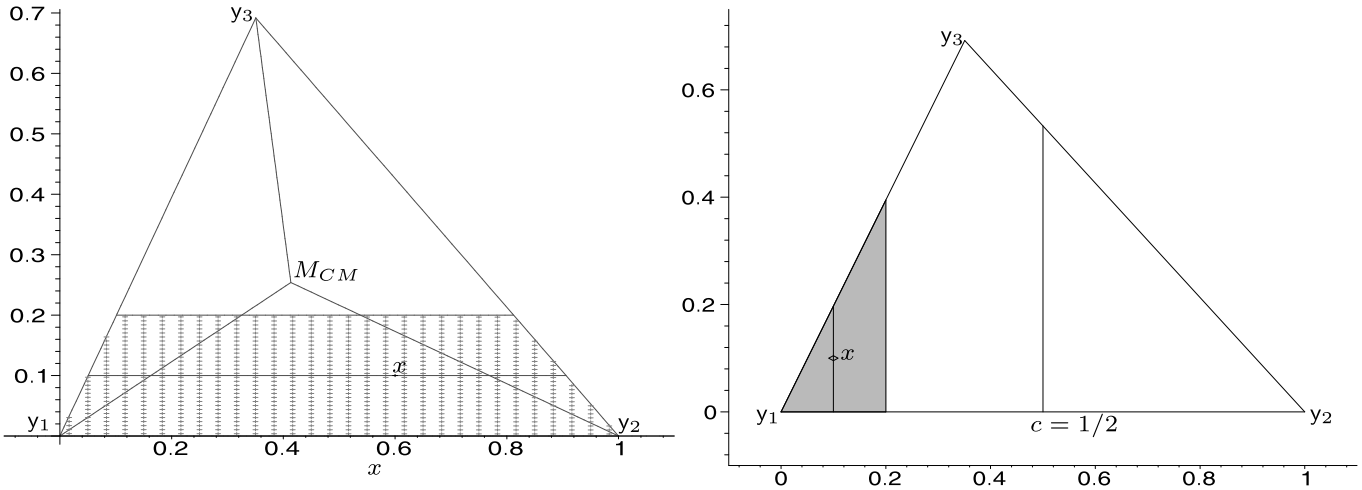

Fig. 26. An example of directional-doubling proximity region with $M = M_{CM}$ (left) and double-X proximity region with $c = 1/2$ (right).

5. Two new proximity maps

In this section, we introduce two new proximity maps and investigate their properties.

5.1. Directional-doubling proximity maps

Without loss of generality, we can assume that $T(\mathcal{Y}_3) = T_b$. Partition the triangle T_b by M-edge regions to obtain $R_M(e_i)$ for $i = 1, 2, 3$. For $z \in R_M(e_i)$, directional-doubling proximity map is defined as

$$N_{DD}(z, M) := \{u \in T_b\colon d(u, e_i) \leqslant 2d(z, e_i)\}.$$

See Fig. 26 (left) with $M = M_{CM}$. If $z \in e_i$, then $N_{DD}(z, M) := e_i$. Notice that if $z \notin e_i$, then $N_{DD}(z, M)$ is a quadrilateral. Among the properties, **P1** and **P2** follow trivially. The line at $z \in R_M(e_i)$ parallel to e_i divides the region into two pieces (half-way in the perpendicular direction to e_i) so **P3** holds in this special sense. Properties **P4** and **P5** both fail, since $N_{DD}(z, M)$ is a quadrilateral. Property **P6** holds if $M \in T(M_1, M_2, M_3)$; otherwise it fails, since $\mathcal{R}_S(N_{DD}, M)$ will have positive area. Property **P7** follows by definition. However, **P8** holds only when $M = M_{CM}$.

Property **P9** follows for N_{DD}, since $N_{DD}(z, M)$ is constructed with the boundary of $T(\mathcal{Y}_3)$ and parallel lines to the edges and by Corollary 2.10, geometry invariance for uniform data follows. That is, the distributions of relative density and the domination number of the corresponding PCD do not depend on the geometry of the triangle $T(\mathcal{Y}_3)$. Hence, it suffices to compute them for the standard equilateral triangle only. Furthermore, we have $\Lambda_0(N_{DD}) = \partial(T(\mathcal{Y}_3))$, since $N_{DD}(x, M)$ has zero area iff $x \in \partial(T(\mathcal{Y}_3))$.

Table 1
The table for the six proximity region families we consider in $\mathbb{R}^d$ with $d > 1$ with respect to the nine properties **P1**-**P9** defined in Section 2.1. $+$ ($-$): the property is (not) satisfied for all parameters of the corresponding proximity region. The symbol $-$ with a numbered superscript means the property is satisfied only for some of the parameters. *: $N_S(\cdot)$ satisfies all the properties in $\mathbb{R}$. **: The left (and right) column is for the proximity region that is based on vertex regions constructed as in Method I (and II). [1]: (property) holds for $M = M_{CC}$ only; [2]: holds for $r = 2$ and $r = \sqrt{2}$ in the loose sense only as described on page 736; [3]: holds for $r \leqslant r_o(M)$ for $r_o(M)$ described on page 736 only; [4]: holds for $M = M_{CM}$ only; [5] and [6]: holds in the special sense described on pages 744 and 745, respectively; [7]: holds for $M \in T(M_1, M_2, M_3)$ only; [8]: holds for $c = 1/2$ only; [9]: holds when $A(R_1(c)) = A(R_2(c))$ only.

Property	Proximity regions							
	$N_S(\cdot)^*$	$N_{AS}(\cdot, M)^{**}$		$N^r_{PE}(\cdot, M)^{**}$		$N^\tau_{CS}(\cdot, M)$	$N_{DD}(\cdot, M)$	$N_{DX}(\cdot, c)$
P1	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
P2	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
P3	$+$	$-$	$-$	$-^{2}$	$-^{2}$	$+$	$-^{5}$	$-^{6}$
P4	$-$	$-$	$-$	$+$	$+$	$+$	$-$	$-$
P5	$-$	$-$	$-$	$+$	$+$	$+$	$-$	$-$
P6	$-$	$-$	$-^{1}$	$-^{3}$	$-^{3}$	$+$	$-^{7}$	$-^{8}$
P7	$-$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
P8	$+$	$-$	$-^{1}$	$-^{4}$	$-$	$+$	$-^{4}$	$-^{9}$
P9	$-$	$-$	$-$	$+$	$-$	$+$	$+$	$-$

5.2. Double-X proximity maps

Without loss of generality, we can assume that $T(\mathcal{Y}_3) = T_b$. Partition the triangle T_b using the vertical line at $c \in (0, 1)$ as in Fig. 26 (right) with $c = 1/2$. Let $R_1(c) := \{(x, y) \in T_b\colon\ x < c\}$ and $R_2(c) := \{(x, y) \in T_b\colon\ x > c\}$. If $(x, y) \in T_b$ with $x = c$, assign (x, y) arbitrarily to one of $R_1(c)$ or $R_2(c)$. We define the double-X proximity map as follows. For $z = (x_o, y_o) \in T_b \setminus \{\mathrm{y}_1, \mathrm{y}_2\}$

$$N_{DX}(z, c) := \begin{cases} \{(x, y) \in T_b\colon\ x \leqslant 2x_o\} & \text{if } z \in R_1(c), \\ \{(x, y) \in T_b\colon\ 1 - x \leqslant 2(1 - x_o)\} & \text{if } z \in R_2(c). \end{cases}$$

If $z = (x_o, y_o) \in \{\mathrm{y}_1, \mathrm{y}_2\}$, then $N_{DX}(z, c) := \{z\}$. See also Fig. 26 (right). Notice that if $z \notin \{\mathrm{y}_1, \mathrm{y}_2\}$, then $N_{DD}(z, M)$ is a right triangle or a quadrilateral. Among the properties, **P1** and **P2** follow trivially. The vertical line at z divides the region into two pieces (half-way along the x-coordinate), so **P3** holds in this special sense. Properties **P4** and **P5** fail to hold, since $N_{DX}(z, c)$ may be a quadrilateral for some $z \in T_b$. Property **P6** holds if $c = 1/2$, otherwise $\mathcal{R}_S(N_{DX}, c)$ has positive area and **P7** also follows by definition. However, **P8** holds only when the regions $R_1(c)$ and $R_2(c)$ are constructed at a point where the vertical line divides the area into two equal pieces.

Property **P10** fails, since $N_{DX}(z)$ is constructed with the boundary of T_b and a line with a specific angle (perpendicular to the largest edge). By Corollary 2.11, geometry invariance for uniform data does not hold. That is, the distributions of relative density and the domination number of the corresponding PCD depend on the geometry of the triangle $T(\mathcal{Y}_3)$. Hence, it does not suffice to compute them for the standard equilateral triangle only, but instead one should compute them for each pair of (c_1, c_2). Moreover, $\Lambda_0(N_{DX}) = \{\mathrm{y}_1, \mathrm{y}_2\}$, since $N_{DX}(x, c)$ has zero area iff $x \in \{\mathrm{y}_1, \mathrm{y}_2\}$.

6. Discussion and conclusions

In this article, we discuss the construction of proximity catch digraphs (PCDs) based on two classes of points $\mathcal{X}$ and $\mathcal{Y}$ in multiple dimensions. Let $\mathcal{X}_n$ and $\mathcal{Y}_m$ be two samples from classes $\mathcal{X}$ and $\mathcal{Y}$, respectively. PCDs are a special type of proximity graphs which have applications in various fields. The class cover catch digraph (CCCD) is the first type of PCD family in the literature [21] which is based on spherical proximity maps (and regions) and has "nice properties" for uniform data in $\mathbb{R}$; in the sense that, the exact and asymptotic distribution of the domination number for CCCDs are available for uniform one-dimensional data. We determine some of the properties of the spherical proximity regions in $\mathbb{R}$ (called *appealing properties*), and use them as guidelines for defining PCDs in higher dimensions. We also characterize the geometry invariance for PCDs based on uniform data. Geometry invariance is important, since it facilitates the computation of quantities (such as relative density or domination number) related to PCDs.

We discuss four PCD families in literature and introduce two new PCD families in this article. We investigate these PCD families in terms of the appealing properties and geometry invariance for uniform data. See Table 1 for the proximity region families with respect to the appealing properties. We provide the asymptotic distribution of relative density and domination number for some of the PCD families. These graph invariants have applications in spatial point pattern analysis and statistical pattern classification. We have demonstrated that the more the properties are satisfied, the better and simpler the asymptotic distribution of relative density. Furthermore, the availability of the asymptotic distribution of domination number is highly correlated with the number of properties satisfied.

The spherical proximity regions were defined with (open) balls only, whereas the new proximity regions are not based on a particular geometric shape or a functional form; that is, the new proximity regions admit any type of region, e.g., circle (ball), arc slice, triangle, a convex or nonconvex polygon, etc. In this sense, the PCDs are defined in a more general setting compared to CCCD. Moreover, the new families of proximity maps we introduce will yield closed regions. Furthermore, the

CCCDs based on balls use proximity regions which are defined by the obvious metric, while the PCDs do not suggest an obvious metric. One main advantage of CCCDs is that they are well defined for all $\mathcal{X}$ points (regardless of they are in the convex hull of $\mathcal{Y}_m$, $\mathcal{C}_H(\mathcal{Y}_m)$, or not) provided that $m \geqslant 1$. On the other hand, the distribution of the domination number and relative density are not analytically tractable for data in $\mathbb{R}^d$ with $d > 1$. All the non-spherical PCDs (i.e., PCDs other than CCCDs) we consider in this article are only well-defined for $\mathcal{X}$ points inside $\mathcal{C}_H(\mathcal{Y}_m)$, but, nevertheless, the distributions of the above graph invariants are tractable for some of them. For testing spatial point patterns, the proportional-edge proximity region with expansion parameter around 1.5–2 have better performance [8], while the central-similarity proximity region with the expansion parameter around 1 [7]. Among these two families, we recommend the proportional-edge PCDs with a parameter within 1.5–2. For large samples, the domination number of the proportional-edge PCD is more sensitive (i.e., more powerful) against the segregation and association alternatives, but for small samples the relative density is more appropriate since the convergence in distribution is faster for the relative density. For points outside the convex hull, a correction factor is introduced for the domination number by Ceyhan [3], a similar factor can be devised for the relative density as well.

Initially, the appealing properties in Section 2.1 were aimed at a particular (important) calculation, namely the distribution of the domination number. However, it turned out that they might be more useful for the geometry invariance of the proximity region and for the distribution of the relative density. For example, the central similarity proximity regions satisfy all the properties, and their relative density has the simplest asymptotic distribution; but the distribution of their domination number is still an open problem. The investigation of the domination number of the arc-slice PCDs sheds some light on that of the spherical PCDs (i.e., CCCDs). By construction, for points in the convex hull, arc-slice PCD is a subdigraph of the CCCD (with the same vertices), hence the domination number of CCCDs is stochastically smaller than that of arc-slice PCDs. Since the upper bound for the arc-slice PCD for data in one triangle is three, for data in the convex hull, CCCDs domination number is bounded by three times the number of Delaunay triangles. Adding more $\mathcal{Y}$ points outside the convex hull so that all $\mathcal{X}$ points are inside the convex hull might provide an upper bound for the domination number of CCCDs as well, but then the applicability for spatial pattern analysis and classification may not be possible.

Most of the discussion in this article is for data in $\mathbb{R}^d$ with $d \leqslant 2$. For higher dimensional data, the geometry invariance results hold as well, and the same properties will be satisfied (with perhaps minor modifications). Moreover, the behavior of the PCDs will be same as in Section 3.4. The asymptotic normality of the relative density for proportional-edge and central similarity PCDs also hold, and domination number of the proportional-edge PCD can be computed as in the two-dimensional case. For example, Ceyhan and Priebe [5] provide the domination number in $\mathbb{R}^3$. However, the calculations of the relative density and domination number might be extremely demanding, if not formidable. There may be a more important problem in very high dimensions, since for large d almost all the points will be outside the convex hull (with high probability). In theory, this is not quite a problem for the asymptotic results, as one simply requires exponentially (in d) large data sets. But in practice this becomes extremely crucial for the applicability of the PCD approach. Perhaps, either PCD approach could be employed after a dimension reduction technique is applied to the data set; or the data can be mapped to a lower dimension by multi-dimensional scaling, and then the PCD approach can be used.

The mechanism to define the proximity maps (and regions) provided in this article can also be used for defining new (perhaps with better properties) proximity map families.

Acknowledgements

I would like to thank anonymous referees, whose constructive comments and suggestions greatly improved the presentation and flow of the paper. This research was supported by the research agency TUBITAK via the Kariyer Project # 107T647.

Appendix A

A.1. Symbols and notation used in the article

$\mathfrak{N}(p,q)$: Neighborhood associated with (unordered) pair of points $p, q \in V$. See page 721.
$\mathfrak{P}$: The property that defines the edge set in $G_{\mathfrak{N},\mathfrak{P}}(V,E)$. See page 721.
$G_{\mathfrak{N},\mathfrak{P}}(V,E)$: Proximity (or neighborhood) graph. See page 721.
$RNG(V)$: Relative neighborhood graph for a set of points $V \subset \mathbb{R}^d$. See page 721.
$NND(V)$: Nearest neighbor digraph for a set of points $V \subset \mathbb{R}^d$. See page 722.
$N(\cdot)$: Proximity map in the most general form defined as $N : \Omega \to \wp(\Omega)$ in measurable space $(\Omega, \mathcal{M})$. See page 722.
$D = (\mathcal{V}, \mathcal{A})$: Digraph with vertex set $\mathcal{V}$ and arc set $\mathcal{A}$. See page 722.
$\gamma(D)$: Domination number for the digraph D. See page 722.
$\gamma_n(r, M)$: The domination number of the PCD based on N^r_{PE} with $\mathcal{X}_n$. See page 743.
$N_{\mathcal{Y}}(\cdot)$: Proximity map based on two classes $\mathcal{X}_n, \mathcal{Y}_m \subseteq \Omega$. See page 723.
$p_a(N_{\mathcal{Y}})$: The probability of having an arc from X_i to X_j, i.e., arc probability for the PCD based on $N_{\mathcal{Y}}$. See page 723.
$p_a(N^r_{PE})$ and $p_a(N^\tau_{CS})$ are also the (asymptotic) means for the relative arc density for proportional-edge and central similarity PCDs. See Eqs. (9) and (12).
$N_S(x)$: Spherical proximity map defined as the open ball $B(x, r(x))$ for all $x \in \mathbb{R} \setminus \mathcal{Y}_m$, where $r(x) = \min_{y \in \mathcal{Y}_m} d(x, y)$. See page 723.

$\mathcal{C}_H(A)$: Convex hull of the set A. See page 724.
$\mathcal{R}_S(N)$: Superset region for the proximity map $N(\cdot)$. See Definition 2.1 on page 724.
$\mathcal{T}_i$: ith Delaunay cell in the Delaunay tessellation of $\mathcal{Y}_m$ in $\mathbb{R}^d$. See page 724.
I_i: ith interval based on $\mathcal{Y}_m$ in $\mathbb{R}$. See page 726.
$\lambda(\cdot)$: The Lebesgue measure on $\mathbb{R}$ (also called $\mathbb{R}$-Lebesgue measure). See page 724.
$\Lambda_0(N)$: The Λ_0-region for the proximity map $N(\cdot)$. See Definition 2.2 on page 725.
$\mathcal{D}_P$: Delaunay tessellation based on a finite data set from a (Poisson) point process. See page 726.
$\mathcal{V}_P$: Poisson Voronoi diagram associated with $\mathcal{D}_P$. See page 726.
T_e: The standard equilateral triangle $T((0,0),(1,0),(1/2,\sqrt{3}/2))$. See page 727.
$T(\mathcal{Y}_3)$: The triangle with vertices $\mathcal{Y}_3 = \{\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3\} \subset \mathbb{R}^2$. See page 727.
T_b: The basic triangle $T((0,0),(1,0),(c_1,c_2))$ with $0 < c_1 \leqslant 1/2$, and $c_2 > 0$ and $(1-c_1)^2 + c_2^2 \leqslant 1$. See page 727.
ϕ_b: The transformation that maps any triangle to T_b. See page 727.
ϕ_e: The transformation that maps T_b to T_e. See page 728.
M_{CC}: The circumcenter of a triangle; M_I: the incenter of a triangle; M_{CM}: the centroid or center of mass of a triangle; and M_O: the orthocenter of a triangle. See page 729.
$\mathcal{V}_C(\mathrm{y}_i)$: The Voronoi cell generated by y_i. See page 731.
$R_M(\mathrm{y})$: The vertex region obtained by using the extensions of the line segments joining y to M (i.e., with Method I). See page 731.
$R_M^{\perp}(\mathrm{y})$: The vertex region obtained by drawing the orthogonal projections to the edges (i.e., with Method II). See page 731.
$R_M(e)$: The edge region for edge e. See page 732.
$\mathcal{S}(F)$: The support of a distribution F. See page 733.
$N_{AS}(x)$: Arc-slice proximity map. See page 733.
$\mathcal{N}_{AS}$: The family of arc-slice proximity regions with center $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$ and vertex regions with Methods I and II. See page 734.
$N_{PE}^r(\cdot, M)$: Proportional-edge proximity map with M-vertex regions. See page 734.
$T_r(x)$: The triangle similar to and with the same orientation as $T(\mathcal{Y}_3)$. See page 734.
$\mathcal{T}^r$: The triangle obtained by drawing the lines $q_i(r,x)$ such that $d(\mathrm{y}_i, e_i) = r\, d(q_i(r,x), \mathrm{y}_i)$ for $i \in \{1,2,3\}$ in $T(\mathcal{Y}_3)$. Vertices of $\mathcal{T}^r$ are $t_1(r)$, $t_2(r)$, and $t_3(r)$. See Fig. 20.
$\mathcal{N}_{PE}^r$: The family of proportional-edge proximity regions with expansion parameter $r \in [1,\infty]$, center $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, and vertex regions with Methods I and II. See page 738.
$\mathfrak{S}(\mathcal{Y}_{d+1})$: The simplex formed by $d+1$ points $\mathcal{Y}_{d+1} = \{\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_{d+1}\}$ in $\mathbb{R}^d$. See page 738.
$\Upsilon(v(x), x)$: The hyperplane parallel to $\varphi(x)$ which contains x where $\varphi(x)$ be the face opposite to vertex $v(x)$ in $\mathfrak{S}(\mathcal{Y}_{d+1})$. See page 738.
$N_{CS}^{\tau}(\cdot, M)$: Central similarity proximity map with M-edge regions. See page 738.
$\mathcal{T}_{\tau}(x)$: The defining triangle for $N_{CS}^{\tau}(x, M)$. See page 738.
$\mathcal{N}_{CS}^{\tau}$: The family of central-similarity proximity regions with expansion parameter $\tau \in [0,1]$ and edge regions based on center $M \in M \in T(\mathcal{Y}_3)^o$. See page 740.
$\rho(D)$: Relative density for digraph D. See page 742.
$\nu(N_{PE}^r)$ and $\nu(N_{CS}^{\tau})$: The (asymptotic) variances for the relative arc density for proportional-edge and central similarity PCDs. See Eq. (10) on page 743.
π_r: The probability $P(\gamma_n(r,M) = 2)$. See Eq. (14) on. See page 743.
$N_{DD}(\cdot, M)$: Directional-doubling proximity map based on M-edge regions. See page 744.
$N_{DX}(\cdot, c)$: Double-X proximity region with centrality parameter c. See Eq. (5.2) on page 745.

References

[1] E. Ceyhan, An investigation of proximity catch digraphs in delaunay tessellations, PhD thesis, The Johns Hopkins University, Baltimore, MD 21218. Also available as technical monograph titled Proximity Catch Digraphs: Auxiliary Tools, Properties, and Applications, VDM Verlag, ISBN 978-3-639-19063-2, 2008.
[2] E. Ceyhan, Extension of one-dimensional proximity regions to higher dimensions, Technical Report # KU-EC-09-2, Koç University, Istanbul, Turkey, arXiv:0902.1306v1 [math.MG], 2009.
[3] E. Ceyhan, Spatial clustering tests based on domination number of a new random digraph family, Communications in Statistics – Theory and Methods (2010), in press.
[4] E. Ceyhan, C. Priebe, Central similarity proximity maps in Delaunay tessellations, in: Proceedings of the Joint Statistical Meeting, Statistical Computing Section, American Statistical Association, 2003.
[5] E. Ceyhan, C.E. Priebe, The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association, Statistics & Probability Letters 73 (2005) 37–50.
[6] E. Ceyhan, C.E. Priebe, On the distribution of the domination number of a new family of parametrized random digraphs, Model Assisted Statistics and Applications 1 (4) (2007) 231–255.
[7] E. Ceyhan, C.E. Priebe, D.J. Marchette, A new family of random graphs for testing spatial segregation, Canadian Journal of Statistics 35 (1) (2007) 27–50.
[8] E. Ceyhan, C.E. Priebe, J.C. Wierman, Relative density of the random r-factor proximity catch digraphs for testing spatial patterns of segregation and association, Computational Statistics & Data Analysis 50 (8) (2006) 1925–1964.
[9] G. Chartrand, L. Lesniak, Graphs & Digraphs, Chapman & Hall/CRC Press LLC, Florida, 1996.

[10] J. DeVinney, The class cover problem and its applications in pattern recognition, PhD thesis, The Johns Hopkins University, Baltimore, MD 21218, 2003.
[11] J. DeVinney, C.E. Priebe, A new family of proximity graphs: Class cover catch digraphs, Discrete Applied Mathematics 154 (14) (2006) 1975–1982.
[12] J. DeVinney, C.E. Priebe, D.J. Marchette, D. Socolinsky, Random walks and catch digraphs in classification, in: Proceedings of the 34th Symposium on the Interface: Computing Science and Statistics, vol. 34, 2002, http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf.
[13] S. Janson, T. Łuczak, A. Rucinński, Random Graphs, Wiley–Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York, 2000.
[14] J.W. Jaromczyk, G.T. Toussaint, Relative neighborhood graphs and their relatives, Proceedings of IEEE 80 (1992) 1502–1517.
[15] C. Kimberling, Encyclopedia of triangle centers, 2008; http://faculty.evansville.edu/ck6/encyclopedia/ETC.html.
[16] E.L. Lehmann, Nonparametrics: Statistical Methods Based on Ranks, Prentice-Hall, Upper Saddle River, NJ, 1988.
[17] D.J. Marchette, C.E. Priebe, Characterizing the scale dimension of a high dimensional classification problem, Pattern Recognition 36 (1) (2003) 45–60.
[18] K.V. Mardia, R. Edwards, M.L. Puri, Analysis of central place theory, Bulletin of International Statistical Institute 47 (1977) 93–110.
[19] A. Okabe, B. Boots, K. Sugihara, S.N. Chiu, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, Wiley, 2000.
[20] M.S. Paterson, F.F. Yao, On nearest neighbor graphs, in: Proceedings of 19th Int. Coll. Automata, Languages and Programming, in: LNCS, vol. 623, Springer, 1992, pp. 416–426.
[21] C.E. Priebe, J.G. DeVinney, D.J. Marchette, On the distribution of the domination number of random class cover catch digraphs, Statistics & Probability Letters 55 (2001) 239–246.
[22] C.E. Priebe, D.J. Marchette, J. DeVinney, D. Socolinsky, Classification using class cover catch digraphs, Journal of Classification 20 (1) (2003) 3–23.
[23] C.E. Priebe, J.L. Solka, D.J. Marchette, B.T. Clark, Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays, Computational Statistics & Data Analysis on Visualization 43 (4) (2003) 621–632.
[24] M. Sen, S. Das, A. Roy, D. West, Interval digraphs: An analogue of interval graphs, Journal of Graph Theory 13 (1989) 189–202.
[25] G.T. Toussaint, The relative neighborhood graph of a finite planar set, Pattern Recognition 12 (4) (1980) 261–268.
[26] Z. Tuza, Inequalities for minimal covering sets in sets in set systems of given rank, Discrete Applied Mathematics 51 (1994) 187–195.
[27] E. Weisstein, Triangle centers, Eric Weisstein's world of mathematics, http://mathworld.wolfram.com/TriangleCenter.html, 2008.