The Area of Intersecting Ellipses

David Eberly
Geometric Tools, LLC
http://www.geometrictools.com/

Created: September 4, 2008
Last Modified: September 5, 2008

Contents

1 Introduction

This document describes an algorithm for computing the area of intersection of two ellipses. The formulas are in closed form, thus providing the *exact* area in terms of real-valued arithmetic. Naturally, the computer evaluation of the trigonometric functions in the formulas has some numerical round-off errors, but the formulas allow you to avoid (1) approximating the ellipses by convex polygons, (2) using the intersection of convex polygons as an approximation to the intersection of ellipses, and (3) using the area of intersection of convex polygons as an approximation to the area of intersection of ellipses.

The algorithm has two main aspects: Computing the points of intersection of the ellipses and computing the area bounded by a line and an elliptical arc.

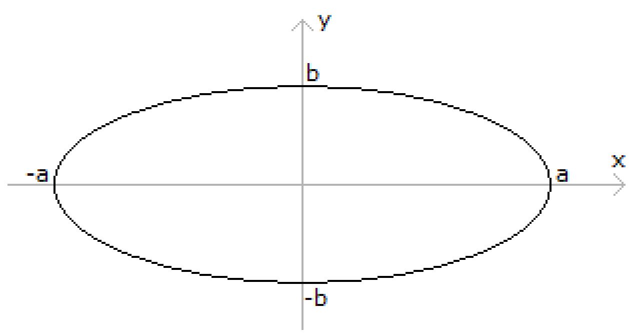
2 Area of an Ellipse

An axis-aligned ellipse centered at the origin is

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \tag{1}$$

where I assume that $a > b$, in which case the major axis is along the x-axis. Figure 2.1 shows such an ellipse.

Figure 2.1 An axis-aligned ellipse centered at the origin with $a > b$.



The area bounded by the ellipse is πab. Using the methods of calculus, the area A is four times that of the area in the first quadrant,

$$A = 4\int_0^a y\,dx = 4\int_0^a b\sqrt{1-(x/a)^2}\,dx \tag{2}$$

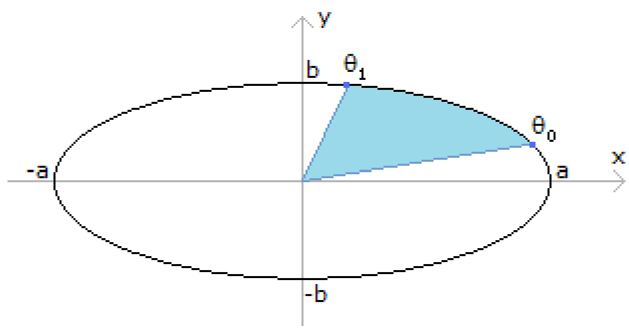
The integral may be computed using the change of variables $x = a\cos\theta$ for $0 \le \theta \le \pi/2$. The differential is $dx = -a\sin\theta\, d\theta$ and the area is

$$\begin{aligned}
A &= 4\int_0^a b\sqrt{1-(x/a)^2}\,dx \\
&= 4b\int_{\pi/2}^0 \sin\theta(-a\sin\theta\,d\theta) \\
&= 4ab\int_0^{\pi/2} \sin^2\theta\,d\theta \\
&= 2ab\int_0^{\pi/2} \left(1-\cos(2\theta)\right)\,d\theta \\
&= 2ab\left(\theta - \tfrac{1}{2}\sin(2\theta)\right)\Big|_0^{\pi/2} \\
&= 2ab\left[\left(\tfrac{\pi}{2} - \tfrac{1}{2}\sin(\pi)\right) - \left(0 - \tfrac{1}{2}\sin(0)\right)\right] \\
&= \pi ab
\end{aligned} \tag{3}$$

3 Area of an Elliptical Sector

An *elliptical arc* is a portion of the ellipse bounded by two points on the ellipse. The arc is delimited by angles θ_0 and θ_1, both in $[-\pi,\pi]$. If the arc does not cross the negative y-axis, we may choose $\theta_0 < \theta_1$. If the arc does cross the negative y-axis, for the purpose of numerical computation we may consider it as an arc from θ_0 to $\theta_1 + 2\pi$. This convention supports an integration formula involving branches of the inverse tangent function, as we will see later. An *elliptical sector* is the region bounded by an elliptical arc and the line segments containing the origin and the endpoints of the arc. Figure 3.1 shows an elliptical arc and the corresponding elliptical sector.

Figure 3.1 An elliptical arc and its corresponding elliptical sector.



The polar-coordinate representation of the arc is obtained by substituting $x = r\cos\theta$ and $y = r\sin\theta$ into Equation (1) and solving for r^2,

$$r^2 = \frac{a^2b^2}{b^2\cos^2\theta + a^2\sin^2\theta} \tag{4}$$

The area of the sector is

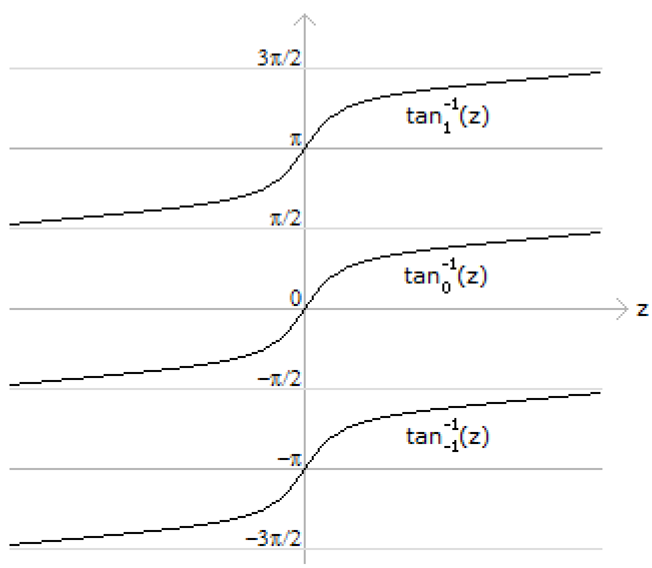
$$A(\theta_0,\theta_1) = \int_{\theta_0}^{\theta_1} \frac{1}{2}r^2\,d\theta = \frac{a^2b^2}{2}\int_{\theta_0}^{\theta_1} \frac{d\theta}{b^2\cos^2\theta + a^2\sin^2\theta} \tag{5}$$

The following formula involves an integral whose value is found in standard books containing tables of integrals; of course, it may also be derived from basic principles of calculus,

$$
\begin{aligned}
A(\theta_0, \theta_1) &= \frac{a^2 b^2}{2} \int_{\theta_0}^{\theta_1} \frac{d\theta}{b^2 \cos^2 \theta + a^2 \sin^2 \theta} \\
&= \frac{ab}{2} \tan^{-1}\left(\frac{a}{b} \tan(\theta)\right)\Big|_{\theta_0}^{\theta_1} \\
&= \frac{ab}{2} \left[\tan^{-1}\left(\frac{a}{b}\tan(\theta_1)\right) - \tan^{-1}\left(\frac{a}{b}\tan(\theta_0)\right)\right]
\end{aligned}
\tag{6}
$$

where $\tan^{-1}(\cdot)$ is the branch of the inverse tangent function that corresponds to the angle at which $\tan(\cdot)$ is evaluated when substituting in the limits of integration. Figure 3.2 shows the graphs of three branches of the inverse tangent function.

Figure 3.2 The graphs of three branches of the inverse tangent function.



The principal branch is labeled $\tan_0^{-1}(z)$ in the figure; its range is $(-\pi/2, \pi/2)$. The other branches are defined by $\tan_k^{-1}(z) = \tan_0^{-1}(z) + k\pi$ for integrer-valued $k \neq 0$.

For example, consider the entire ellipse when $\theta_0 = -\pi$ and $\theta_1 = \pi$. The area is

$$
\begin{aligned}
A(-\pi, \pi) &= \frac{ab}{2} \left[\tan_1^{-1}\left(\frac{a}{b}\tan(\pi)\right) - \tan_{-1}^{-1}\left(\frac{a}{b}\tan(-\pi)\right)\right] \\
&= \frac{ab}{2} \left[\tan_1^{-1}(0) - \tan_{-1}^{-1}(0)\right] \\
&= \frac{ab}{2} [(\pi) - (-\pi)] \\
&= \pi ab
\end{aligned}
\tag{7}
$$

As another example, consider the half ellipse in the first and fourth quadrants when $\theta_0 = -\pi/2$ and $\theta_1 = \pi/2$. The tangent function is undefined at $\pm\pi/2$, so the integral is computed in the limiting sense,

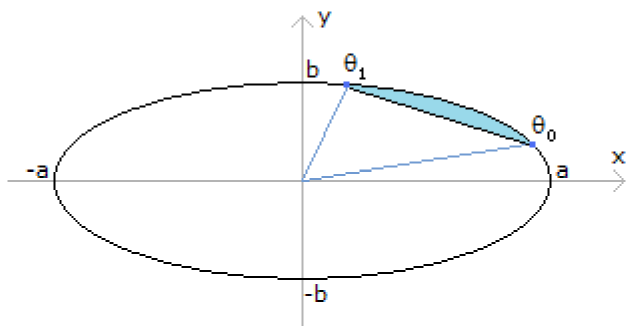
$$\begin{aligned} A(-\pi/2, \pi/2) &= \lim_{\varepsilon \to 0^+} A(-\pi/2+\varepsilon, \pi/2-\varepsilon) \\ &= \tfrac{ab}{2} \lim_{\varepsilon \to 0^+} \left[\tan_0^{-1}\left(\tfrac{a}{b}\tan(\pi/2-\varepsilon)\right) - \tan_0^{-1}\left(\tfrac{a}{b}\tan(-\pi/2+\varepsilon)\right)\right] \\ &= \tfrac{ab}{2}\left[\tan_0^{-1}(+\infty) - \tan_0^{-1}(-\infty)\right] \\ &= \tfrac{ab}{2}[(\pi/2) - (-\pi/2)] \\ &= \tfrac{1}{2}\pi ab \end{aligned} \tag{8}$$

The plus-sign superscript on 0 in the limit notation indicates that ε decreases to zero through positive values. The area is half that of the entire ellipse, as expected.

4 Area Bounded by a Line Segment and an Elliptical Arc

Figure 4.1 shows the region bounded by an elliptical arc and the line segment connecting the arc's endpoints.

Figure 4.1 The region bounded by an elliptical arc and the line segment connecting the arc's endpoints.



The area of this region is the area of the elliptical sector minus the area of the triangle whose vertices are the origin, $(0,0)$, and the arc endpoints $(x_0, y_0) = (a\cos\theta_0, b\sin\theta_0)$ and $(x_1, y_1) = (a\cos\theta_1, b\sin\theta_1)$. The triangle area is

$$\frac{1}{2}|x_1 y_0 - x_0 y_1| = \frac{ab}{2}|\cos\theta_1 \sin\theta_0 - \cos\theta_0 \sin\theta_1| = \frac{ab}{2}|\sin(\theta_1 - \theta_0)| \tag{9}$$

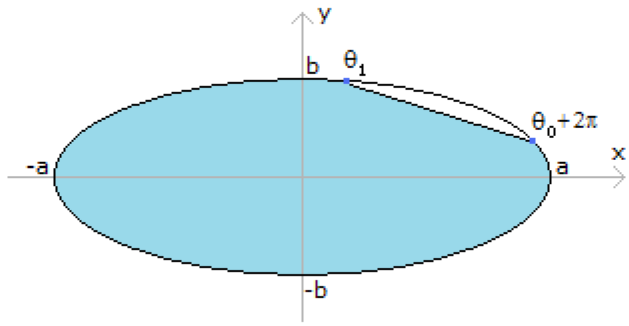
If $\alpha(\theta_0, \theta_1)$ denotes the area of the aforementioned region, then

$$\begin{aligned} \alpha(\theta_0, \theta_1) &= A(\theta_0, \theta_1) - \tfrac{ab}{2}|\sin(\theta_1 - \theta_0)| \\ &= \tfrac{ab}{2}\left[\tan^{-1}\left(\tfrac{a}{b}\tan(\theta_1)\right) - \tan^{-1}\left(\tfrac{a}{b}\tan(\theta_0)\right) - |\sin(\theta_1 - \theta_0)|\right] \end{aligned} \tag{10}$$

where $A(\theta_0, \theta_1)$ is the area of the sector as defined in Equation (6) and the other term on the right-hand side is the area of the triangle as defined in Equation (9).

The region of interest could be that bounded by the line segment and the elliptical arc that is spanned counterclockwise from θ_1 to $\theta_0 + 2\pi$. Figure 4.2 illustrates.

Figure 4.2 The other region bounded by an elliptical arc and the line segment connecting the arc's endpoints.



The area of this region is

$$\alpha(\theta_1, \theta_0 + 2\pi) = \pi ab - \alpha(\theta_0, \theta_1) \tag{11}$$

The bounded region has the area of the ellipse minus the area of the smaller region bounded by the line segment and elliptical arc. In Equation (10), the origin is outside the bounded region. In Equation (11), the origin is inside the bounded region.

5 Intersection Points of Ellipses

The ellipses may be written as quadratic forms,

$$(\mathbf{P} - \mathbf{C}_i)^{\mathrm{T}} R_i^{\mathrm{T}} D_i R_i (\mathbf{P} - \mathbf{C}_i) = 1, \quad i = 0, 1 \tag{12}$$

The center of ellipse i is $\mathbf{C}_i$, a 2×1 point. The orientation is $R_i = [\mathbf{U}_{i0}\ \mathbf{U}_{i1}]$, a 2×2 rotation matrix whose first column $\mathbf{U}_{i0}$ is the 2×1 direction of the major axis and whose second column $\mathbf{U}_{i1}$ is the 2×1 direction of the minor axis. The 2×2 matrix $D_i = \mathrm{Diag}(1/a_i^2, 1/b_i^2)$ is a diagonal matrix where a_i is the distance along the major axis from the center to the ellipse and b_i is the distance along the minor axis from the center to the ellipse. The 2×1 point $\mathbf{P}$ is any point on the ellipse.

Ignoring the i subscript for the sake of clarity, an ellipse $(\mathbf{P} - \mathbf{C})^{\mathrm{T}} R^{\mathrm{T}} D R (\mathbf{P} - \mathbf{C}) = 1$ is parameterized by

$$\mathbf{P}(\theta) = \mathbf{C} + (a \cos\theta)\mathbf{U}_0 + (b \sin\theta)\mathbf{U}_1, \quad \theta \in (-\pi, \pi] \tag{13}$$

The point $\mathbf{C}$ may be viewed as the origin of a coordinate system whose axes have directions $\mathbf{U}_0$ and $\mathbf{U}_1$. If (x, y) is a 2-tuple in the coordinate system, the point in the original coordinate system is $\mathbf{P} = \mathbf{C} + x\mathbf{U}_0 + y\mathbf{U}_1$, in which case $x = \mathbf{U}_0 \cdot (\mathbf{P} - \mathbf{C})$ and $y = \mathbf{U}_1 \cdot (\mathbf{P} - \mathbf{C})$. For the parameterized ellipse, $x = a\cos\theta$ and $y = b\sin\theta$, so $(x/a)^2 + (y/b)^2 = 1$. Thus, the ellipse is axis-aligned in the coordinate system induced by $\mathbf{C}$, $\mathbf{U}_0$, and $\mathbf{U}_1$.

The ellipses may also be written as quadratic equations,

$$\begin{aligned} s_0 + s_1 x + s_2 y + s_3 x^2 + s_4 xy + s_5 y^2 = 0 \\ t_0 + t_1 x + t_2 y + t_3 x^2 + t_4 xy + t_5 y^2 = 0 \end{aligned} \tag{14}$$

where (x, y) is the 2-tuple version of $\mathbf{P}$ in Equation (12). This formulation is convenient for computing the points of intersection of the ellipses; that is, the points are generated by solving simultaneously the two quadratic equations.

The quadratic equations may be viewed as quadratic polynomials in x with coefficients that depend on y,

$$\begin{aligned} Q_0(x,y) = f(x) = (s_0 + s_2 y + s_5 y^2) + (s_1 + s_4 y)x + (s_3)x^2 = \sigma_0 + \sigma_1 x + \sigma_2 x^2 = 0 \\ Q_1(x,y) = g(x) = (t_0 + t_2 t + t_5 y^2) + (t_1 + t_4 y)x + (t_3)x^2 = \tau_0 + \tau_1 x + \tau_2 x^2 = 0 \end{aligned} \tag{15}$$

The two polynomials $f(x)$ and $g(x)$ have a common root if and only if the Bézout determinant is zero,

$$(\sigma_2\tau_1 - \sigma_1\tau_2)(\sigma_1\tau_0 - \sigma_0\tau_1) - (\sigma_2\tau_0 - \sigma_0\tau_2)^2 = 0. \tag{16}$$

This determinant is constructed by

$$0 = \sigma_2 g(x) - \tau_2 f(x) = (\sigma_2\tau_1 - \sigma_1\tau_2)x + (\sigma_2\tau_0 - \sigma_0\tau_2) \tag{17}$$

and

$$0 = \tau_1 f(x) - \sigma_1 g(x) = (\sigma_2\tau_1 - \sigma_1\tau_2)x^2 + (\sigma_0\tau_1 - \sigma_1\tau_0), \tag{18}$$

Equation (17) is solved for x and substituted it into Equation (18) to produce Equation (16). When the Bézout determinant is zero, the common root of $f(x)$ and $g(x)$ is

$$\bar{x} = \frac{\sigma_2\tau_0 - \sigma_0\tau_2}{\sigma_1\tau_2 - \sigma_2\tau_1}.$$

The common root to $f(x) = 0$ and $g(x) = 0$ is obtained from the linear equation $\sigma_2 g(x) - \tau_2 f(x) = 0$ by solving for x.

Equation (16) is a quartic polynomial in y, say,

$$B(y) = u_0 + u_1 y + u_2 y^2 + u_3 y^3 + u_4 y^4 \tag{19}$$

where

$$\begin{aligned} u_0 &= d_{31}d_{10} - d_{30}^2 \\ u_1 &= d_{34}d_{10} + d_{31}(d_{40} + d_{12}) - 2d_{32}d_{30} \\ u_2 &= d_{34}(d_{40} + d_{12}) + d_{31}(d_{42} - d_{51}) - d_{32}^2 - 2d_{35}d_{30} \\ u_3 &= d_{34}(d_{42} - d_{51}) + d_{31}d_{45} - 2d_{35}d_{32} \\ u_4 &= d_{34}d_{45} - d_{35}^2 \end{aligned} \tag{20}$$

where $d_{ij} = s_i t_j - s_j t_i$. For each $\bar{y}$ solving $B(\bar{y}) = 0$ solve $Q_0(x, \bar{y}) = 0$ for up to two values $\bar{x}$. Keep only the valid solutions, those for which $Q_0(\bar{x}, \bar{y}) = 0$ and $Q_1(\bar{x}, \bar{y}) = 0$.

6 Area of Intersecting Ellipses

The quartic polynomial of Equation (19) has an even number of real-valued roots: 0, 2, or 4. However, these can be repeated. In geometric terms, repeated real-valued roots correspond to intersection points where the two ellipses are tangent. A distinct real-valued root corresponds to an intersection point where the ellipses intersect transversely.

In the following discussion, the ellipses are named E_0 and E_1. The general logic for the area computation is shown next. The subsections after this pseudocode describe why the function is structured as it is.

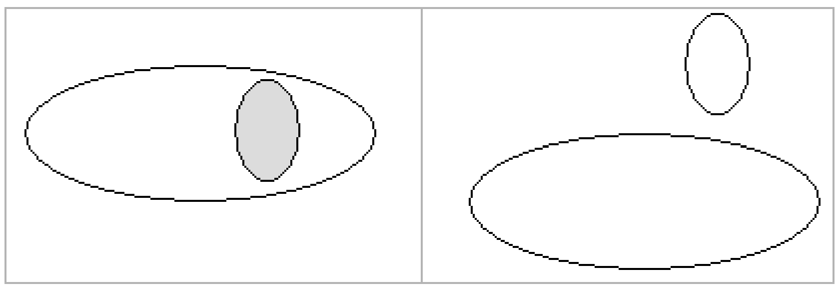
```
real AreaOfIntersection (Ellipse E0, Ellipse E1)
{
    Polynomial2 Q0 = E0.GetQuadraticRepresentation();  // Q0(x,y)
    Polynomial2 Q1 = E0.GetQuadraticRepresentation();  // Q1(x,y)
    Polynomial1 B = GetBezoutDeterminant(Q0, Q1);  // B(y)

    // Compute the roots of B.  The input to ComputeRoots is B.  The output numDistinctRoots is the number of distinct
    // real-valued roots.  The output root[] stores the distinct roots, where only array elements 0 through
    // numDistinctRoots-1 are valid.  The output multiplicity[] stores the number of times the roots occur.  For a
    // distinct root, the multiplicity is 1.  For a repeated root, the multiplicity is 2.
    int numDistinctRoots, multiplicity[4];
    real root[4];
    ComputeRoots(B, numDistinctRoots, root, multiplicity);

    // Compute the intersection points.  The points are ordered counterclockwise about their centroid.
    Point2 intr[4];
    ComputeIntersections(E0, E1, Q0, Q1, numDistinctRoots, root, multiplicity, intr);

    if (numDistinctRoots == 0)
    {
        // Returns area(E0) [E0 is contained in E1], area(E1) [E1 is contained in E0], or zero [E0 and E1 are separated].
        return AreaOfIntersectionCS(E0, E1);
    }
    else if (numDistinctRoots == 1)  // multiplicity[0] must be 2.
    {
        return AreaOfIntersectionCS(E0, E1);
    }
    else if (numDistinctRoots == 2)
    {
        if (multiplicity[0] == 2)  // Two roots, each repeated.  multiplicity[1] must be 2.
        {
            return AreaOfIntersectionCS(E0, E1);
        }
        else  // Two distinct roots.  Region bounded by two arcs, one from each ellipse.
        {
            return AreaOfIntersection2(E0, E1, intr[0], intr[1]);
        }
    }
    else if (numDistinctRoots == 3)
    {
        if (multiplicity[0] == 2)
        {
            return AreaOfIntersection2(E0, E1, intr[1], intr[2]);
        }
        else if (multiplicity[1] == 2)
        {
            return AreaOfIntersection2(E0, E1, intr[2], intr[0]);
        }
        else // multiplicity[2] == 2
        {
            return AreaOfIntersection2(E0, E1, intr[0], intr[1]);
        }
    }
    else // numDistinctRoots == 4
    {
        return AreaOfIntersection4(E0, E1, intr);
    }
}
```

The assumption is that E_0 and E_1 are represented as quadratic forms, as defined by Equation (12), and that the quadratic polynomials of Equation (14) must be computed. Of course, you can store as much information as you like in the `Ellipse` data structure to avoid having to compute various quantities at run time.

6.1 0 Intersection Points

One ellipse is contained strictly in the other, or the ellipses (as solids) are separated. Figure 6.1 illustrates.

Figure 6.1 Left: One ellipse is contained by the other. Right: The ellipses are separated.



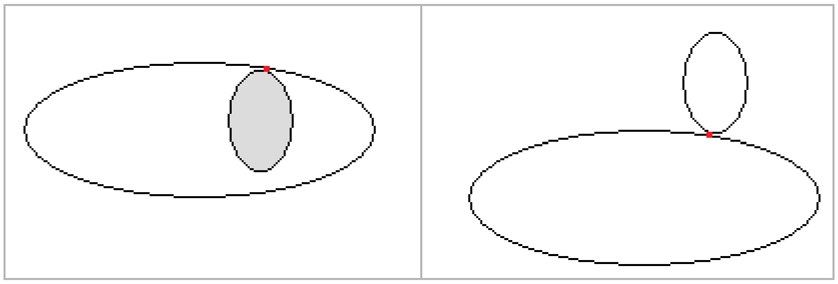
In the case of containment, the area of intersection is the area of the smaller ellipse. In the case of separation, the area of intersection is zero. The logic is

```
real AreaOfIntersectionCS (Ellipse E0, Ellipse E1)
{
    if (E0.Contains(E1.center))
    {
        return Area(E1);
    }
    else if (E1.Contains(E0.center))
    {
        return Area(E0);
    }
    else
    {
        return 0;
    }
}
```

6.2 1 Intersection Point

One ellipse is contained in the other but the two ellipses are tangent at the point of intersection, or the ellipses (as solids) are separated except for a single point of tangency. Figure 6.2 illustrates.

Figure 6.2 Left: One ellipse is contained by the other but they are tangent at a single point. Right: The ellipses are separated except for a single point of tangency. The intersection point is drawn as a red dot.

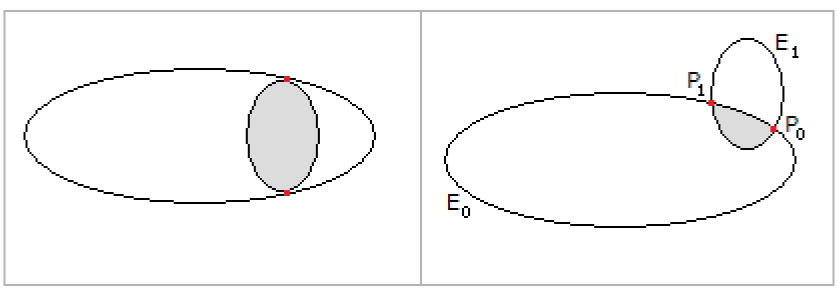


The area of intersection is computed using the function `AreaOfIntersectionCS` that was described in the subsection for 0 intersection points.

6.3 2 Intersection Points

One ellipse is contained in the other but the two ellipses are tangent at the point of intersection, or the ellipses intersect at two distinct points. Figure 6.3 illustrates.

Figure 6.3 Left: One ellipse is contained by the other but they are tangent at two points. Right: The ellipses intersect at two distinct points. The intersection points are drawn as red dots.



For the case shown in the left of the figure, the area of intersection is computed using the function `AreaOfIntersectionCS` because one ellipse is contained in the other.

The more interesting case is shown in the right of the figure. The region of intersection is bounded by two elliptical arcs, one from each ellipse. In the coordinate system induced by E_0, the points of intersection are

$$\mathbf{P}_j = \mathbf{C}_0 + x_j \mathbf{U}_{00} + y_j \mathbf{U}_{01}, \quad j = 0, 1 \tag{21}$$

where $x_j = a_0 \cos\theta_j$ and $y_j = b_0 \sin\theta_j$. The area of the region bounded by the E_0-arc and the line segment connecting $\mathbf{P}_0$ and $\mathbf{P}_1$ is computed using either Equation (10) or Equation (11) depending on whether $\mathbf{C}_0$ is outside or inside the region. The angles are counterclockwise ordered as $\theta_0 < \theta_1$. The a and b in the area equations are replaced by a_0 and b_0.

In the coodinate system induced by E_1, the points of intersection are

$$\mathbf{P}_j = \mathbf{C}_1 + x_j \mathbf{U}_{10} + y_j \mathbf{U}_{11}, \quad j = 0, 1 \tag{22}$$

where $x_j = a_1 \cos\phi_j$ and $y_j = b_1 \sin\phi_j$. The area of the region bounded by the E_1-arc and the line segment connecting $\mathbf{P}_0$ and $\mathbf{P}_1$ is computed using either Equation (10) or Equation (11) depending on whether $\mathbf{C}_1$ is outside or inside the region. The angles are counterclockwise ordered as $\phi_0 < \phi_1$. The a and b in the area equations are replaced by a_1 and b_1. The angles ϕ_j replace the θ_j in the formulas, but be careful to use the correct order as the formulas require.

The function

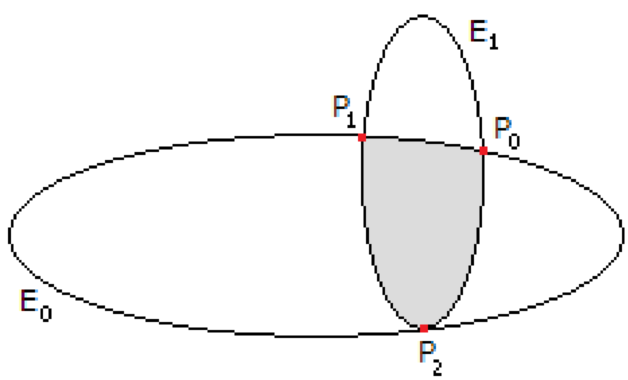
```
real AreaOfIntersection2 (Ellipse E0, Ellipse E1, Point2 P0, Point P1);
```

computes the area of the subregion associated with $\langle \mathbf{P}_0, \mathbf{P}_1 \rangle$ using ellipse E_0. It also computes the area of the subregion associated with $\langle \mathbf{P}_1, \mathbf{P}_0 \rangle$ using ellipse E_1. The returned result is the sum of the two areas.

6.4 3 Intersection Points

The two ellipses intersect tangentially at one point and transversely at two points. Figure 6.4 illustrates.

Figure 6.4 Ellipses that intersect in 3 points.

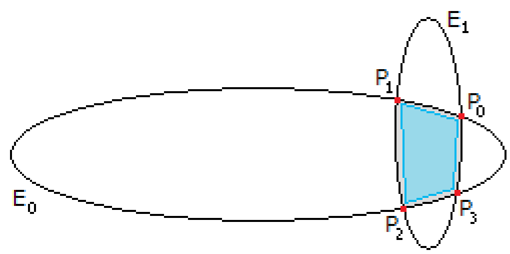


The point of tangency is somewhat irrelevant here. The region of intersection is still bounded by two elliptical arcs, one from each ellipse, so the algorithm for 2 intersection points still applies. In the figure, $\mathbf{P}_2$ is not required in the area calculation.

6.5 4 Intersection Points

The two ellipses intersect transversely at four points. Figure 6.5 illustrates.

Figure 6.5 Ellipses that intersect in 4 points.



The region of intersection is the union of a convex quadrilateral and four special regions, each region bounded by an elliptical arc and the line segment connecting the endpoints of the arc.

The function

```
real AreaOfIntersection4 (Ellipse E0, Ellipse E1, Point2 P[4]);
```

computes the area of the subregion associated with $\langle \mathbf{P}_0, \mathbf{P}_1 \rangle$ and the area of the subregion associated with $\langle \mathbf{P}_2, \mathbf{P}_3 \rangle$. Ellipse E_1 is used for these area calculations. Similarly, the function computes the area of the subregion associated with $\langle \mathbf{P}_1, \mathbf{P}_2 \rangle$ and the area of the subregion associated with $\langle \mathbf{P}_3, \mathbf{P}_0 \rangle$. Ellipse E_0 is used for these area calculations. The area of the quadrilateral is computed as the sum of areas of two triangles, $\langle \mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2 \rangle$ and $\langle \mathbf{P}_3, \mathbf{P}_0, \mathbf{P}_2 \rangle$.