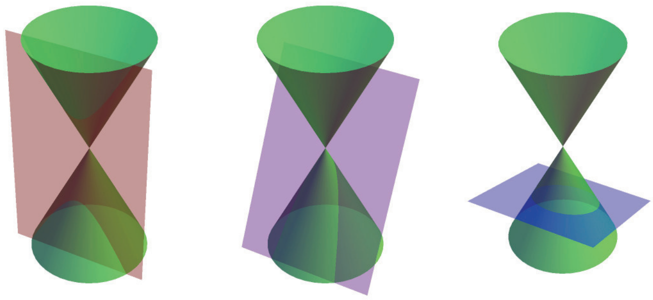
Conic sections

Conics have recurred throughout this book in both geometric and algebraic settings. Hence, I have decided to dedicate the final chapter to them. As the only conics appearing on IMO geometry problems are invariably circles, the results proved in this chapter are largely irrelevant. Nevertheless, the material is sufficiently interesting to be worthy of inclusion.

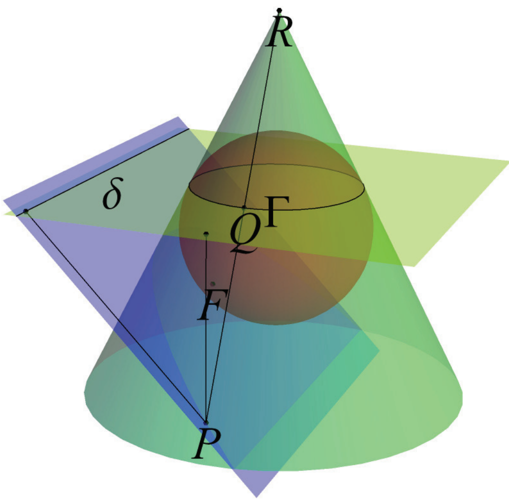
Sections of cones

With the obvious exception of the circle, the conics were first discovered by the Greek mathematician Menaechmus who contemplated slicing a right circular cone C with a flat plane Λ. Indeed, the term 'conic' is an abbreviation of *conic section*.



It is more natural to consider C as the double cone with equation $x^2 + y^2 = z^2$. If Λ cuts both cones, then the conic section is a *hyperbola*. If it cuts only one cone in a closed curve, it is an *ellipse*. The intermediate case, where the plane is inclined at exactly the same slope as the cone, results in a *parabola*.

Observe that $x^2 + y^2 = z^2$ is the equation of a projective circle; this explains why all conic sections are equivalent under projective transformations.



We define a Dandelin sphere Ω to be a sphere tangent to both C (at a circle Γ) and Λ (at a point F, namely the *focus*). The plane containing Γ intersects Λ at a line δ, known as the *directrix*.

1. Prove that the directrix is the polar of the focus.

For an arbitrary point P on the conic, we let PR meet Γ at Q.

2. Prove that $PQ = PF$.

3. Let A be the foot of the perpendicular from P to the plane containing Γ. Let D be the foot of the perpendicular from P to δ. Show that $\frac{PQ}{PD}$ is independent of the location of P.

By combining the two previous theorems, we establish the *focus-directrix property* of a conic section.

- For every point P on a conic section, the ratio $\frac{PF}{P\delta} = \varepsilon$ remains constant. ε is known as the *eccentricity* of the conic. **[Focus-directrix property]**

The type of conic section can be determined by its eccentricity.

4. Show that $\varepsilon < 1$ for an ellipse, $\varepsilon = 1$ for a parabola and $\varepsilon > 1$ for a hyperbola.

Conics on a plane

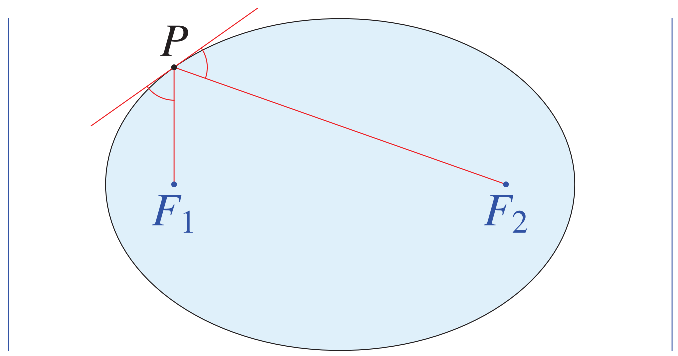
The focus-directrix property enables us to give conic sections a Cartesian treatment. By allowing the directrix to be the x-axis and scaling the conic so that the focus is at $(0, 1)$, the equation of a conic becomes $x^2 + (y-1)^2 = \varepsilon^2 y^2$. We can see that a conic section is a quadratic curve (although this is obvious from the projective definition). More remarkably, the converse is also true: all non-degenerate quadratic curves are conic sections.

5. Prove that, if $\varepsilon \neq 1$, the conic has two lines of reflectional symmetry.

Hence, for ellipses and hyperbolae, we can reflect the focus and directrix in the line of symmetry to obtain an alternative focus and directrix. Returning to the Dandelin spheres, the other focus corresponds to placing the sphere below the plane instead of above it. A parabola can be regarded as an ellipse/hyperbola with a focus on the line at infinity.

The Cartesian equation also makes it evident that ellipses are indeed ‘squashed (affine transformed) circles’. We can translate the ellipse so that the lines of symmetry are coordinate axes, giving us the equation for an ellipse.

- $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is the general equation for an ellipse. If we reverse the sign of $\frac{y^2}{b^2}$, we obtain a hyperbola instead. **[Cartesian equations for ellipses and hyperbolae]**

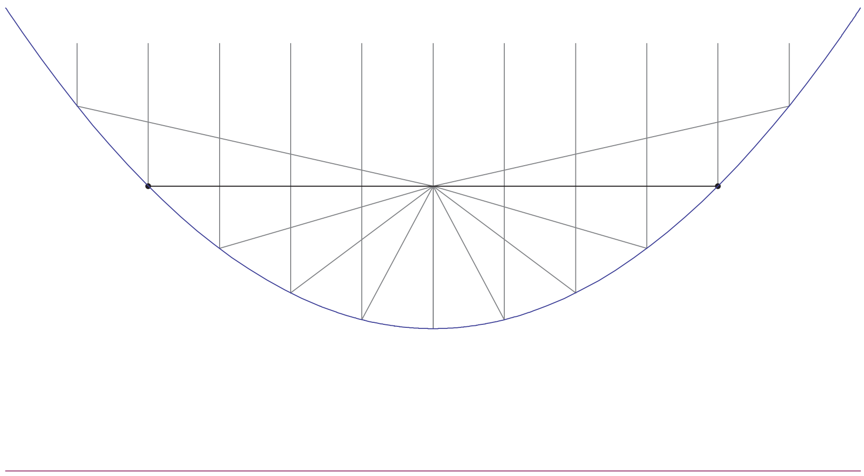


Consider an ellipse, with its two foci (F_1 and F_2) and directrices. Let P be a variable point on the ellipse.

6. Prove that $PF_1 + PF_2$ is constant.

7. Show that the angles between the tangent at P and the lines PF_1 and PF_2 are equal. **[Reflector property of the ellipse]**

A parabola can be considered to be the limit of ellipses with one focus fixed and the other tending towards infinity. This gives us the reflector property of the parabola, which states that a pencil of rays originating from the focus is reflected to a pencil of parallel lines perpendicular to the directrix. This was known to Archimedes, and formed the basis of a mechanism for igniting the sails of enemy ships by reflecting sunlight from polished metal shields. Nowadays, it is used in Newtonian telescopes for focusing light from infinity.



Let the focus be the origin, and the directrix be represented with $x = d$ in Cartesian coordinates. Consider the polar coordinates $\langle r, \theta\rangle = (r\cos\theta,\ r\sin\theta)$.

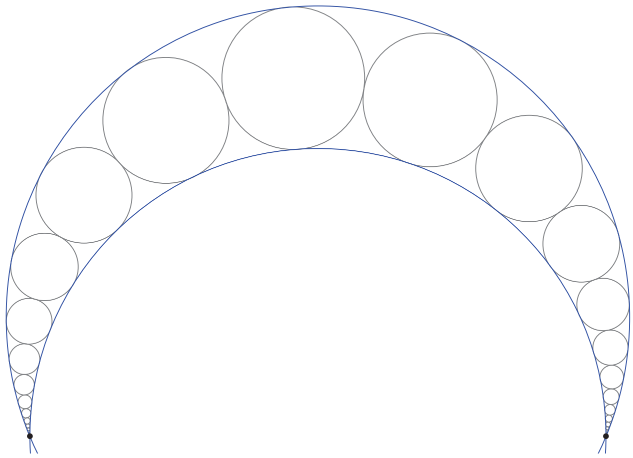
8. Prove that the equation for a conic in polar coordinates is $r = \frac{l}{1+\varepsilon\cos\theta}$, where $l = \varepsilon d$. **[Polar equation of a conic]**

This parametrisation of a conic will prove useful when verifying Kepler's laws of planetary motion in the next section.

The length $l = \varepsilon d$ is known as the *semi-latus rectum*, as it is half of the length of the line segment parallel to the directrix passing through the focus and meeting the conic twice. The latus rectum is shown in black on the diagram of the parabolic reflector.

The return of Steiner's porism

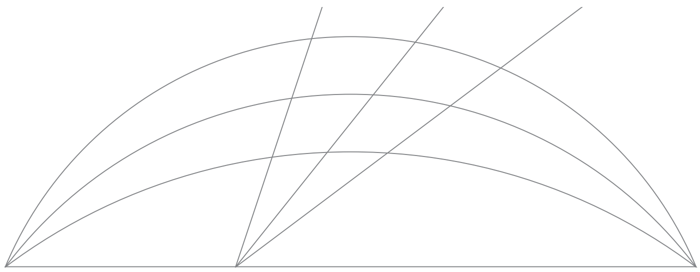
In Steiner's porism, we considered the family of circles internally tangent to Γ_2 and externally tangent to Γ_1. Let's suppose these two circles now intersect. Instead of a finite Steiner chain, we obtain an infinite set of circles bounded by Γ_1 and Γ_2.



9. Show that the centres of the circles lie on an ellipse, the foci of which are the centres of Γ_1 and Γ_2.

10. Hence demonstrate that the radius of the variable circle is proportional to the distance between its centre and the radical axis of Γ_1 and Γ_2.

11. Three circular arcs, γ_1, γ_2 and γ_3, connect the points A and C. These arcs lie in the same half-plane defined by the line AC in such a way that γ_2 lies between γ_1 and γ_3. Three rays, h_1, h_2 and h_3, emanate from a point B on the line AC, resulting in a grid of four curvilinear quadrilaterals as shown in the diagram below. Prove that if one can inscribe a circle in each of three of the curvilinear quadrilaterals, then a circle can be inscribed in the fourth. **[IMO 2010 shortlist, Question G7, Géza Kós]**



Géza created many more problems on this theme, all of which are amenable to embedding in three-dimensional space. One equivalent problem is where the situation is on the surface of a sphere, which can be transformed into the original problem through stereographic projection. Similarly, there are variants in hyperbolic space.

Kepler's laws of planetary motion

The vast majority of the content of this book is exclusively in the realms of pure mathematics. Nevertheless, conic sections naturally occur as the paths traced by objects in gravitational fields. The elliptical orbits of planets were first proposed by Johannes Kepler, as a refinement of earlier (mostly Greek) ideas that celestial bodies travel in perfect circles. Isaac Newton later inferred his law of gravitation from Kepler's laws; the derivation is not too difficult, although it relies heavily upon differential calculus.

Consider an object of negligible mass moving around a fixed object O of large mass due to gravitational attraction. An example of this is the Earth orbiting the Sun. We aim to show that the path must be a conic section, by showing that the polar equation of a conic satisfies Newton's laws of gravitation. Suppose an object is initially at P and moves to Q (very close to P). Let $A = [POQ]$ be the area 'swept out' by the object, and consider the derivative $\frac{dA}{dt}$, known as the *areal velocity*.

12. If an object moves in a straight line at constant velocity, show that $\frac{dA}{dt}$ is constant.

13. If the acceleration of an object is entirely radial (towards or away from O) at all times, then show that $\frac{dA}{dt}$ again remains constant. **[Conservation of angular momentum]**

Indeed, the converse is also true: if areal velocity is conserved, then acceleration is entirely radial. By integrating $\frac{dA}{dt}$ with respect to time, we obtain Kepler's second law.

- A planet P in orbit around the Sun O sweeps out equal areas in equal intervals of time. **[Kepler's second law]**

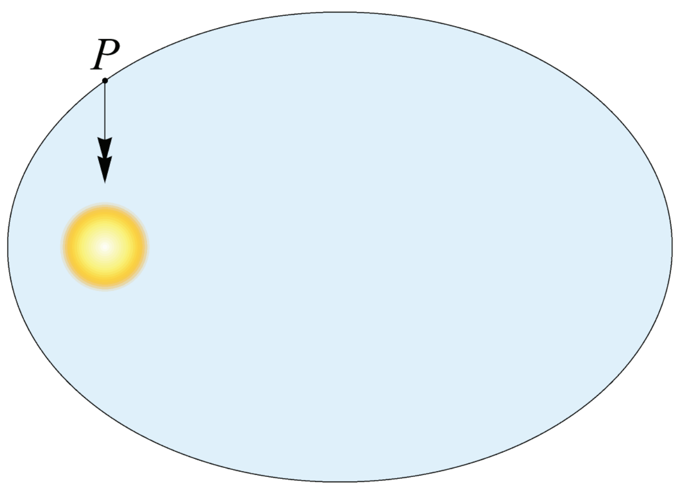
If $P = \langle r, \theta \rangle$, then $\frac{dA}{dt} = \frac{1}{2} r^2 \frac{d\theta}{dt}$. So, the value of $r^2 \frac{d\theta}{dt}$ must remain constant. Let's refer to this value (twice the areal velocity) as k.

In ordinary circular motion, the radial acceleration is given by $-r\left(\frac{d\theta}{dt}\right)^2$. Hence, in the general case, radial acceleration equals $a = \frac{d^2 r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2$.

14. Show that, if the planet P follows the path of a conic with focus O and semi-latus rectum l, then $a = -\frac{k^2}{r^2 l}$. **[Newton's inverse square law]**

Conversely, if we assume the inverse square law $a = -\frac{GM}{r^2}$, then we can choose a conic with centre O, passing through P in the appropriate direction, and with a latus rectum of $l = \frac{k^2}{GM}$. As Newton's law of universal gravitation is deterministic, the conic must be the **unique** solution. Hence, the converse is also true: all planets obeying the inverse square law travel in conic orbits. If the orbit is cyclic (and therefore closed), then it must be an ellipse.

- A planet P describes an ellipse, one focus of which is the Sun O. **[Kepler's first law]**



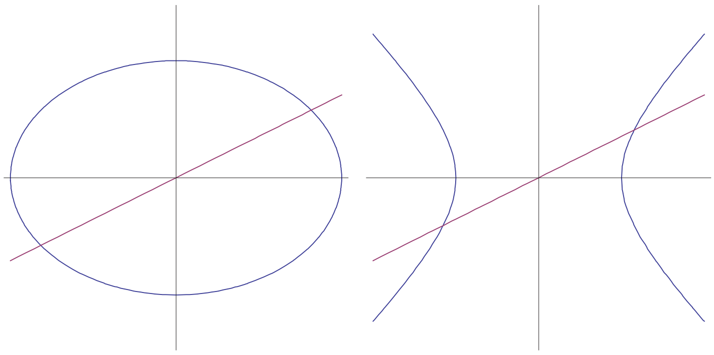
By considering the equation $r = \frac{l}{1+\varepsilon \cos\theta}$, the value of r is minimised at $\frac{l}{1+\varepsilon}$ (the *perihelion*) and maximised at $\frac{l}{1-\varepsilon}$

(the *aphelion*). When the eccentricity is zero, the orbit is circular.

In general, two bodies experiencing gravitational attraction will orbit each other in coplanar conic orbits, where the barycentre of the system (assumed to be stationary) is their common focus. For three or more bodies, the equations cannot be solved algebraically, and the system behaves chaotically (arbitrarily small initial perturbations lead to arbitrarily large effects). Indeed, it has been shown to be undecidable, so no computer or Turing machine is capable of calculating the movements with perfect precision.

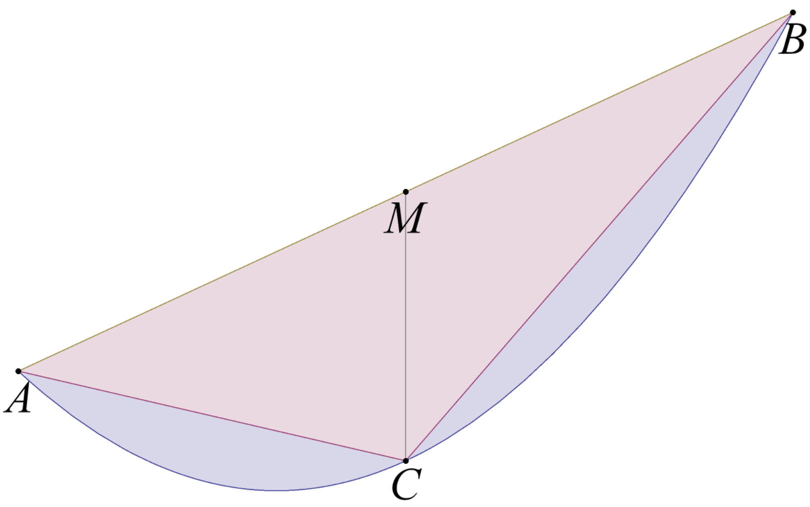
Areas of conics

If we take the pole of the line at infinity, we obtain the *centre* of a conic. For parabolae, this point is at infinity, therefore does not lie on the affine plane. For ellipses and hyperbolae, however, the centre lies on the plane and can be taken as the origin. It is then possible to apply a rotation about the origin to place the conic in *standard position*.



The ellipse has Cartesian equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. The line $y = x \tan\theta$ meets the curve at $P = (a\cos\theta,\ b\sin\theta)$. The area bounded by the curve, the line OP and the positive x-axis is given by $\frac{1}{2}ab\theta$. In particular, when $\theta = 2\pi$, the total area of the ellipse is $\pi a b$.

Similarly, the hyperbola has Cartesian equation $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$. The line $y = x\tanh\phi$ meets the curve at $P = (a\cosh\phi,\ b\sinh\phi)$. The area bounded by the curve, the line OP and the positive x-axis is given by $\frac{1}{2}ab\phi$. The hyperbolic functions are defined in a similar way to the trigonometric functions, with $\cosh\phi = \frac{1}{2}\left(e^{\phi} + e^{-\phi}\right)$ and $\sinh\phi = \frac{1}{2}\left(e^{\phi} - e^{-\phi}\right)$.



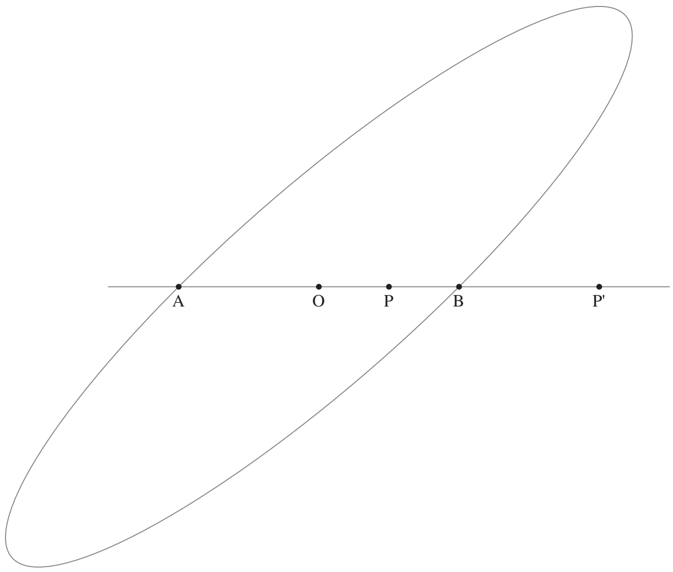
The area of a parabolic segment is much easier to calculate, as it can be obtained by integration of the equation of the parabola, $y = \frac{x^2}{4l}$, with respect to x. Archimedes instead used completely Euclidean methods in his *Quadrature* *of the Parabola*, determining the area recursively by adding together the area of triangle $A\,B\,C$ with the areas of the parabolic segments below $A\,C$ and $B\,C$. M is the midpoint of $A\,B$, and C is the intersection of the parabola with the perpendicular from M to the directrix. The triangle constructed by repeating this process with $A\,C$ instead of $A\,B$ has one eighth of the area of the original triangle. By summing an infinite geometric series, the total area is equal to $[A\,B\,C]\left(1 + \frac{2}{8} + \frac{4}{64} + \frac{8}{512} + \ldots\right) = \frac{4}{3}[A\,B\,C]$.

Unlike areas, the arc lengths of ellipses are not easy to compute. The circumference of an ellipse is $4\,a\,E(\varepsilon)$, where E is the *complete elliptic integral of the second kind*. With the exception of the circle, where $E(0) = \frac{\pi}{2}$ and the circumference is $2\,\pi\,r$, $E(\varepsilon)$ cannot be expressed in terms of basic functions.

Inversion in arbitrary conics

In the chapter about the Riemann sphere, we considered inversion in a circle. It is, however, possible to invert about any conic section. We have a seven-parameter set of inversions we can apply, as the centre of inversion and conic can be chosen independently.

- Let Γ be a non-degenerate conic, and O be a point not on Γ. For any point P other than O, we draw the line l through O and P, and let it meet Γ at A and B. We then define P' to be the projective harmonic conjugate of P with respect to $A\,B$. **[Inversion in a conic]**



Equivalently, we can define P' as the intersection of the polar of P with the line $O\,P$.

To investigate the properties of conic inversion, apply a projective transformation to make Γ a circle and O its centre. Then, the basic theorems applying to ordinary inversion translate into projective versions.

- O inverts to an entire line Ω, namely the polar of O, and *vice-versa*. (If O is the centre of the conic, then this is the line at infinity.) We treat this line as a single point, so the projective plane becomes topologically equivalent to a sphere. Allow Ω to intersect Γ at the points I and J.

In this perverse world of conic inversion, Ω behaves like the point-line at infinity and I and J are analogous to the circular points. This enables us to convert theorems in circle inversion to their conic counterparts.

- Straight lines passing through O remain invariant under inversion.

- Conics containing O, I and J invert to straight lines not passing through O, and *vice-versa*.

- Conics containing I and J (but not O) invert to other conics containing I and J (but not O).

In circle inversion, angles between curves remain constant (or, more precisely, are reversed). In conic inversion, this must be converted into a projective statement.

15. If P and Q invert to P' and Q', respectively, then show that $P\,Q\,P'\,Q'\,I\,J$ are conconic.

16. Let curves C and D intersect at P. Let C' and D' be the inverse curves with respect to Γ, and let P' be the inverse of P. The tangent to C at P and the tangent to D' at P' intersect at R. Similarly, the tangent to D at P and the tangent to C' at P' intersect at S. Show that $P\,P'\,R\,S\,I\,J$ are conconic. **[Preservation of generalised angle]**

Solutions

1. Apply a projective transformation to take δ to infinity, then apply an affine transformation to return Ω to being a sphere. The cone is tangent to Ω, so remains a right circular cone. The plane containing Γ becomes parallel to the plane Λ (i.e. horizontal). By symmetry, F is now the centre of the conic (which is a circle), and therefore the pole of the line at infinity δ.

2. They are both tangents from P to Ω, therefore of equal length.

3. The angle $\angle APQ$ in the right-angled triangle is constant (equal to half the angle at the vertex of the cone), so $\frac{PA}{PQ}$ is independent of the location of P. Similarly, $\frac{PD}{PA}$ is also constant, by considering the right-angled triangle PAD.

4. Observe that $\varepsilon = \frac{PQ}{PD} = \frac{\cos\angle HPD}{\cos\angle HPQ}$. When $\varepsilon = 1$, the plane Λ is inclined at the same slope as the cone, thus creating a parabola. When $\varepsilon < 1$, the plane is shallower than this, so the conic is an ellipse. Conversely, when the eccentricity exceeds 1, we have a hyperbola.

5. The equation of the conic is $x^2 + \left(1-\varepsilon^2\right) y^2 - 2y + 1 = 0$. We can complete the square, resulting in $x^2 + \left(1-\varepsilon^2\right)\left(y - \frac{1}{1-\varepsilon^2}\right)^2 - \frac{\varepsilon^2}{1-\varepsilon^2}$. This is symmetric about the lines $x = 0$ and $y = \frac{1}{1-\varepsilon^2}$.

6. Let the feet of the perpendiculars from P to the directrices be D_1 and D_2. We have $PF_1 + PF_2 = \varepsilon(PD_1 + PD_2) = \varepsilon(D_1D_2)$.

7. Let Q be another point on the ellipse, very close to P. Let $F_1P = u$, $F_1Q = u + \delta$ and $F_2Q = v$. By the previous theorem, $F_2P = v + \delta$. Let $PQ = \gamma$. We apply the cosine rule, to get $\cos\angle F_1PQ = \frac{u^2+\gamma^2-(u+\delta)^2}{2u\gamma} = \frac{\gamma^2-2u\delta-\delta^2}{2u\gamma}$. As Q approaches P, the γ^2 and δ^2 terms become negligible, and this cosine equates to $-\frac{\delta}{\gamma}$. This is the same as $\cos\angle F_2QP$, so the rays F_1P and F_2P describe equal angles with the normal to the curve. This establishes the reflector property.

8. From the focus-directrix property, the equation for the conic is $r = \varepsilon(d - x) = \varepsilon(d - r\cos\theta)$. Rearranging this equation gives us the formula for r.

9. Let the variable circle have centre P and radius r, and let Γ_i have radius R_i and centre O_i. We have $PO_1 = r + R_1$ and $PO_2 = R_2 - r$. Adding the equations gives $PO_1 + PO_2 = R_1 + R_2$, which is constant. Hence, the locus of centres is an ellipse with foci O_1 and O_2.

10. $r = PO_1 - R_1$ is a linear function of the distance to the focus, which is a linear function of the distance to the directrix, which is a linear function of the distance to the radical axis (which is parallel to the directrix). As the radius of the variable circle tends to zero as it approaches the radical axis, this linear function must have a constant term of 0. Hence, the radius of the circle is proportional to the distance between its centre and the radical axis.

11. Let the three circles be C_1, C_2 and C_3, such that (without loss of generality) C_2 and C_3 are tangent to both h_1 and h_2, and C_1 and C_2 are tangent to both γ_1 and γ_2. Let a fourth circle C_4 be tangent to γ_2 and γ_3, and let its centre lie on the line BC_1. Let d_i denote the perpendicular distance between the centre of C_i and the line AC, and let r_i denote its radius. We want to show that $\frac{r_4}{r_1} = \frac{d_4}{d_1}$, as this means that the circles are homothetic with centre of homothety B, so C_4 is tangent to both h_2 and h_3. Using this idea of homothety, we have $\frac{r_3}{r_2} = \frac{d_3}{d_2}$. Similarly, the previous exercise gives us $\frac{r_4}{r_3} = \frac{d_4}{d_3}$ and $\frac{r_2}{r_1} = \frac{d_2}{d_1}$. We can multiply these three equations to prove that $\frac{r_4}{r_1} = \frac{d_4}{d_1}$, so C_4 is indeed inscribed in the remaining curvilinear quadrilateral. For further discussion of the problem, see http://www.imo-official.org/problems/IMO2010SL.pdf. This includes a three-dimensional interpretation of the problem where cones of equal gradient are erected on the circles.

12. Let d be the perpendicular distance from the locus of motion to O. Then $[O\,P\,Q] = \frac{1}{2}\,v\,d\,t$, so $\frac{dA}{dt} = \frac{1}{2}\,v\,d$ is constant.

13. If P continues in a straight line at its present velocity, let the new position after time t be denoted Q. If P instead is accelerated towards or away from O, then its new position is denoted Q'. As the acceleration is in the direction of O, and t is very small (technically, the limit as $t \to 0$), $O\,P$ is parallel to $Q\,Q'$. So, $[O\,Q\,P] = [O\,Q'\,P]$ and thus $\frac{dA}{dt}$ is unaffected by the acceleration. So, it must remain constant, as in the previous scenario.

14. We begin with the polar form of a conic, $r = \frac{l}{1+\varepsilon\cos\theta}$. We then rearrange to obtain $\frac{l}{r} = 1 + \varepsilon\cos\theta$, and differentiate both sides with respect to time. This gives us $-\frac{l}{r^2}\,\frac{dr}{dt} = -\varepsilon\sin\theta\,\frac{d\theta}{dt}$. Multiplying both sides by $-\frac{r^2}{l}$ gives us $\frac{dr}{dt} = \frac{k\,\varepsilon}{l}\sin\theta$. Proceeding to differentiate again results in $\frac{d^2r}{dt^2} = \frac{k\,\varepsilon}{l}\cos\theta\,\frac{d\theta}{dt} = k\left(\frac{1}{r} - \frac{1}{l}\right)\frac{d\theta}{dt}$, where the last stage involves substituting the equation of the conic back into the equation. Acceleration is then $a = k\left(\frac{1}{r} - \frac{1}{l}\right)\frac{d\theta}{dt} - r\left(\frac{d\theta}{dt}\right)^2 = k\left(\frac{1}{r} - \frac{1}{l}\right)\frac{k}{r^2} - \frac{k^2}{r^3} = -\frac{k^2}{r^2\,l}$.

15. Project I and J to the circular points at infinity, so O is the centre of the circle Γ. $P\,Q\,P'\,Q'$ are concyclic, thus $P\,Q\,P'\,Q'\,I\,J$ are conconic.

16. Again, project I and J to the circular points at infinity, so O is the centre of the circle Γ. Then, this statement equates to $P\,P'\,R\,S$ being a cyclic quadrilateral, which is obvious from the fact that angles are preserved in circle inversion.