Computing the Region Areas of Euler Diagrams Drawn with Three Ellipses

Luana Micallef and Peter Rodgers

School of Computing, University of Kent, UK
{L.Micallef,P.J.Rodgers}@kent.ac.uk

Abstract. Ellipses generate accurate area-proportional Euler diagrams for more data than is possible with circles. However, computing the region areas is difficult as ellipses have various degrees of freedom. Numerical methods could be used, but approximation errors are introduced. Current analytic methods are limited to computing the area of only two overlapping ellipses, but area-proportional Euler diagrams in diverse application areas often have three curves. This paper provides an overview of different methods that could be used to compute the region areas of Euler diagrams drawn with ellipses. We also detail two novel analytic algorithms to instantaneously compute the exact region areas of three general overlapping ellipses. One of the algorithms decomposes the region of interest into ellipse segments, while the other uses integral calculus. Both methods perform equally well with respect to accuracy and time.

Keywords: Euler diagram, Venn diagram, ellipses, area, area-proportional

1 Introduction

Area-proportional Euler diagrams are used in various areas, such as genetics [24] and information search [8], to represent both the set relations (by the intersecting closed curves) and their cardinalities (by the area of the regions) [5]. Often these diagrams are drawn for data with two or three sets and are Venn diagrams, like those in Fig. 1, such that all the possible curve intersections are depicted [28].

Euler diagrams drawn with circles facilitate user comprehension due to their smoothness, symmetry and good continuation that make the curves easily distinguishable even at intersections [3]. Most area-proportional Euler diagrams are thus drawn with circles [28]. However, circles have limited degrees of freedom (i.e., a centre and a radius) and draw accurate diagrams for any data with two sets [5] but not three [6]. Polygons are more flexible and draw accurate diagrams for any 3-set data [6], but their non-smooth curves impede comprehension [2].

Ellipses have more degrees of freedom (i.e., a centre, a semi-minor axis, a semi-major axis, and an angle of rotation) than circles, but like circles are symmetric and smooth. Their helpful features have been long noted [5], but due to difficulties in computing the region areas of the diagram, a drawing method that uses ellipses (known as eulerAPE) was only recently devised [19].

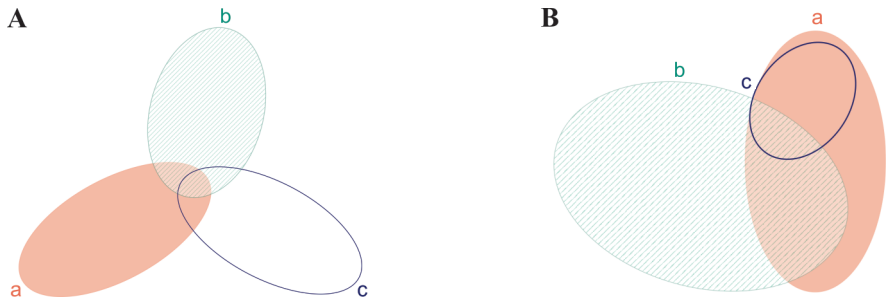

Fig. 1. Area-proportional Venn diagrams for three sets, a, b and c, and set relation cardinalities (A) $\{a = 100, b = 100, c = 100, ab = 5, ac = 5, bc = 5, abc = 5\}$ and (B) $\{a = 17671, b = 65298, c = 3, ab = 19904, ac = 12951, bc = 4, abc = 4816\}$. These diagrams were accurately and automatically generated using eulerAPE [19].

Approximate methods could be used to compute the region areas of Euler diagrams. However, these methods introduce error [15] that could distort the diagram and thus violate Tufte's graphical integrity principle that "representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented" [27]. Though humans are biased to area judgement [7], such accuracy and integrity is still important. Firstly, it is still unclear how areas in such diagrams are perceived and what inaccuracies are not noticeable. Secondly, it is highly unlikely for an area to be perceived in a same way by different individuals [18]. Tufte emphasized that "graphical excellence begins with telling the truth about the data" [27] and for this reason, he highly criticized metrics that scale objects (e.g., on maps [21]) based on how they could be perceived.

Analytic methods can compute the exact region areas of Euler diagrams, but none of the current methods are appropriate for three general ellipses. Drawing methods for area-proportional Euler diagrams use optimization search techniques [6] and so, analytic methods that compute the region areas should be computationally efficient to allow for the generation of accurate diagrams in a time that maintains user's attention.

In this paper, we discuss current methods to compute the region areas of Euler diagrams with ellipses (Section 3) and we describe our two novel analytic methods to instantaneously (in 10 milliseconds) compute the region areas of Euler diagrams with three general ellipses (Section 4). Using one of our analytic methods, eulerAPE (`http://www.eulerdiagrams.org/eulerAPE`) [19] draws accurate area-proportional Venn diagrams with ellipses for a large majority of random 3-set data (86%, N=10,000) in a relatively fast time (97% of the diagrams within 1 second). The two analytic methods for computing the region areas of three intersecting ellipses have not been previously detailed. Hence, this forms the research contribution of this paper.

We start by introducing basic concepts and definitions in relation to ellipses.

2 Basic Concepts and Definitions

An ellipse is a simple closed curve characterized by (Fig. 2A): a centre, (γ_1,γ_2); two semi-axes, α and β, where $\alpha \geq \beta$; an angle of rotation, θ, where $0 \leq \theta < 2\pi$. Since $\alpha \geq \beta$, α and β are respectively referred to as the semi-major axis and semi-minor axis. An *ellipse* is in *canonical form* if $(\gamma_1,\gamma_2) = (0,0)$ and $\theta = 0$, so that in the Cartesian coordinate system the ellipse is centred on the origin and the ellipse's semi-axes are along the x-axis and y-axis, as shown in Fig. 2B. An ellipse with any (γ_1,γ_2) and θ is *a general ellipse*.

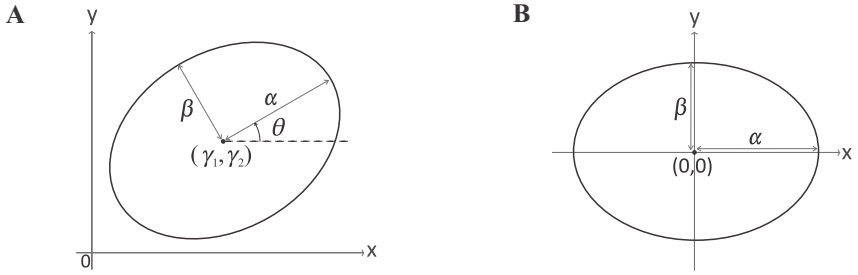

Fig. 2. An ellipse and its properties in (A) general form and (B) canonical form.

The *area of an ellipse* is

$$\pi\alpha\beta\,. \tag{1}$$

The *curve of an ellipse* using a *polar coordinate system* with pole (or origin) (γ_1,γ_2) and the polar axis a ray from (γ_1,γ_2) passing through α is defined as

$$\rho^2 = \frac{\alpha^2\beta^2}{\beta^2\cos^2\varphi + \alpha^2\sin^2\varphi} \tag{2}$$

where φ is the polar angle between α and a ray from (γ_1,γ_2) passing through a point on the ellipse .

An *ellipse arc* ($\frown$) is a connected portion of the ellipse curve (e.g., $\frown MN$ in Fig. 3A). An *ellipse sector* ($\triangledown$) is the space bounded by an ellipse arc and two line segments between the ellipse's centre and the arc's endpoints (e.g., $\triangledown MNO$ in Fig. 3B). The *area of an ellipse sector* can be defined using (2) as

$$\begin{aligned}\frac{1}{2}\int_{\varphi_1}^{\varphi_2}\rho^2 d\varphi &= \frac{\alpha^2\beta^2}{2}\int_{\varphi_1}^{\varphi_2}\frac{d\varphi}{\beta^2\cos^2\varphi + \alpha^2\sin^2\varphi} \\ &= \frac{\alpha\beta}{2}\tan^{-1}\left(\frac{\alpha}{\beta}\tan\varphi\right)\Bigg]_{\varphi_1}^{\varphi_2}\end{aligned} \tag{3}$$

where φ_1 and φ_2 are the polar angles of the endpoints of the arc from α with respect to (γ_1,γ_2) and $0 < \varphi_2 - \varphi_1 \leq 2\pi$.

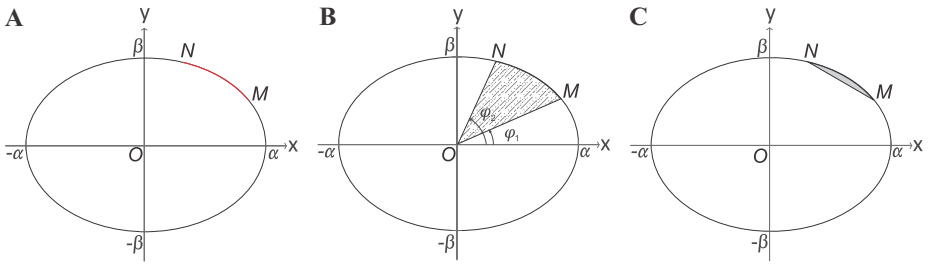

Fig. 3. (A) An ellipse arc, (B) an ellipse sector, and (C) an ellipse segment.

The value returned by (3) is the area of the ellipse sector from φ_1 to φ_2 in an anticlockwise direction along the ellipse curve. Thus, to compute the area of $\triangledown MNO$ in Fig. 3B using (3), φ_1 and φ_2 should respectively be the polar angles of M and N from α, which in this case is along the x-axis, with respect to (γ_1,γ_2), which in this case is (0,0) or O. If instead, φ_1 and φ_2 are respectively the polar angles of N and M from α with respect to O, the area computed by (3) is the area of the ellipse minus the area of $\triangledown MNO$.

An *ellipse segment* ($\frown$) is the space bounded by an ellipse arc and a chord that share the same endpoints (e.g., $\frown MN$ in Fig. 3C). An ellipse sector ($\triangledown MNO$) is composed of an ellipse segment ($\frown MN$) with the same ellipse arc as the sector and a triangle ($\triangle MNO$). Thus, the *area of an ellipse segment* can be defined as

$$\text{Area of } \frown MN = \text{ Area of } \triangledown MNO \; - \; \text{Area of } \triangle MNO\,. \tag{4}$$

The area of $\triangledown MNO$ in (4) is (3) for the same arc endpoints as that of $\frown MN$.

The polar coordinates representation of an ellipse curve in (2) assumes that (γ_1,γ_2) is at the pole and α is along the polar axis of the coordinate system. So (2) cannot represent a general ellipse.

The *curve of a general ellipse* can be defined *parametrically* as

$$\begin{aligned} x(t) &= \gamma_1 + \alpha\cos\theta\cos t - \beta\sin\theta\sin t \\ y(t) &= \gamma_2 + \alpha\sin\theta\cos t + \beta\cos\theta\sin t \\ &\textit{where } t \textit{ is the parameter and } 0 \le t \le 2\pi \textit{ and } (x(t), y(t)) \\ &\textit{are the Cartesian coordinates of a point on the ellipse curve}\,. \end{aligned} \tag{5}$$

In (5), t is the angular parameter that determines the position of a particle moving along the ellipse curve, so that every value of t determines the Cartesian coordinates $(x(t),\, y(t))$ of a point on the ellipse curve. As shown in Fig. 4, given a point $(x,\, y)$ on the ellipse curve, the corresponding value of t is determined by drawing a line perpendicular to α that passes through $(x,\, y)$ and intersects with the auxiliary circle (i.e., a circle with radius α and centre (γ_1,γ_2) that is the circumscribed circle of the ellipse; in red in Fig. 4) at a point P, and by then

computing the anticlockwise angle from α to the line passing through (γ_1,γ_2) and P with respect to (γ_1,γ_2). The latter angle is the value of t. Since in Fig. 4, $\theta = 0$ and point (x, y) is on the top right quarter of the ellipse curve, the value of t for (x, y) can be computed using any one of the following

$$t = \cos^{-1}\left(\frac{x-\gamma_1}{\alpha}\right) \quad \text{or} \quad t = \sin^{-1}\left(\frac{y-\gamma_2}{\beta}\right) . \tag{6}$$

If $\theta = 0$ and the point (x, y) is on any other quarter of the ellipse curve, the equations in (6) have to be adapted. For instance, if (x, y) is on the upper left quarter of the ellipse curve, t is π minus the value of t in (6). If $\theta \neq 0$, the point (x, y) has to be rotated by -θ about (γ_1, γ_2) before an equation of t in (6) or adaptations of them can be used.

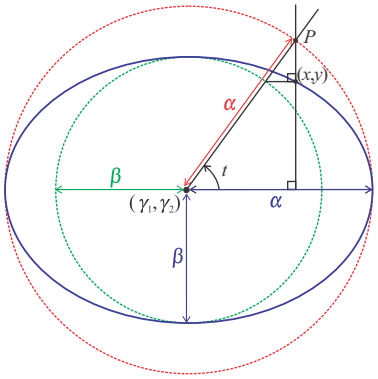

Fig. 4. Computing parameter t of the parametric representation of the curve of a general ellipse, given the Cartesian coordinates (x, y) of a point on the ellipse curve.

Alternatively, the *curve of a general ellipse* can be defined by the set of points (x, y) on the Cartesian plane that satisfy the *implicit polynomial equation*

$$\frac{\left((x-\gamma_1)\cos\theta + (y-\gamma_2)\sin\theta\right)^2}{\alpha^2} + \frac{\left((y-\gamma_2)\cos\theta - (x-\gamma_1)\sin\theta\right)^2}{\beta^2} = 1 . \tag{7}$$

This is more complex to handle compared with the parametric representation in (5) and the implicit polynomial equation that should be satisfied by the set of points (x, y) that define the curve of an ellipse in canonical form, that is

$$\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1 . \tag{8}$$

We now discuss available methods, after which we introduce our novel analytic methods.

3 Available Methods

3.1 Approximate Methods

A numerical quadrature method can be used to compute the region areas of any number of intersecting curves irrespective of their shape. The drawing method venneuler [28] uses this method to draw area-proportional Euler diagrams with circles. In venneuler, each circle is drawn on a 200×200 pixel plane and each pixel on each plane is set to 1 if it is in the circle or 0 if not. The area of each region in the diagram is the sum of the value of all the pixels on all the planes that are located in that region.

Such numerical approximation introduce errors in the calculation of the region areas, as shown in Fig. 5 for an Euler diagram drawn with ellipses. In Fig. 5A and Fig. 5B, the same diagram is drawn on two identical 200×200 grids with a difference in the placement of the diagrams on the grids. The magnified views of regions *ab*, *ac*, *bc* and *abc* indicate that according to Fig. 5A there are 0 grid squares or pixels that are only in region *bc*, while according to Fig. 5B there are 3. Similar differences can be noted for the other regions due to a difference in the positioning of the diagrams on the grids. Using a larger grid and thinner outlines for the curves could reduce inaccuracies, but such issues will still be inevitable. Also, small regions (e.g., region *bc* in Fig. 5) could be considered missing even though they are depicted. This could impede the drawing method from handling diagrams with small regions and impede its optimization process from taking a path that could lead to an improved solution. If, like venneuler, the drawing method uses a steepest descent optimization approach, the analytic gradient cannot be computed and the generated diagram is less likely to be accurate.

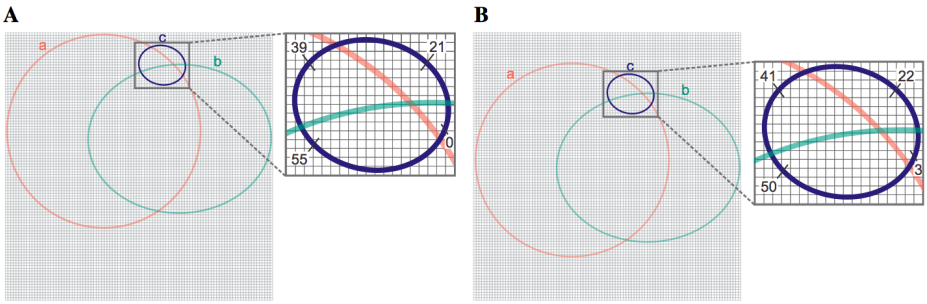

Fig. 5. The region areas of intersecting ellipses using a numerical quadrature method.

Checking whether a point is in a closed curve could be computationally complex and expensive. With circles, venneuler takes the Cartesian coordinates of the pixel's location on the plane and applies them to the implicit polynomial equation of the circle to verify whether the distance between the circle's centre and the pixel is less than the circle's radius. If the latter is true, the pixel is in the circle. However, checking if such a pixel is in a general ellipse and conducting this check for all the pixels on all the planes is computationally expensive, as ellipses have more degrees of freedom than circles and the implicit polynomial

equation of the curve of a general ellipse (7) is more complex. If instead, the implicit polynomial equation of the curve of an ellipse in canonical form (8) is used, both the ellipse and the pixel would have to be translated and rotated before this simpler equation is used. Alternatively, a secant line passing through the ellipse's centre and the pixel to be checked could be drawn, and the intersection points of this line with the ellipse could be computed. If the distance between the ellipse's centre and the intersection point that has the same polar angle (from α about (γ_1,γ_2)) as that of the pixel is greater than the distance between the ellipse's centre and the pixel, then the pixel is in the ellipse. However, all of these methods are more computationally expensive than those for circles.

Alternatively, the ellipses could be represented using regular convex polygons and a standard polygon intersection algorithm could be used to compute the region areas. For instance, the drawing method VennMaster computes the region areas of its Euler diagrams with polygons by applying the Gaussian integration theorem [17]. However, the region areas are not computed accurately and small regions are often considered missing. The polygons should have numerous vertices and small edges otherwise the curves would be highly non-smooth, leading to unreliable approximations of the region areas. Polygon intersection algorithms for polygons with numerous vertices are also more computationally expensive than the numerical quadrature method [25].

Monte Carlo methods could also be used to approximate the region areas, but due to repeated random sampling, they could be computationally expensive.

3.2 Analytic Methods

Analytic methods to compute the region areas of two [4] or three [5] intersecting circles are available. For more circles, approximation methods are often used. For ellipses, only two analytic methods are available: one by Eberly [9] and another by Hughes and Chraibi [14].

Both methods are restricted to two ellipses. Eberly's method is further restricted to ellipses in canonical form. Hughes and Chraibi's method can handle any two general ellipses with any centre and angle of rotation. Both methods compute the area of the overlapping region by first obtaining the area of the two ellipse segments making up the region (as shown later in Fig. 6B). The area of each ellipse segment is obtained by computing the area of an ellipse sector and then subtracting the area of a triangle, as discussed in Section 2. The derivation of the area of a sector of an ellipse in canonical form is obtained using integral calculus. The representation of the ellipse curve is defined in polar coordinates by Eberly and parametrically by Hughes and Chraibi. To handle general ellipses, Hughes and Chraibi first translate and rotate the general ellipses so they are transformed into canonical form, and then compute the area of the required ellipse segment using the same equation as that of ellipses in canonical form.

Though an analytic method might seem computationally expensive due to the various degrees of freedom of ellipses, Hughes and Chraibi's method has been efficient enough to compute the area of two overlapping ellipses for simulations of dynamic systems, such as an orbiting satellite with a solar calibrator [16].

4 Our Analytic Methods

We devised two general analytic methods for three ellipses, both of which can handle ellipses that are not necessarily in canonical form. These include:

M1. Decomposition into ellipse segments;
M2. Using integral calculus.

These methods differ in the way they compute the area of the overlapping region between two ellipses and that of an ellipse segment. Similar to Eberly's [9] and Hughes and Chraibi's [14] methods, M1 decomposes the region of interest into ellipse segments and uses an equation of the segment area of an ellipse in canonical form. M2 uses integral calculus to directly derive the equation of the required enclosed region area without the need of any geometric transformations. M1 and M2 are discussed later.

The general algorithm for computing the region areas of three intersecting ellipses in a Venn diagram using either M1 or M2 is as follows:

Algorithm 1: *ComputeRegionAreasOfThreeIntersectingEllipses* (d)

Input: A Venn diagram d with three ellipses.
Output: *regionAreas*, a set of areas one for every region in d.

1. **for each** ellipse e in d **do**
2. $ellipseAreas[e] \leftarrow$ area of e by (1)
3. **end for**
4. **for each** pair of ellipses e_1 and e_2 in d **do**
5. $points[e_1,e_2] \leftarrow$ intersection points of e_1 and e_2 by Hill's method [13]
6. $overlapAreas[e_1,e_2] \leftarrow$ area of the overlapping region between e_1 and e_2 by M1 or M2 and $points[e_1,e_2]$
7. $interiorPoints[e_1,e_2] \leftarrow$ the intersection point in $points[e_1,e_2]$ that is inside the third ellipse of d
8. **end for**
9. Decompose the region in all three ellipses of d into ellipse segments es_1, es_2, es_3 and triangle ts (as in Fig. 6D) using *interiorPoints*, which defines the arc endpoints of es_1, es_2, es_3 and the vertices of t
10. $regionAreas[e_1,e_2,e_3] \leftarrow \sum$ areas of es_1, es_2, es_3 by (4) and ts
11. **for each** pair of ellipses e_1 and e_2 in d **do**
12. $regionAreas[e_1,e_2] \leftarrow overlapAreas[e_1,e_2] - regionAreas[e_1,e_2,e_3]$, where e_3 is the other ellipse in d
13. **end for**
14. **for each** ellipse e in d **do**
15. $otherAreas \leftarrow \sum regionAreas[e,e_1]$, $regionAreas[e,e_2]$, $regionAreas[e,e_1,e_2]$, where e_1 and e_2 are the two other ellipses in d
16. $regionAreas[e] \leftarrow ellipseAreas[e] - otherAreas$
17. **end for**
18. **return** *regionAreas*

Algorithm 1 has been implemented for both M1 and M2 and for any possible representation of a Venn diagram with three ellipses that are not necessarily in canonical form and that intersect pairwise exactly twice. M1 and M2 as well as the chosen method to compute the intersection points of the ellipses are discussed in the next sections.

If the general intersecting ellipses do not always intersect each other exactly twice, an extended version of Algorithm 1 would have to be implemented, so that the various ways in which the ellipses can intersect would be handled (there can be from zero up to four intersection points between two ellipses [9]). The method we have chosen to compute the intersection points [13] returns all the intersection points (i.e., zero up to four) between any two general ellipses. So the algorithm can easily be extended by: (i) identifying the way each pair of ellipses intersect from the number of intersection points between the two ellipses; (ii) decomposing the relevant regions into ellipse segments and basic geometry shapes like triangles or rectangles whenever necessary; (iii) using either M1 or M2 to find the area of the overlapping region between two ellipses and the area of ellipse segments; (iv) using (1) to compute the area of the ellipses; (v) using basic algebra to add and subtract areas wherever necessary to calculate the region areas. In such cases, a region in the diagram could be made up of multiple connected components and so, its area would be the sum of the area of all of these components.

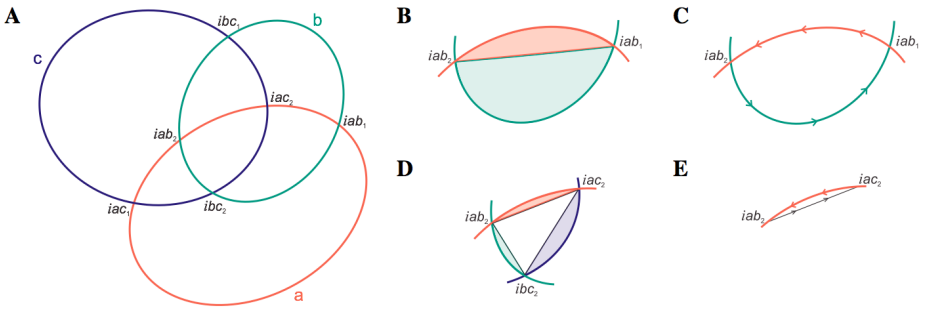

Fig. 6. Computing the area of the overlapping region between two ellipses and the area of the region that is in all three ellipses using M1 and M2.

4.1 M1: Decomposition Into Ellipse Segments

To compute the area of the overlapping region of two general ellipses, M1 decomposes the region into two ellipse segments, as in Fig. 6B. As explained in Algorithm 1 and Fig. 6D, the region in all three ellipses is similarly decomposed into ellipse segments and a triangle.

Equations defining the area of a segment of an ellipse in canonical form can easily be derived and could be less complex than the ones for a general

ellipse. Such equations are already available. For example, Eberly [9] and Hughes and Chraibi [14] provide two different equations based on the area of a sector of an ellipse in canonical form. Eberly's equation uses the polar coordinates representation of the ellipse curve, while that of Hughes and Chraibi uses the parametric representation of the ellipse curve.

M1 uses (4) to compute the area of an ellipse segment which, similar to Eberly's equation, uses the polar coordinates representation of the ellipse curve in (2). As shown in Fig. 6B and Fig. 6D, the arc endpoints of an ellipse segment are two of the intersection points of the three overlapping ellipses, which define the arc in an anticlockwise direction along the ellipse curve (the importance of this direction is explained in Section 2 in relation to (3)). The arc endpoints and the ellipse are rotated by an angle of -θ about (γ_1, γ_2) and translated to (0,0), so that the transformed ellipse is in canonical form. Rotation and translation are affine transformations and so, the ellipse's properties, its area and the area of the required ellipse segment are preserved. The polar angle of each transformed endpoint from α with respect to (0,0) is then computed, and together with the transformed ellipse, these polar angles are used to compute the area of the required ellipse segment using (4).

4.2 M2: Using Integral Calculus

M2 derives an equation of the area of the required enclosed region from definite integrals that compute the area under an ellipse curve (or line) between two given points. The equation of the curve of a general ellipse is used and so, M2 does not require any of the geometric transformations used in M1 to get the ellipse in canonical form.

Let e_1 and e_2 be two ellipses that intersect at points i_1 and i_2, such that the overlapping region is enclosed by an ellipse arc from i_1 and i_2 in an anticlockwise direction along e_1 and an ellipse arc from i_1 and i_2 in an anticlockwise direction along e_2. In M2, the area of the overlapping region between e_1 and e_2 is

$$\text{Area under curve } e_1 \text{ from } i_1 \text{ to } i_2 \; + \; \text{Area under curve } e_2 \text{ from } i_2 \text{ to } i_1 \; . \tag{9}$$

This is illustrated in Fig. 6C for ellipses a and b in Fig. 6A, where a and b are respectively e_1 and e_2 in (9), and iab_1 and iab_2 are respectively i_1 and i_2 in (9).

To compute the area of the region located in all three ellipses, M2 decomposes the region into ellipse segments and a triangle, as in M1 and as shown in Fig. 6D, where the arc endpoints of the ellipse segments and the vertices of the triangle are defined by the intersection points of the three overlapping ellipses. Given a segment of an ellipse e with an arc from i_1 and i_2 in an anticlockwise direction along e and a secant line l intersecting e at i_1 and i_2, M2 defines the area of the ellipse segment as

$$\text{Area under curve } e \text{ from } i_1 \text{ to } i_2 \; + \; \text{Area under line } l \text{ from } i_2 \text{ to } i_1 \; . \tag{10}$$

This is illustrated in Fig. 6E for the segment of ellipse a between iac_2 and iab_2 that is located in all three ellipses of Fig. 6A, where a is e in (9) and iac_2 and iab_2 are respectively i_1 and i_2 in (9).

The area under a curve between two given points is defined by a definite integral. The curve of a general ellipse can be defined in different ways (Section 2). However, it is not always so straightforward to handle some representations. For instance, the implicit polynomial in (7) has to be converted to an explicit polynomial before it can be used to find the area under an ellipse curve. The parametric representation in (5) can be used as is and determining the antiderivative of this representation is straightforward. So M2 uses the parametric representation of the curve of the general ellipse in (5), that is $x(t)$ and $y(t)$, and defines the area under the curve of a general ellipse e from (x_1, y_1) to (x_2, y_2) on e as follows:

If $|x_1 - x_2| > |y_1 - y_2|$

If $y = F(x)$ is the explicit definition of e as a function of x and $x(t_1) = x_1$ and $x(t_2) = x_2$, the area under the curve e from (x_1, y_1) to (x_2, y_2) is

$$\begin{aligned}
&\int_{x_1}^{x_2} F(x)dx = \int_{t_1}^{t_2} F(x(t))x'(t)dt = \int_{t_1}^{t_2} y(t)x'(t)dt = \\
&\int_{t_1}^{t_2} (\gamma_2 + \alpha\sin\theta\cos t + \beta\cos\theta\sin t)\,(-\alpha\cos\theta\sin t - \beta\sin\theta\cos t)\,dt = \\
&K_1\sin 2t + K_2\sin t + K_3\cos 2t + K_4\cos t + K_5 t\Big]_{t_1}^{t_2} \\
&where\ \ K_1 = \frac{\alpha\beta\cos 2\theta}{4}, \quad K_2 = -\gamma_2\beta\sin\theta, \quad K_3 = \frac{\left(\alpha^2 + \beta^2\right)\sin 2\theta}{8}, \\
&\qquad K_4 = \gamma_2\alpha\cos\theta, \ \ K_5 = -\frac{\alpha\beta}{2}.
\end{aligned} \tag{11}$$

Else (i.e., $|x_1 - x_2| \le |y_1 - y_2|$)

If $x = F(y)$ is the explicit definition of e as a function of y and $y(t_1) = y_1$ and $y(t_2) = y_2$, the area under the curve e from $(x_1, {}_1)$ to (x_2, y_2) is

$$\begin{aligned}
&\int_{y_1}^{y_2} F(y)dy = \int_{t_1}^{t_2} F(y(t))y'(t)dt = \int_{t_1}^{t_2} x(t)y'(t)dt = \\
&\int_{t_1}^{t_2} (\gamma_1 + \alpha\cos\theta\cos t - \beta\sin\theta\sin t)\,(-\alpha\sin\theta\sin t + \beta\cos\theta\cos t)\,dt = \\
&K_1\sin 2t + K_2\sin t + K_3\cos 2t + K_4\cos t + K_5 t\Big]_{t_1}^{t_2} \\
&where\ \ K_1 = \frac{\alpha\beta\cos 2\theta}{4}, \quad K_2 = \gamma_1\beta\cos\theta, \quad K_3 = \frac{\left(\alpha^2 + \beta^2\right)\sin 2\theta}{8}, \\
&\qquad K_4 = \gamma_1\alpha\sin\theta, \ \ K_5 = \frac{\alpha\beta}{2}.
\end{aligned} \tag{12}$$

The value of t_1 and t_2 corresponding to (x_1, y_1) and (x_2, y_2) respectively is computed as discussed in Section 2 in relation to (5) and (6).

Similarly, the area under a line l from (x_1, y_1) to (x_2, y_2) on l where m and c are respectively the gradient and y-intercept of l is defined as:

If $|x_1 - x_2| > |y_1 - y_2|$

If $y = mx + c$ explicitly defines l as a function of x, the area under l from (x_1, y_1) to (x_2, y_2) is

$$\int_{x_1}^{x_2} (mx + c)dx = \frac{mx^2}{2} + cx \Big]_{x_1}^{x_2} \tag{13}$$

Else (i.e., $|x_1 - x_2| \leq |y_1 - y_2|$)

If $x = \frac{y - c}{m}$ explicitly defines l as a function of y, the area under l from (x_1, y_1) to (x_2, y_2) is

$$\frac{1}{m} \int_{y_1}^{y_2} (y - c)dy = \frac{1}{m} \left(\frac{y^2}{2} - cy \right) \Big]_{y_1}^{y_2} \tag{14}$$

However, to compute the region area of the overlapping ellipses, the intersection points of the ellipses are required.

4.3 Computing the Intersection Points of the Ellipses

All the intersection points of the ellipses in a diagram can be obtained by computing the intersection points of every ellipse pair.

An ellipse is a curve and so, the various methods of computing the intersection points of two curves [11, 25] can be adapted for the intersection points of two ellipses. Numerical methods, such as Bézier and internal subdivision [29], Bézier clipping [26] and the Newton-Raphson method, could also be used. However, as explained earlier, numerical approximation methods introduce error and can distort the diagram.

The most common analytic methods include: (i) the resultant-based method using for instance Bezout's resultant [9, 10]; (ii) the Gröbner basis method [11]; (iii) the matrix-based method [13]. These methods have been used for various areas (e.g., the resultant-based method [14]; the Gröbner basis method [23]; the matrix-based method [1]).

Methods (i) and (ii) are more complex than (iii) as the roots of a quartic polynomial have to be solved using Ferrari's solution or other methods [12]. For example, Hughes and Chraibi's [14] analytic method for the region areas of two

intersecting ellipses uses a numerical implementation of Ferrari's solution [22] to find the roots of the quartic polynomial. However, we only require the real roots and so, finding all the roots of a quartic polynomial is a waste of computational resources [12].

Using the matrix representation of conic sections and homogeneous transformation matrices, method (iii) can compute the intersection points of the two ellipses without referring to any quartic polynomials. This method has now been extended to efficiently identify whether two solid ellipsoid intersect [1]. Thus, for our analytic methods, we use Hill's method [13] to compute the ellipses' intersection points.

5 Evaluation of Our Methods

M1, M2 and an approximate method were used to compute the region areas of 8000 Venn diagrams with three ellipses whose properties were randomly generated and which intersect exactly twice pairwise. Some of these randomly generated diagrams had very small region areas that are barely visible like regions c and bc in Fig. 1B. The computed areas were then compared. The approximate method represented the ellipses as regular convex polygons and used a standard polygon intersection algorithm to find the area of the regions and to compute the intersection points.

The same areas were obtained by our analytic methods, M1 and M2 (with an occasional insignificant difference of less than 10^{-4}). The average percentage error of the areas provided by the approximate method with respect to any one of our analytic methods was 1.04%. M1 and M2 returned the same area for small regions like those in Fig. 1B, but the approximate method disregarded these regions as if they were missing and the represented set relations were non-existent.

M1 and M2 computed the areas in around 10 milliseconds on an Intel Core i7 CPU @2GHz with 8GB RAM, OS X 10.8.4 and Java Platform 1.6.0_51. This response time is 10 times less than the 0.1 second limit for an instantaneous response [20], and similar to that of numerical methods (e.g., venneuler's [28] response time is 1 millisecond using a numerical quadrature method on a MacBook Pro @2.5Ghz with 2GB RAM and Java Platform 1.5). Thus, our analytic methods can also be used for any other application, including simulations of dynamic systems, where both accuracy and efficiency is important.

6 Conclusion

We have discussed methods to compute the region areas of Euler diagrams drawn with ellipses. We first reviewed current available methods, namely approximate and analytic methods, and we then described our two novel analytic methods to compute the region areas of three general intersecting ellipses.

Methods that automatically draw area-proportional Euler diagrams with three curves require a mechanism to compute the region areas of diagrams in

relatively fast time. A recent drawing method eulerAPE [19] demonstrated the benefits of ellipses in drawing such diagrams accurately and with smooth curves. However, the region areas must first be computed accurately and an analytic method is thus required. Current analytic methods are restricted to two ellipses and so our methods are the first to calculate region areas for three ellipses. One method decomposes the regions into ellipse segments and the other uses integral calculus, but both return accurate region areas instantaneously (in 10 milliseconds), even when the diagram has very small regions as in Fig. 1B.

We implemented Algorithm 1 that measures the region areas of Venn diagrams with three curves and ellipses that intersect exactly twice pairwise. We describe how to handle other Euler diagrams with three curves, but this extended version is not yet implemented.

The regions of an Euler diagram with any number of ellipses can be decomposed into ellipse segments and possibly a polygon (e.g., a triangle or a quadrilateral). The area of the overlapping region between two ellipses and that of an ellipse segment can be computed with our methods, while the area of an ellipse and that of a polygon can be computed using standard geometry formula. Thus, our methods can easily be extended to instantaneously compute the region areas of Euler diagrams with any number of ellipses. Methods that draw area-proportional Euler diagrams with more than three ellipses can then be devised.

References

1. Salvatore Alfano and Meredith L Greer. Determining If Two Solid Ellipsoids Intersect. *Journal of Guidance, Control, and Dynamics*, 26(1):106–110, 2003.
2. F Benoy and P Rodgers. Evaluating the comprehension of Euler diagrams. *Proceedings of the 11th International Conference on Information Visualization (IV)*, pages 771–780, 2007. IEEE.
3. A Blake, G Stapleton, P Rodgers, L Cheek, and J Howse. The impact of shape on the perception of euler diagrams. *Proceedings of the 8th International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Artificial Intelligence 8578*, pages 123–137, 2014. Springer.
4. S Chow and P Rodgers. Constructing area-proportional Venn and Euler diagrams with three circles. *Proceedings of the 2nd International Workshop on Euler Diagrams*, 2005.
5. S Chow and F Ruskey. Drawing Area-Proportional Venn and Euler Diagrams. *Proceedings of the 11th International Symposium on Graph Drawing (GD 2003), Lecture Notes in Computer Science 2912*, pages 466–477, 2004. Springer.
6. Stirling Christopher Chow. *Generating and Drawing Area-Proportional Venn and Euler Diagrams*. PhD thesis, University of Victoria, Victoria, BC, Canada, 2007.
7. William S Cleveland. *The Elements of Graphing Data, Revised Edition*. Hobart Press, Summit, NJ, USA, 1994.
8. Tuan Dang, Anushka Anand, and Leland Wilkinson. FmFinder: Search and Filter Your Favorite Songs. *Advances in Visual Computing, Lecture Notes in Computer Science*, 7431:348–358, 2012. Springer.
9. D Eberly. The Area of Intersecting Ellipses. `http://www.geometrictools.com/Documentation/AreaIntersectingEllipses.pdf`, 2010. Geometric Tools, LLC.

10. D Eberly. Intersection of Ellipses. `http://www.geometrictools.com/Documentation/IntersectionOfEllipses.pdf`, 2011. Geometric Tools, LLC.
11. Gerald E Farin, Josef Hoschek, and Myung-Soo Kim. *Handbook of Computer Aided Geometric Design*. Elsevier (North-Holland), Amsterdam, The Netherlands, 2002.
12. Don Herbison-Evans. Solving Quartics and Cubics for Graphics. In Alan W Paeth, editor, *Graphics Gems V*, pages 3–15. Morgan Kaufmann, USA, 1995.
13. Kenneth J Hill. Matrix-based Ellipse Geometry. In Alan W Paeth, editor, *Graphics Gems V*, pages 72–77. Morgan Kaufmann, USA, 1995.
14. GB Hughes and M Chraibi. Calculating Ellipse Overlap Area. `http://works.bepress.com/gbhughes/17`, 2011. arXiv:1106.3787 [physics.comp-ph].
15. Eugene Isaacson. *Analysis of numerical methods*. Courier Dover Publications, Mineola, NY, USA, 1994.
16. S Kent, ME Kaiser, SE Deustua, J Smith, and et al. Photometric calibrations for 21st century science. 2009. arXiv:0903.2799v1 [astro-ph.CO].
17. Hans A Kestler, Andr Mller, Johann M Kraus, Malte Buchholz, Thomas M Gress, Hongfang Liu, David W Kane, Barry R Zeeberg, and John N Weinstein. VennMaster: Area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC Bioinformatics*, 9:67, 2008. BioMed Central Ltd.
18. Hans-Joachim Meihoefer. The visual perception of the circle in thematic maps/experimental results. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(1):63–84, 1973. UT Press.
19. Luana Micallef and Peter Rodgers. eulerAPE: Drawing Area-Proportional 3-Venn Diagrams using ellipses. *PLoS ONE*, 9(7):e101717, 2014. Public Library of Science. `http://www.eulerdiagrams.org/eulerAPE`.
20. RB Miller. Response time in man-computer conversational transactions. *Proceedings of the December 9-11, 1968 (AFIPS) fall joint computer conference, part I*, pages 267–277, 1968. ACM.
21. Daniel R Montello. Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. *Cartography and Geographic Information Science*, 29(3):283–304, 2002. Cartography and Geographic Information Society.
22. Terence RF Nonweiler. Algorithm 326: Roots of low-order polynomial equations. *Communications of the ACM*, 11(4):269–270, 1968. ACM.
23. Federico Pernici. *Two Results in Computer Vision using Projective Geometry*. PhD thesis, University of Florence, Florence, Italy, 2005.
24. Christian RA Regenbrecht, Marc Jung, Hans Lehrach, and James Adjaye. The molecular basis of genistein-induced mitotic arrest and exit of self-renewal in embryonal carcinoma and primary cancer cell lines. *BMC Medical Genomics*, 1:49, 2008. BioMed Central Ltd.
25. Philip Schneider and David H Eberly. *Geometric tools for computer graphics*. Morgan Kaufmann, San Francisco, CA, USA, 2002.
26. Thomas W Sederberg and Tomoyuki Nishita. Curve intersection using Bézier clipping. *Computer-Aided Design*, 22(9):538–549, 1990. Elsevier.
27. Edward Rolf Tufte. *The Visual Display of Quantitative Information, 1st Edition*. Graphics Press, Cheshire, CT, USA, 1983.
28. Leland Wilkinson. Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 18(2):321–331, 2012. IEEE.
29. CK Yap. Complete subdivision algorithms, I: Intersection of Bezier curves. *Proceedings of the 22nd Annual Symposium on Computational Geometry (SoCG)*, pages 217–226, 2006. ACM.