Journal of Combinatorics and Number Theory
Volume 4, Number 3, pp. 25–39
ISSN 1942-5600

On Invariant Area Formulas and Lattice Point Bounds for the Intersection of Hyperbolic and Elliptic Regions

Mike Bennett, Kirill Y. Lazebnik, Patrick X. Rault* and Jeffrey A. Singer
Department of Mathematics, State University of New York
Geneseo, United States of America

Abstract

We provide an explicit invariant formula for the area of intersection of the regions interior to ellipses and hyperbolas centered at the origin. A further explicit formula is given for the area of union of the interior of two ellipses centered at the origin. We construct our regions using nonsingular real quadratic forms f and g, with the region in question being bounded by potentially four curves of the form $f(x,y) = \pm 1$ and $g(x,y) = \pm 1$. Our formulas make use of the following invariants of these quadratic forms under the group of linear transformations: discriminant, resultant, and polarization of discriminant. In addition, we provide an invariant upper bound for the number of lattice points in a region bounded by ellipses. This fills a void in the literature, as noted in the recent paper [4] by the third author which merely gives an upper bound for the area computed here in the hyperbolic case. Indeed, the aforementioned paper specifically states that an exact formula for the area will provide a route to generalized results.

2010 Mathematics Subject Classification: Primary 51M25; Secondary 11P21

1. Introduction

Conic regions have been studied since the time of the ancient Greeks. The only conic which forms a bounded region is the ellipse, whose area is given by the well known formula $2\pi/\sqrt{-D}$, where D is its discriminant. Despite the long history of conic regions, our main theorem is the first to provide exact formulas for the area of intersection of elliptic and hyperbolic regions, such as those in Figure 1 below.

Indeed, a literature search reveals [1], which gives a computational algorithm for computing the area of intersection of specified horizontal or vertical ellipses. A similar article, [2], gives another algorithm with several concrete applications including "simulations for both the satellite solar calibrator and force-based pedestrian dynamic model" which they caution requires "efficient calculation of the overlap area between two ellipses;" our exact formulas should give the best possible efficiency in these applications.

*E-mail address: rault@geneseo.edu; Website: http://geneseo.edu/~rault

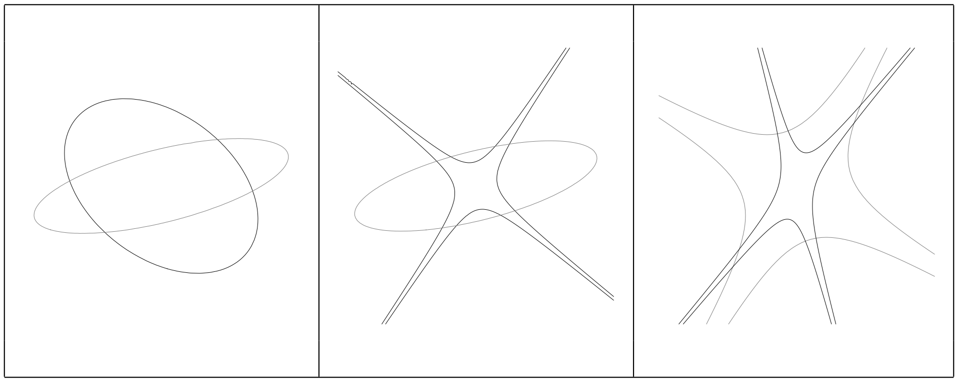

Figure 1. Regions interior to ellipses and hyperbolas.

Theorem 1.1. *Let f and g be real binary quadratic forms with resultant R, discriminants $D(f)$ and $D(g)$, and polarization of the discriminant P. Consider the region of all $(x,y) \in \mathbb{R}^2$ such that $|f(x,y)| \leq 1$ and $|g(x,y)| \leq 1$. If $RD(f)D(g) \neq 0$, then the area of the region is equal to the following.*

1. *If there exist $u_f, u_g \in \{-1,1\}$ such that both $f(x,y) = u_f$ and $g(x,y) = u_g$ represent ellipses, then the area is*

$$\frac{2}{\sqrt{-D(f)}} \arccos\left[\frac{D(f)-u_f u_g P}{\sqrt{4R}}\right] + \frac{2}{\sqrt{-D(g)}} \arccos\left[\frac{D(g)-u_f u_g P}{\sqrt{4R}}\right].$$

2. *If there exists a unit $u_f \in \{-1,1\}$ for which $f(x,y) = u_f$ is an ellipse and $g(x,y) = 1$ is a hyperbola, then the area depends on the number of intersection points.*

   - *If there exists a unit $u_g \in \{-1,1\}$ for which $f(x,y) = u_f$ intersects $g(x,y) = u_g$ but not $g(x,y) = -u_g$, then the area is*

$$\frac{2}{\sqrt{-D(f)}} \arccos\left[\frac{D(f)-u_f u_g P}{\sqrt{4R}}\right] + \frac{2}{\sqrt{D(g)}} \operatorname{arccosh}\left|\frac{D(g)-u_f u_g P}{\sqrt{4R}}\right|.$$

   - *If $f(x,y) = u_f$ intersects both $g(x,y) = 1$ and $g(x,y) = -1$, then the area is*

$$\frac{2}{\sqrt{-D(f)}} \left(\arccos\left[\frac{D(f)-u_f P}{\sqrt{4R}}\right] - \arccos\left[\frac{-D(f)+u_f P}{\sqrt{4R}}\right]\right) + \frac{2}{\sqrt{D(g)}} \left(\operatorname{arccosh}\left|\frac{D(g)-u_f P}{\sqrt{4R}}\right| + \operatorname{arccosh}\left|\frac{D(g)+u_f P}{\sqrt{4R}}\right|\right).$$

3. *If both $f(x,y) = 1$ and $g(x,y) = 1$ are hyperbolas, then the area depends on the number of intersection points (between the four hyperbolas $f(x,y) = \pm 1$ and $g(x,y) = \pm 1$) and the arrangement of the asymptotes.*

   - *If there exist $u_f, u_g \in \{-1,1\}$ for which the curve $f(x,y) = u_f$ intersects the curve $g(x,y) = u_g$, but $f(x,y) = -u_f$ intersects neither $g(x,y) = 1$ nor $g(x,y) = -1$, and the curve $g(x,y) = -u_g$ intersects $f(x,y) = 1$ nor $f(x,y) = -1$, then there are 4 intersection points and the area is*

$$\frac{2}{\sqrt{D(f)}} \operatorname{arccosh}\left|\frac{D(f)-u_f u_g P}{\sqrt{4R}}\right| + \frac{2}{\sqrt{D(g)}} \operatorname{arccosh}\left|\frac{D(g)-u_f u_g P}{\sqrt{4R}}\right|.$$

- *If there exists a unit $u \in \{-1,1\}$ such that the curve $f(x,y) = u$ intersects both $g(x,y) = 1$ and $g(x,y) = -1$, but $f(x,y) = -u$ intersects neither $g(x,y) = 1$ nor $g(x,y) = -1$, and if the asymptotes of $f(x,y) = 1$ lie "between" the asymptotes of $g(x,y) = 1$ (i.e. if rotating about the origin we pass in sequence to asymptotes of $g(x,y) = \pm 1$ followed by two asymptotes of $f(x,y) = \pm 1$, then recross those of $g(x,y) = \pm 1$), then the area is*

$$\frac{2}{\sqrt{D(f)}}\left(\operatorname{arccosh}\left|\frac{D(f)-uP}{\sqrt{4R}}\right| - \operatorname{arccosh}\left|\frac{D(f)+uP}{\sqrt{4R}}\right|\right) + \frac{2}{\sqrt{D(g)}}\left(\operatorname{arccosh}\left|\frac{D(g)-uP}{\sqrt{4R}}\right| + \operatorname{arccosh}\left|\frac{D(g)+uP}{\sqrt{4R}}\right|\right).$$

- *If each of $f(x,y) = 1$ and $f(x,y) = -1$ intersect both $g(x,y) = 1$ and $g(x,y) = -1$, then exactly one asymptote of $f(x,y) = 1$ lies "between" the two asymptotes of $g(x,y) = 1$ (any notion of "between" is valid here), and there is some choice of $u_1, u_2, u_3, u_4 \in \{-1,1\}$ for which the area is*

$$\frac{2}{\sqrt{|D(f)|}}\left(\operatorname{arccosh}\left|\frac{D(f)+u_1P}{\sqrt{4R}}\right| + u_2\operatorname{arccosh}\left|\frac{D(f)-u_1P}{\sqrt{4R}}\right|\right) + \frac{2}{\sqrt{|D(g)|}}\left(\operatorname{arccosh}\left|\frac{D(g)+u_3P}{\sqrt{4R}}\right| + u_4\operatorname{arccosh}\left|\frac{D(g)-u_3P}{\sqrt{4R}}\right|\right).$$

A secondary benefit of our technique, most notably the use of Lemma 2.1, is the area of union of the interiors of two ellipses. Specifically, the formula is a mere sign-change from the ellipse intersection formula found in Theorem 1.1 above. Note that the infinitude of area bounded by a hyperbola prohibits a full generalization of Theorem 1.1 for unions.

Theorem 1.2. *Let f and g be real definite binary quadratic forms with resultant R, discriminants $D(f)$ and $D(g)$, and polarization of the discriminant P. Consider the region of all $(x,y) \in \mathbb{R}^2$ such that either $|f(x,y)| \le 1$ or $|g(x,y)| \le 1$: this is a union of two ellipses. Assume that $RD(f)D(g) \neq 0$ and that there exist $u_f, u_g \in \{-1,1\}$ for which both $u_f f(x,y)$ and $u_g g(x,y)$ are positive definite. Then the area of the region is equal to the following.*

$$\frac{2}{\sqrt{-D(f)}}\arccos\left[\frac{-D(f)+u_fu_gP}{\sqrt{4R}}\right] + \frac{2}{\sqrt{-D(g)}}\arccos\left[\frac{-D(g)+u_fu_gP}{\sqrt{4R}}\right].$$

Note that $D_j > 0$ for hyperbolas, which arise from indefinite polynomials, and $D_j < 0$ for ellipses, which arise from definite polynomials. Hence we have

$$\frac{2}{\sqrt{-D_j}}\arccos\left[\frac{D_j+uP}{\sqrt{4R}}\right] = \frac{2}{\sqrt{D_j}}\operatorname{arccosh}\left|\frac{D_j+uP}{\sqrt{4R}}\right|.$$

Upper bounds for the areas of, and number of lattice points in, the intersection of hyperbolic regions are given in the paper [4], which focuses on approximations for the number of integer lattice points in these regions. The paper left open a question about a general formula for the exact area of these regions.

Our result helps to generalize the lattice point approximations of [4]. Indeed, using one of their results we give the following invariant bound for lattice points in the intersection and union of the interior of two ellipses:

Theorem 1.3. *Let f and g be real definite binary quadratic forms with resultant R, discriminants $D(f)$ and $D(g)$, and polarization of the discriminant P. Let E_f denote the region of all $(x,y) \in \mathbb{R}^2$ such that $|f(x,y)| \leq 1$ and E_g denote the region of all $(x,y) \in \mathbb{R}^2$ such that $|g(x,y)| \leq 1$; each of these is the interior of an ellipse. Assume that $RD(f)D(g) \neq 0$ and that there exist $u_f, u_g \in \{-1,1\}$ for which both $u_f f(x,y)$ and $u_g g(x,y)$ are positive definite. Let $N_\cap$ denote the number of integer lattice points inside $E_f \cap E_g$, and $N_\cup$ denote the number of lattice points inside $E_f \cup E_g$. Then we have the following two upper bounds:*

$$N_\cap \leq 2 + \frac{4}{\sqrt{-D(f)}} \arccos\left[\frac{D(f)-u_f u_g P}{\sqrt{4R}}\right] + \frac{4}{\sqrt{-D(g)}} \arccos\left[\frac{D(g)-u_f u_g P}{\sqrt{4R}}\right]$$

and

$$N_\cup \leq 2 + \frac{4}{\sqrt{-D(f)}} \arccos\left[\frac{-D(f)+u_f u_g P}{\sqrt{4R}}\right] + \frac{4}{\sqrt{-D(g)}} \arccos\left[\frac{-D(g)+u_f u_g P}{\sqrt{4R}}\right].$$

The point here is that these upper bounds are *invariant* under the action of $\mathrm{SL}_2(\mathbb{Z})$. It is conjectured that these formulas could be improved to resemble those found for hyperbolic regions in [4]. Indeed, the lattice point calculations of [4] serve a higher purpose of calculating asymptotic formulas for the height-counting function of the rational curves in [5], in which the author states that “a strengthening of this prior result would also strengthen [the main theorem].”

2. Preliminaries

Consider two nonsingular real quadratic forms (i.e. homogenous polynomials) f_1 and f_2, written as $f_i(x,y) = a_i x^2 + b_i xy + c_i y^2$. Each of these give rise to one ellipse or two quadratic curves $C_i^\pm : f_i(x,y) = \pm 1$. Note that when f_i is definite only one choice of $\pm$ gives rise to an ellipse $C_i^\pm$ (while $C_i^\mp$ is empty), but when f_i is indefinite, both choices of $\pm$ will give rise to hyperbolas $C_i^\pm$.

We will use the following invariants throughout our computations.

Definition 1. For $i \in \{1,2\}$, let $f_i(x,y) = a_i x^2 + b_i xy + c_i y^2$ denote quadratic forms over any field.

1. The <u>discriminant</u> $D_i = D(f_i)$ of f_i is the coefficient polynomial $D_i = b_i^2 - 4a_i c_i$.
2. The <u>polarization</u> (of the discriminant) $P = P(f_1, f_2)$ of the forms is the coefficient polynomial $P = b_1 b_2 - 2a_1 c_2 - 2a_2 c_1$.
3. The <u>resultant</u> $R = R(f_1, f_2)$ of the forms is the determinant of the Sylvester matrix of f_1 and f_2.

The definitions of discriminant and resultant are standard, and the definition of polarization can be found in [3]. For an alternate formulation in terms of determinants involving the linear factors of f_1 and f_2, see [4] or [5]. The resultant satisfies the following remarkable syzygy, which simplifies our calculations:

$$4R = P^2 - D_1 D_2.$$

Observe that these coefficient polynomials are invariant under a matrix transformation $\mathbb{A}$ acting on $[x,y]$, with weights 2 or 4 as follows: $D(f_i \circ \mathbb{A}) = D(f_i)\det(\mathbb{A})^2$, $P(f_1 \circ \mathbb{A}, f_2 \circ \mathbb{A}) = P(f_1, f_2)\det(\mathbb{A})^2$, and $R(f_1 \circ \mathbb{A}, f_2 \circ \mathbb{A}) = R(f_1, f_2)\det(\mathbb{A})^4$. Note further that there is a nice correspondence under negation.

$$\begin{aligned} D(\pm f_i) &= D(f_i) \\ P(\pm_1 f_1, \pm_2 f_2) &= \pm_1 (\pm_2 P(f_1, f_2)) \\ R(\pm_1 f_1, \pm_2 f_2) &= R(f_1, f_2) \end{aligned} \tag{2.1}$$

We note that in order for C_i^+ or C_i^- to be an ellipse we need $D_i < 0$. In order for C_i^+ and C_i^- to be hyperbolas we need $D_i > 0$.

In polar coordinates our conics are given by the following equations.

$$\begin{aligned} C_1^{u_1} : (r_1(\theta))^2 &= \tfrac{u_1 \sec^2\theta}{a_1 + b_1 \tan\theta + c_1 \tan^2\theta} \\ C_2^{u_2} : (r_2(\theta))^2 &= \tfrac{u_2 \sec^2\theta}{a_2 + b_2 \tan\theta + c_2 \tan^2\theta} \end{aligned} \tag{2.2}$$

where u_1 and u_2 are discrete variables in $\{-, +\}$.

To find the angles at which our conics intersect we set $(r_1(\theta))^2, (r_2(\theta))^2$ equal giving us the two quadratic equations

$$(u_2c_2 - u_1c_1)\tan^2\theta + (u_2b_2 - u_1b_1)\tan\theta + (u_2a_2 - u_1a_1) = 0. \tag{2.3}$$

We simplify this expression by writing $a = u_2a_2 - u_1a_1, b = u_2b_2 - u_1b_1$, and $c = u_2c_2 - u_1c_1$.

Employing the quadratic formula we have at most two pairs of tangent values:

$$\tan\theta_{i,\pm} = \frac{-b \pm \sqrt{(D_1 + D_2) - 2u_1u_2P}}{2c}.$$

Each intersection angle ϕ at which our conics intersect corresponds to another angle of intersection $\phi + \pi$ of our two conics because of the periodicity of the tangent function. Hence at most 8 angles of intersection are possible. Furthermore a glance at some examples shows us that our two conics will not always intersect at eight points; in cases where our conics intersect in four points, the variables u_1 and u_2 in Equation 2.3 are constant.

Lemma 2.1. *Consider curves C_1 and C_2 each of which is either an ellipse or hyperbola centered at the origin. Assume further that C_1 and C_2 intersect in exactly 4 points. Let A_{in} denote the arcs of C_1 which intersect C_2 at its endpoints and is closer to the origin than C_2 (viewed along rays from the origin), and let A_{out} denote the remaining arcs of U, which are further from the origin than is C_2.*

For each i, let f_i denote the real quadratic form for which C_i is the curve $f_i(x,y) = 1$. Let D_i denote the discriminant of f_i, P the polarization (of the discriminant) of f_1 and f_2, and R the resultant of f_1 and f_2. If R is nonzero, then

1. *If C_1 is an ellipse, then the area between the arc A_{in} and the origin, and the area between A_{out} and the origin, are respectively*

$$\frac{2}{\sqrt{-D_1}} \arccos\left[\frac{D_1 - P}{\sqrt{4R}}\right]$$

or

$$\frac{2}{\sqrt{-D_1}} \arccos\left[\frac{-D_1+P}{\sqrt{4R}}\right].$$

2. *If C_1 is a hyperbola, then exactly one of A_{in} and A_{out} is a connected arc; the area between this arc and the origin is exactly*

$$\frac{2}{\sqrt{D_1}} \operatorname{arccosh}\left|\frac{D_1-P}{\sqrt{4R}}\right|.$$

Because some of the techniques used within the proof of Lemma 2.1 will be reused, we now divert our attention to its proof.

3. Proof of Lemma 2.1

For $i=1$ and $i=2$ let $f_i = a_i x^2 + b_i xy + c_i y^2$ denote a real quadratic form with nonzero discriminant D_i, and let C_i denote the curve $f_i(x,y)=1$. Assume further that the pair f_1 and f_2 have nonzero resultant R and have polarization P, and that $|C_1 \cap C_2| = 4$.

Let θ_1 and θ_2 denote a pair of adjacent intersection points of C_1 and C_2. Via rotation, we assume without loss of generality that $-\pi/2 < \theta_1 < 0 < \theta_2 < \pi/2$. We recall Equation 2.2, the formula for C_1 in polar coordinates, and manipulate it to yield

$$(r(\theta))^2 = \frac{\frac{\sqrt{|D_1|}}{2\sqrt{c_1}}(b_1\sec^2\theta + 2c_1\sec^2\theta\tan\theta)(a_1+b_1\tan\theta+c_1\tan^2\theta)^{-\frac{3}{2}}}{\sqrt{|D_1|}\sqrt{1+\frac{D_1}{4c_1(a_1+b_1\tan\theta+c_1\tan^2\theta)}}}.$$

We rewrite this formula using $v := \sqrt{\dfrac{|D_1|}{4c_1(a_1+b_1\tan\theta+c_1\tan^2\theta)}}$.

Then $r^2(\theta)\, d\theta = \begin{cases} \frac{-2\ dv}{\sqrt{-D_1(1-v^2)}}, & \text{if } C_1 \text{ is an ellipse;} \\ \frac{-2\ dv}{\sqrt{D_1(1+v^2)}}, & \text{if } C_1 \text{ is a hyperbola.} \end{cases}$

For simplicity of notation, we set v_i as $\sqrt{D_1/\left(4c_1(a_1+b_1\tan\theta_i+c_1\tan^2\theta_i)\right)}$ if C_1 is an ellipse or $\sqrt{-D_1/\left(4c_1(a_1+b_1\tan\theta_i+c_1\tan^2\theta_i)\right)}$ if C_1 is a hyperbola. Then the arc area between C_1 and the origin, with angle between θ_1 and θ_2, is exactly the following.

$$\begin{aligned} \text{Area} &= \int_{\theta_1}^{\theta_2}\int_0^r \rho\, d\rho\ \ d\theta = \frac{1}{2}\int_{\theta_1}^{\theta_2} r^2\, d\theta = \frac{-1}{|D_1|}\int_{v_1}^{v_2}\frac{dv}{\sqrt{1\mp v^2}} \\ &= \begin{cases} -\frac{1}{\sqrt{-D_1}}\arcsin v\Big|_{v_1}^{v_2}, & \text{if } C_1 \text{ is an ellipse;} \\ -\frac{1}{\sqrt{D_1}}\operatorname{arcsinh} v\Big|_{v_1}^{v_2}, & \text{if } C_1 \text{ is a hyperbola.} \end{cases} \end{aligned} \tag{3.4}$$

Note that our restriction $-\pi/2 < \theta_1 < \theta_2 < \pi/2$ guarantees that $v \neq 0$ and therefore that we do not cross the y-axis branch cut of arcsin; note that arcsinh has no branch cut. Indeed we avoid using arccos and arccosh due to the branch cuts of $y = \frac{-b_1 x}{2c_1}$.

Note that $\arcsin v_2 - \arcsin v_1 = \pm\arccos\left(v_1v_2 + \sqrt{(1-v_1^2)(1-v_2^2)}\right)$ allows us to combine terms in the elliptic case, while in the hyperbolic case we use $\operatorname{arcsinh} v_2 - \operatorname{arcsinh} v_1 = \pm\operatorname{arccosh}\left(\sqrt{(v_1^2+1)(v_2^2+1)} - v_1v_2\right)$. Recalling our definitions for v_1 and v_2 while noting that areas are necessarily positive, we arrive at two related formulas with a common argument. Specifically, in the case when C_1 is an ellipse we have

$$\text{Area} = \frac{1}{\sqrt{-D_1}}\arccos\left(\frac{2c_1\tan\theta_1\tan\theta_2 + b_1(\tan\theta_1+\tan\theta_2)+2a_1}{2\sqrt{(a_1+b_1\tan\theta_1+c_1\tan^2\theta_1)(a_1+b_1\tan\theta_2+c_1\tan^2\theta_2)}}\right), \tag{3.5}$$

and when C_1 is a hyperbola we have

$$\text{Area} = \frac{1}{\sqrt{-D_1}}\operatorname{arccosh}\left(\frac{2c_1\tan\theta_1\tan\theta_2 + b_1(\tan\theta_1+\tan\theta_2)+2a_1}{2\sqrt{(a_1+b_1\tan\theta_1+c_1\tan^2\theta_1)(a_1+b_1\tan\theta_2+c_1\tan^2\theta_2)}}\right). \tag{3.6}$$

Note that $X = \tan\theta_1$ and $X = \tan\theta_2$ are roots of the polynomial found in Equation 2.3.

$$(c_2-c_1)X^2 + (b_2-b_1)X + (a_2-a_1) = 0 \tag{3.7}$$

Hence $\frac{-(b_2-b_1)}{c_2-c_1} = \tan\theta_1 + \tan\theta_2$, and $\frac{a_2-a_1}{c_2-c_1} = \tan\theta_1\tan\theta_2$.

Let u denote the sign of $c_2 - c_1$. Recalling our assumption that $\theta_1 < 0 < \theta_2$, we note that $\frac{a_2-a_1}{c_2-c_1} = \tan\theta_1\tan\theta_2 < 0$. Therefore the sign of $a_2 - a_1$ is $-u$. Since $1/\sqrt{a_1}$ is the x-intercept of C_1, clearly we know that $a_1 > 0$. Furthermore, if C_2 is closer to the origin on the arc from θ_1 to θ_2 then we know that $a_2 > a_1$; in this case $u = -$. In the remaining case we have $a_1 > a_2$ and we say that C_1 is closer to the origin (even if C_2 is a hyperbola crossing an asymptote on this arc); in this case $u = +$. We conclude that u is positive if and only if C_1 is closer to the origin than C_2.

After substitution, the argument of Equations 3.5 and 3.6 becomes

$$u\left(\frac{2c_1(a_2-a_1) - b_1(b_2-b_1) + 2a_1(c_2-c_1)}{2\sqrt{(c_2-c_1)^2(a_1+b_1\tan\theta_1+c_1\tan^2\theta_1)(a_1+b_1\tan\theta_2+c_1\tan^2\theta_2)}}\right).$$

After rearrangement and involvement of Definition 1, and doubling for the symmetric arc from $\theta_1+\pi$ to $\theta_2+\pi$, we come to the following concise formula for the area.

$$\text{Area} = \begin{cases} \frac{2}{\sqrt{-D_1}}\arccos\left(u\frac{D_1-P}{\sqrt{4R}}\right), & \text{if } C_1 \text{ is an ellipse;} \\ \frac{2}{\sqrt{D_1}}\operatorname{arccosh}\left(u\frac{D_1-P}{\sqrt{4R}}\right), & \text{if } C_1 \text{ is a hyperbola.} \end{cases}$$

In the hyperbolic case, we note that the real parts of $\operatorname{arccosh}(x)$ and $\operatorname{arccosh}(-x)$ are equal, and indeed their real value is exactly $\operatorname{arccosh}|x|$. Hence for the correct real area of a hyperbolic arc we choose $\frac{2}{\sqrt{D_1}}\operatorname{arccosh}\left|\frac{D_1-P}{\sqrt{4R}}\right|$, regardless of the value of u.

This completes the proof of the desired formula.

4. Case-by-Case Proof for Each Configuration

Lemma 2.1 is the key ingredient to computing the area for the configurations in Sections 4.1, 4.2, and 4.3. For the alternating asymptote configuration of Section 4.4, we use a similar method of proof to that found within Lemma 2.1.

The following arcco_i notation will provide a convenient way of writing the formulas for ellipses and hyperbolas in the same form.

Definition 2. The $\operatorname{arcco}_i x = \operatorname{arcco}(C_i^{\pm}, x)$ function takes the values of arccos or arccosh depending, respectively, on whether $C_i^{\pm}$ is an ellipse or a pair of hyperbolas.

$$\operatorname{arcco}_i x := \begin{cases} \arccos x, & \text{if } C_i^{\pm} \text{ is an ellipse;} \\ \operatorname{arccosh} |x|, & \text{if } C_i^{\pm} \text{ is a hyperbola.} \end{cases}$$

We are ready to complete the proof of the area formulas. We will consider each configuration of curves separately, dividing into four canonical cases. Note that several of our conclusions are direct corollaries of the lemma.

4.1. Four Intersection Points

Consider the case when the four curves $C_1^{\pm}$ and $C_2^{\pm}$ have exactly four intersection points. These points necessarily come from one curve $C_1^{\pm}$ and one curve $C_2^{\pm}$, each with independently fixed choices of $\pm$. This could take the form of two ellipses, an ellipse an a hyperbola, or two hyperbolas; these configurations are demonstrated in the following graphs in Figure 2 below.

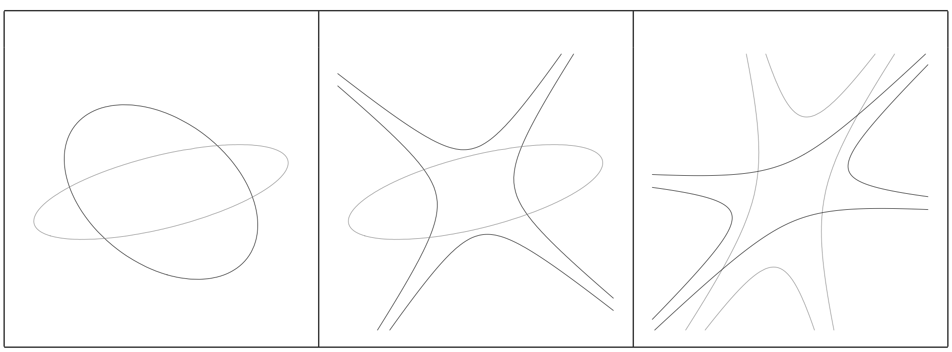

Figure 2. Configurations intersecting in four points.

Indeed, if both choices are $+$, then the formulas from Lemma 2.1 hold and the area of intersection is exactly

$$\frac{2}{\sqrt{|D_1|}} \operatorname{arcco}_1 \left[\frac{D_1 - P}{\sqrt{4R}} \right] + \frac{2}{\sqrt{|D_2|}} \operatorname{arcco}_2 \left[\frac{D_2 - P}{\sqrt{4R}} \right]. \tag{4.8}$$

where $\operatorname{arcco}_i x$ is the notational convenience from Definition 2 which represents either arccos or arccosh depending respectively on whether we are considering an ellipse or a hyperbola. If any choices are $-$, then we recall from Equation 2.1 that $P(\pm_1 f_1, \pm_2 f_2) =$

$\pm_1(\pm_2 P(f_1, f_2))$ while the invariants D and R remain unchanged. Hence, if both choices are $-$, then Equation 4.8 gives the area, and if the choices are opposite (i.e. one is $+$ while the other is $-$), then and the area is

$$\frac{2}{\sqrt{|D_1|}}\operatorname{arcco}_1\left[\frac{D_1+P}{\sqrt{4R}}\right]+\frac{2}{\sqrt{|D_2|}}\operatorname{arcco}_2\left[\frac{D_2+P}{\sqrt{4R}}\right]. \tag{4.9}$$

Furthermore, if C_1 and C_2 are both ellipses, then we can calculate the union of their enclosed areas using Lemma 2.1 by reversing the sign of the argument of $\operatorname{arcco}_i x$. Specifically, the area of the union in the case when both signs are positive or negative (respectively) is

$$\frac{2}{\sqrt{|D_1|}}\operatorname{arcco}_1\left[\frac{-D_1+P}{\sqrt{4R}}\right]+\frac{2}{\sqrt{|D_2|}}\operatorname{arcco}_2\left[\frac{-D_2+P}{\sqrt{4R}}\right], \tag{4.10}$$

or

$$\frac{2}{\sqrt{|D_1|}}\operatorname{arcco}_1\left[\frac{-D_1-P}{\sqrt{4R}}\right]+\frac{2}{\sqrt{|D_2|}}\operatorname{arcco}_2\left[\frac{-D_2-P}{\sqrt{4R}}\right]. \tag{4.11}$$

This completes the proof of Theorem 1.2. Furthermore, we have proven part 1 and the first bullets of parts 2 and 3 of Theorem 1.1.

4.2. Eight Intersection Points: Ellipse and Hyperbolas

Consider the case when $C_1^{\pm}$ is a pair of hyperbolas and C_2^{+} is an ellipse which intersects $C_1^{\pm}$ in 8 points total. This configuration is demonstrated in the following graph in Figure 3.

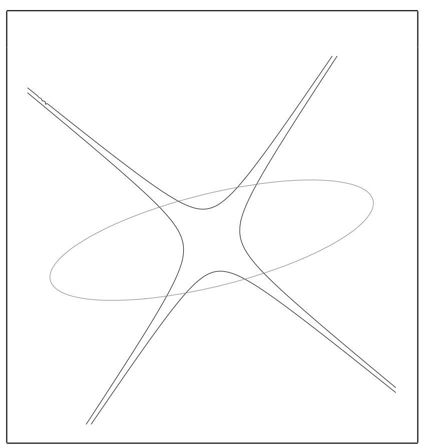

Figure 3. Ellipse and symmetric hyperbolas, intersecting in 8 points.

As seen previously in Section 4.1, the area of intersection of the regions bounded by the curves C_1^{+} and C_2^{+} is exactly

$$\frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1-P}{\sqrt{4R}}\right|+\frac{2}{\sqrt{|D_2|}}\arccos\left[\frac{D_2-P}{\sqrt{4R}}\right],$$

which overcounts the intended intersection area by the finite arc between the intersection points of C_1^{-} and C_2^{+}. By Lemma 2.2, the area of this arc along C_2^{+} is $\frac{2}{\sqrt{|D_2|}}\arccos\left[\frac{-D_2-P}{\sqrt{4R}}\right]$, while the area of the arc along C_1^- is $\frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1+P}{\sqrt{4R}}\right|$. Hence the total area overcounted is exactly their difference,

$$\frac{2}{\sqrt{|D_2|}}\arccos\left[\frac{-D_2-P}{\sqrt{4R}}\right]-\frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1+P}{\sqrt{4R}}\right|.$$

The total area of intersection can now be computed as:

$$\frac{2\left(\operatorname{arccosh}\left|\frac{D_1-P}{\sqrt{4R}}\right|+\operatorname{arccosh}\left|\frac{-D_1-P}{\sqrt{4R}}\right|\right)}{\sqrt{|D_1|}}+\frac{2\left(\arccos\left[\frac{D_2-P}{\sqrt{4R}}\right]-\arccos\left[\frac{-D_2-P}{\sqrt{4R}}\right]\right)}{\sqrt{|D_2|}}.$$

If C_2^- is an ellipse then the sign of the polarization is reversed and the area of intersection can be found similarly.

$$\frac{2\left(\operatorname{arccosh}\left|\frac{D_1+P}{\sqrt{4R}}\right|+\operatorname{arccosh}\left|\frac{-D_1+P}{\sqrt{4R}}\right|\right)}{\sqrt{|D_1|}}+\frac{2\left(\arccos\left[\frac{D_2+P}{\sqrt{4R}}\right]-\arccos\left[\frac{-D_2+P}{\sqrt{4R}}\right]\right)}{\sqrt{|D_2|}}$$

This completes the proof of part 2 of Theorem 1.1.

4.3. Eight Intersection Points: Hyperbolas with Adjacent Asymptotes

Consider two hyperbolas C_1^+ and C_2^+, for which between two asymptotes of C_1^+ the curve C_2^+ has no asymptotes (note that any sense of "between" applies here). Figure 4 illustrates this configuration.

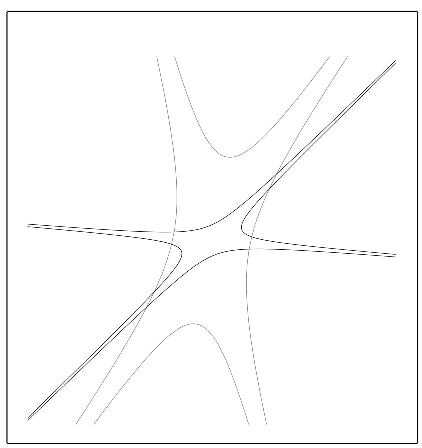

Figure 4. Hyperbolas, adjacent asymptotes.

We first consider the case when either (a) both C_1^+ and C_2^+ or (b) both C_1^- and C_2^- contain a bounded region. This configuration is virtually identical to that in the preceding section, with the ellipse replaced by one hyperbola. Indeed, the area of this region is the familiar formula,

$$\frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1-P}{\sqrt{4R}}\right|+\frac{2}{\sqrt{|D_2|}}\operatorname{arccosh}\left|\frac{D_2-P}{\sqrt{4R}}\right|$$

and the area overcounted is

$$\pm\left(\frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1+P}{\sqrt{4R}}\right| - \frac{2}{\sqrt{|D_2|}}\operatorname{arccosh}\left|\frac{D_2+P}{\sqrt{4R}}\right|\right),$$

where the choice of $\pm$ guarantees that this overcounted area is positive; indeed, it is positive exactly when C_1 is closest to the origin on the overcounted region.

As in the previous section, we note that changing the initial conditions to begin with a finite region bounded by either (a) both C_1^- and C_2^+ or (b) both C_1^+ and C_2^- merely changes the sign of the polarization P in the final result. Thus the area is either the sum of the above formulas,

$$\frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1-P}{\sqrt{4R}}\right| + \frac{2}{\sqrt{|D_2|}}\operatorname{arccosh}\left|\frac{D_2-P}{\sqrt{4R}}\right|$$
$$\pm\left(\frac{2}{\sqrt{|D_2|}}\operatorname{arccosh}\left|\frac{D_2+P}{\sqrt{4R}}\right| - \frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1+P}{\sqrt{4R}}\right|\right),$$

or the same formula with the sign in front of P reversed,

$$\frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1+P}{\sqrt{4R}}\right| + \frac{2}{\sqrt{|D_2|}}\operatorname{arccosh}\left|\frac{D_2+P}{\sqrt{4R}}\right|$$
$$\pm\left(\frac{2}{\sqrt{|D_2|}}\operatorname{arccosh}\left|\frac{D_2-P}{\sqrt{4R}}\right| - \frac{2}{\sqrt{|D_1|}}\operatorname{arccosh}\left|\frac{D_1-P}{\sqrt{4R}}\right|\right).$$

This completes the proof of the second bullet of part 3 of Theorem 1.1.

4.4. Eight Intersection Points: Hyperbolas with Alternating Asymptotes

Consider two hyperbolas C_1^+ and C_2^+, for which between two asymptotes of C_1^+ the curve C_2^+ has exactly one asymptote. We assume further that the y-intercepts of $C_2^\pm$ are closer to 0 than those of $C_1^\pm$ (via rotation we assume without loss of generality that these y-intersects exist). This final configuration is shown in Figure 5 below, with the darker symmetric pair of hyperbolas representing $C_2^\pm$.

For this case the results of Lemma 2.1 are not useful: a choice of $C_1^\pm$ and a choice of $C_2^\pm$ does not yield a bounded region. However, our proof in this case will be similar to that of Lemma 2.1.

Consider the increasing sequence of intersection angles in the right half plane, $-\pi/2 < \theta_1 < \theta_2 < \theta_3 < \theta_4 < \pi/2$, which via rotation we assume without loss of generality that no angle is equal to $\pm\pi/2$. Note that each angle θ has a corresponding angle $\theta+\pi$ in the left half-plane. By our assumption that the y-intercepts of $C_2^\pm$ are closer to 0 than those of $C_1^\pm$, we conclude that $C_1^\pm$ is closer to the origin than $C_2^\pm$ on the arcs from θ_1 to θ_2 and from θ_3 to θ_4. We assume further that C_1^+ is closest along the arc in question from θ_1 to θ_2, and thus that C_1^- is closest from θ_3 to θ_4. Lastly, we assume that C_2^+ crosses the y-axis, and hence that the relevant arc of C_2^- is the one from θ_2 to θ_3. This configuration is shown in the following labeled Figure 5.

We now recall Equation 2.2, the polar coordinate representation, to write the area of the

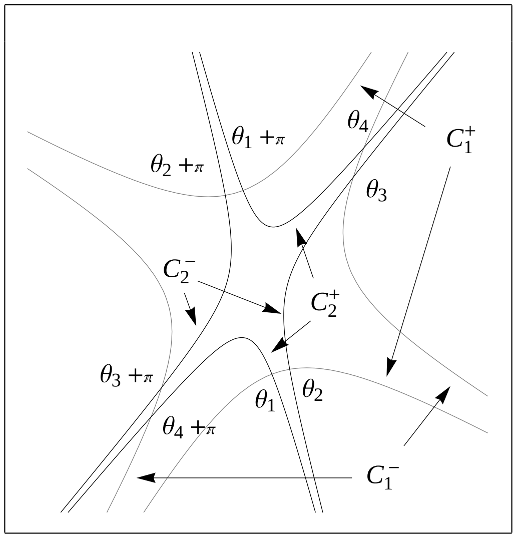

Figure 5. Hyperbolas, alternating asymptotes.

two arcs from θ_1 to θ_2 and θ_3 to θ_4:

$$\begin{aligned}\text{Area} &= \frac{1}{2}\left(\int_{\theta_1}^{\theta_2} r^2\,d\theta + \int_{\theta_3}^{\theta_4} r^2\,d\theta\right)\\ &= \frac{1}{2}\left(\int_{\theta_1}^{\theta_2}\left|\frac{\sec^2(\theta)}{a_1+b_1\tan\theta+c_1\tan^2\theta}\right|\,d\theta + \int_{\theta_3}^{\theta_4}\left|\frac{\sec^2(\theta)}{a_1+b_1\tan\theta+c_1\tan^2\theta}\right|\,d\theta\right).\end{aligned}$$

Note that the absolute value leaves the integrand unchanged on the interval $[\theta_1,\theta_2]$ along C_2^+ and reverses the sign on the interval $[\theta_3,\theta_4]$ along C_1^-. This, together with the antiderivative given in Equation 3.4 during the proof of Lemma 2.1, we arrive at the equation

$$\text{Area} = -\frac{1}{\sqrt{D_1}}\operatorname{arcsinh}\left(\sqrt{\frac{D_1}{4c_1(a_1+b_1\tan\theta+c_1\tan^2\theta)}}\right)\Bigg|_{\theta_1}^{\theta_3}$$

$$+\frac{1}{\sqrt{D_1}}\operatorname{arcsinh}\left(\sqrt{\frac{D_1}{4c_1(a_1+b_1\tan\theta+c_1\tan^2\theta)}}\right)\Bigg|_{\theta_4}^{\theta_2}.$$

As in the computation in the proof of Lemma 2.1, we have the difference formula $\operatorname{arcsinh} v_i - \operatorname{arcsinh} v_j = \pm\operatorname{arccosh}\left(\sqrt{(v_i^2+1)(v_j^2+1)} - v_i v_j\right)$. Setting v_i as $\sqrt{-D_1/(4c_1(a_1+b_1\tan\theta_i+c_1\tan^2\theta_i))}$ yields

$$\frac{\pm\operatorname{arccosh}\left(\frac{2c_1\tan\theta_1\tan\theta_3+b_1(\tan\theta_1+\tan\theta_3)+2a_1}{2\sqrt{(a_1+b_1\tan\theta_1+c_1\tan^2\theta_1)(a_1+b_1\tan\theta_3+c_1\tan^2\theta_3)}}\right)}{\sqrt{D_1}} \tag{4.12}$$

for the first term, involving θ_1 and θ_3, and a corresponding formula for the second, involving θ_2 and θ_4; note that the $\pm$ is dependent on which of $\operatorname{arcsinh} v_i$ and $\operatorname{arcsinh} v_j$ is larger (for $(i,j)\in\{(1,3),(2,4)\}$).

Recall that C_1^+ and C_2^+ intersect at θ_1 and that C_1^- and C_2^- intersect at θ_3. Hence $X = \tan\theta_1$ and $X = \tan\theta_3$ are roots of the polynomial of Equation 2.3 with $u_1 = u_2$: $(c_2 - c_1)X^2 + (b_2 - b_1)X + (a_2 - a_1) = 0$. Similarly, $X = \tan\theta_2$ and $X = \tan\theta_4$ are roots of the polynomial of Equation 2.3 with $u_1 = -u_2$: $(c_2 + c_1)X^2 + (b_2 + b_1)X + (a_2 + a_1) = 0$. Hence $\tan\theta_1 + \tan\theta_3 = \frac{-(b_2-b_1)}{c_2-c_1}$, $\tan\theta_1 \tan\theta_3 = \frac{a_2-a_1}{c_2-c_1}$, $\tan\theta_2 + \tan\theta_4 = \frac{-(b_2+b_1)}{c_2+c_1}$, and $\tan\theta_2 \tan\theta_4 = \frac{a_2+a_1}{c_2+c_1}$.

In the proof of Lemma 2.1 the sign of $c_2 \pm c_1$ was important; by rotation we assume without loss of generality that it is nonzero. Since the y-intercepts of C_2^+ are closer to the origin than those of $C_1^\pm$, we have $|c_2| > |c_1|$ and $c_2 > 0$; hence $c_2 \pm c_1$ is positive. After substitution and rearrangement using this newly determined fact, the argument of Equation 4.12 becomes

$$\frac{2c_1(a_2 \pm a_1) - b_1(b_2 \pm b_1) + 2a_1(c_2 \pm c_1)}{2\sqrt{(c_2 \pm c_1)^2(a_1 + b_1\tan\theta_i + c_1\tan^2\theta_i)(a_1 + b_1\tan\theta_j + c_1\tan^2\theta_j)}},$$

for $(i, j, \pm) \in \{(1,3,-),(2,4,+)\}$. After rearrangement, invoking Definition 1, and doubling for the symmetric arc from $\theta_i + \pi$ to $\theta_j + \pi$, we come to the following concise formula for the area:

$$\text{Area} = \pm\frac{2}{\sqrt{D_1}} \operatorname{arccosh}\left(\frac{D_1 - P}{\sqrt{4R}}\right) \pm \frac{2}{\sqrt{D_1}} \operatorname{arccosh}\left(\frac{D_1 + P}{\sqrt{4R}}\right).$$

Since the area is ultimately positive, it is clear that the formula is truly

$$\text{Area} = \frac{2}{\sqrt{D_1}} \operatorname{arccosh}\left(\frac{D_1 \pm_1 P}{\sqrt{4R}}\right) \pm_2 \frac{2}{\sqrt{D_1}} \operatorname{arccosh}\left(\frac{D_1 \mp_1 P}{\sqrt{4R}}\right).$$

The area of the arcs along $C_2^\pm$ follows the same formula (up to the choice of $\pm$), since the formula is invariant under the group of rotations. Hence the total area follows a similar formula to that found in the previous section:

$$\frac{2}{\sqrt{|D_1|}}\left(\operatorname{arccosh}\left|\frac{D_1 \pm_1 P}{\sqrt{4R}}\right| \pm_2 \operatorname{arccosh}\left|\frac{D_1 \mp_1 P}{\sqrt{4R}}\right|\right)$$
$$+\frac{2}{\sqrt{|D_2|}}\left(\operatorname{arccosh}\left|\frac{D_2 \pm_3 P}{\sqrt{4R}}\right| \pm_4 \operatorname{arccosh}\left|\frac{D_2 \mp_3 P}{\sqrt{4R}}\right|\right).$$

Fixing exactly one of $\pm_2$ or $\pm_4$ to be positive (while the other is possibly negative), we note that the authors have found curves with every possible combination of choices of $\pm_1$ and $\pm_3$; indeed replacing f_1 by $-f_1$ or f_2 by $-f_2$ merely changes the sign of the polarization. However, it is an open question whether there exist curves with $\pm_2 = - = \pm_4$.

This completes the proof of the final case of Theorem 1.1.

5. Proof of Theorem 1.3

We return to the case of two ellipses, which we shall denote $f = u_f$ and $g = u_g$, for some $u_f, u_g \in \{-1, 1\}$, with respective interiors by E_f and E_g.

In order to prove Theorem 1.3, we recall a result of the third author: Lemma 2.2 in [4].

Lemma 5.2. *Let P be a closed convex plane region symmetric about the origin with area A and containing L primitive integral lattice points. Then $L \leq 2A+2$.*

Recall from Section 4.1 that we computed the areas of $E_f \cap E_g$ and $E_f \cup E_g$ in Equations 4.8-4.11. We recall the formulas here.

$$\left|E_f \cap E_g\right| = \frac{2}{\sqrt{-D(f)}} \arccos\left[\frac{D(f)-u_f u_g P}{\sqrt{4R}}\right] + \frac{2}{\sqrt{-D(g)}} \arccos\left[\frac{D(g)-u_f u_g P}{\sqrt{4R}}\right]$$

$$\left|E_f \cup E_g\right| = \frac{2}{\sqrt{-D(f)}} \arccos\left[\frac{-D(f)+u_f u_g P}{\sqrt{4R}}\right] + \frac{2}{\sqrt{-D(g)}} \arccos\left[\frac{-D(g)+u_f u_g P}{\sqrt{4R}}\right]$$

Coupled with Lemma 5.2, we obtain the desired result.

$$\left|\mathbb{Z}^2 \cap E_f \cap E_g\right| \leq 2 + \frac{4}{\sqrt{-D(f)}} \arccos\left[\frac{D(f)-u_f u_g P}{\sqrt{4R}}\right] + \frac{4}{\sqrt{-D(g)}} \arccos\left[\frac{D(g)-u_f u_g P}{\sqrt{4R}}\right]$$

$$\left|\mathbb{Z}^2 \cap (E_f \cup E_g)\right| \leq 2 + \frac{4}{\sqrt{-D(f)}} \arccos\left[\frac{-D(f)+u_f u_g P}{\sqrt{4R}}\right] + \frac{4}{\sqrt{-D(g)}} \arccos\left[\frac{-D(g)+u_f u_g P}{\sqrt{4R}}\right]$$

6. Conclusion

It is interesting that the literature is apparently devoid of exact formulas for intersections of such simple regions as these ellipses and hyperbolas. Indeed, it is an open question for the remaining (nonhomogeneous) conic, not considered here: the parabola. The third author is especially interested in the following open question: given degree d homogenous polynomials $f(x,y)$ and $g(x,y)$ with real (respectively integer) coefficients, what is the area of (resp. number of integer lattice points in) the region formed by the intersection of the interiors of the regions $|f(x,y)| = 1$ and $|g(x,y)| = 1$ (resp, for given real B, $|f(x,y)| = B$ and $|g(x,y)| = B$). Indeed, such a result would generalize the results found in [4], which would in turn improve on the results of [5].

The invariants D_1, D_2, and P found within this paper are likely to generate the ring of invariants for real binary quadratic forms; indeed they generate the fourth invariant found here via the remarkable syzygy $R = P^2 - 4D_1D_2$. While Hilbert's Basis Theorem guarantees a finite basis (with a higher dimension expected upon increasing the degree), it is an open question what the actual basis is.

As mentioned in the introduction, there is apparent need of quick algorithms (such as the one found in [2]) for computing the area bounded by multiple ellipses. Our exact formulas are likely to be more efficient than an algorithm. However, an explicit formula is unknown for the area of intersection of ellipses with arbitrary centers. Furthermore, intersection formulas for three or more conic regions may also be applicable.

References

[1] D. Eberly, *The Area of Intersecting Ellipses*, http://www.geometrictools.com (2008).

[2] G. Hughes and M. Chraibi, *Calculating Ellipse Overlap Areas*, arXiv:1106.3787v1 [physics.comp-ph] (2011).

[3] P. J. Olver, *Classical invariant theory*, 5th edn. Cambridge University Press, Cambridge, 1999.

[4] P. X. Rault, On uniform bounds for lattice points in intersections of hyperbolic plane regions, *J. Comb. Number Theory* 2 (3) (2010), 209–215.

[5] P. X. Rault, On uniform bounds for rational points on rational curves and thin sets, *JP J. Algebra Number Theory Appl.* 23 (2) (2011), 171–185.

MA