The Highlight Ovals

Adrien Bartoli

Abstract

We propose a model of plane ovals based on two foci defined in the 3D space. Geometrically, some of these ovals are the isocontours of a planar specular highlight for a point light source according to Phong's reflectance model. We hence call them the highlight ovals. More precisely, we define a geometric highlight oval as the locus of points p whose product of distances to the foci is proportional to the dot-product of the vectors from p to the foci. When the proportionality is taken up to sign, this gives a quartic, whose roots define the algebraic highlight oval. Similarly to the Cartesian oval, the algebraic formulation hence includes a pair of geometric highlight ovals. Similarly to the Cassinian oval, the geometric definition includes non-convex closed curves. The highlight ovals have four intrinsic parameters, namely the height of the two foci relative to the oval's plane, the distance between the two foci's projection on the oval's plane, and a scale factor fixing the proportionality and related to the isocontour's intensity. We thoroughly study the topology and convexity of the family of algebraic highlight ovals for any combination of its four intrinsic parameters and show how this characterizes the geometric highlight ovals, and thus the isocontours of specular highlights. We report an extensive experimental evaluation showing that the highlight ovals form a reliable model of specular highlights.

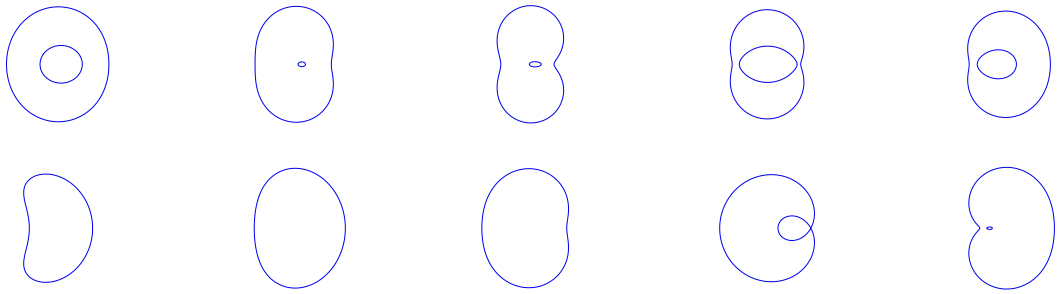


Figure 1: **Examples of algebraic highlight ovals.** Algebraic highlight ovals are algebraic curves which generalize the isocontours of planar specular highlights according to Phong's reflectance model. The algebraic highlight ovals which correspond to physically realizable specular highlights are said to be in Phong conditions and in non-Phong conditions otherwise. For instance, the observer and light source are always located on the same side of the observed surface in Phong conditions, but may be located on different sides in non-Phong conditions. The top row shows examples obtained in Phong conditions. These are made of two nested closed curves. The inner curve is convex and corresponds to the highlight's isocontours. The bottom row shows examples obtained in non-Phong conditions.

1 Introduction

Loosely speaking, an oval is a convex closed plane curve whose shape looks similar to that of an egg (Dixon, 1987). The most well-known models of ovals are probably the Cassinian and the Cartesian ovals (Lawrence, 1972), which both involve two foci. Geometrically, a Cassinian oval is the locus of points whose product of distances to the foci equals a constant and a Cartesian oval is the locus of points whose weighted sum of distances to the foci equals a constant. Both geometric definitions can be formulated algebraically using quartics. There also exist cubic and higher-degree algebraic models of ovals, such as the elliptic curves (Milne,

2006). We now make two clarifications. The first one is about the discrepancy that may occur between the algebraic and the geometric loci, as in general the algebraic locus forms a superset of the geometric locus. For instance, the algebraic locus of the quartic for the Cartesian oval includes two geometric loci. This occurs because the algebraic formulation requires one to eliminate the square roots involved in the distances by squaring, which also eliminates the weight's and constant's signs. We raise the ambiguity that this discrepancy introduces by qualifying the terms as geometric or algebraic as and when needed. For instance, we use the expressions *geometric oval* and *algebraic oval*. The second clarification we make is about the use of the term oval, as most types of oval include non-convex curves. This happens for some geometric ovals, and thus necessarily for the corresponding algebraic ovals. For instance, a geometric Cassinian oval may look like a pair of circles or the infinite sign ∞. Therefore, following the literature, we use the term oval in a loose sense, to name the curve obtained from the geometric or algebraic curve families.

We propose a model of ovals which we call the highlight ovals, illustrated in figure 1. We chose this name because our model includes the isocontours of specular highlights created by a point light source on a glossy planar surface, as we show in the next section using Phong's reflectance model. We study the algebraic highlight ovals and classify them in two categories. The first category is formed of the algebraic highlight ovals which have a part that strictly matches a highlight isocontour, as illustrated in the top row of figure 1. The second category contains all the other algebraic highlight ovals, as illustrated in the bottom row of figure 1. We say that the algebraic highlight ovals from the first and second categories are respectively in Phong conditions and non-Phong conditions. The highlight ovals are based on two space-foci and are defined in the (x, y) plane $\mathcal{S}$ lying at $z = 0$. The Phong conditions specifically mean that the two space-foci are located on different sides of, and not on, plane $\mathcal{S}$. A point lies on an algebraic highlight oval if the product of its distances to the foci equals the dot product of the vectors between the point and the foci, up to a defined scale and sign. Similarly to the algebraic Cassinian and Cartesian ovals, the algebraic highlight ovals depend on two foci and are defined using quartics. The key difference is that the foci are defined in the 3D space rather than in the 2D plane. The Cassinian and Cartesian ovals have respectively two and three intrinsic parameters, while the highlight ovals have four.

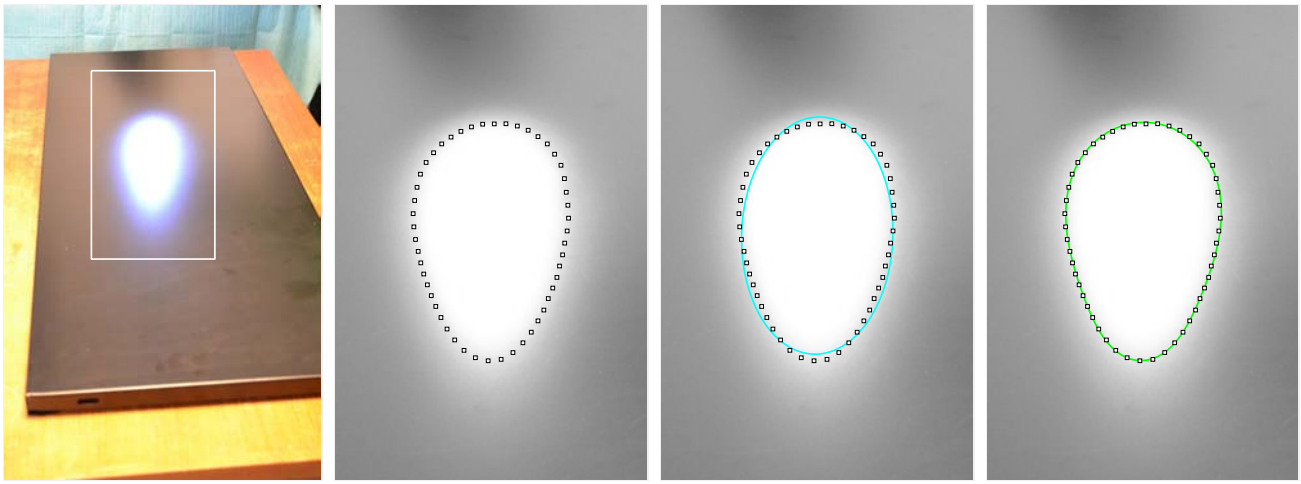


Figure 2: **Example of model fitting to a real image with saturation.** (from left to right) The image was taken by a digital camera in regular settings, causing the specular highlight to sature the sensor. The box around the specular highlight has size 404×616 pixels. The sample points were extracted from the image. The ellipse and algebraic highlight oval obtained fitting residuals of 4.20 px and 0.93 px respectively.

Our primary motivation to introduce and study the highlight ovals stems from their tight connection with the specular highlights. Specular highlights play a very important role in the perception of shape (Blake and Bülthoff, 1990). They typically form a bright blob, for instance when created by a point light source on a glossy planar surface, as illustrated by the example of figure 2. This bright blob is due to the camera sensor being saturated by the strong light reflected at the specular highlight, and forms the most common instance

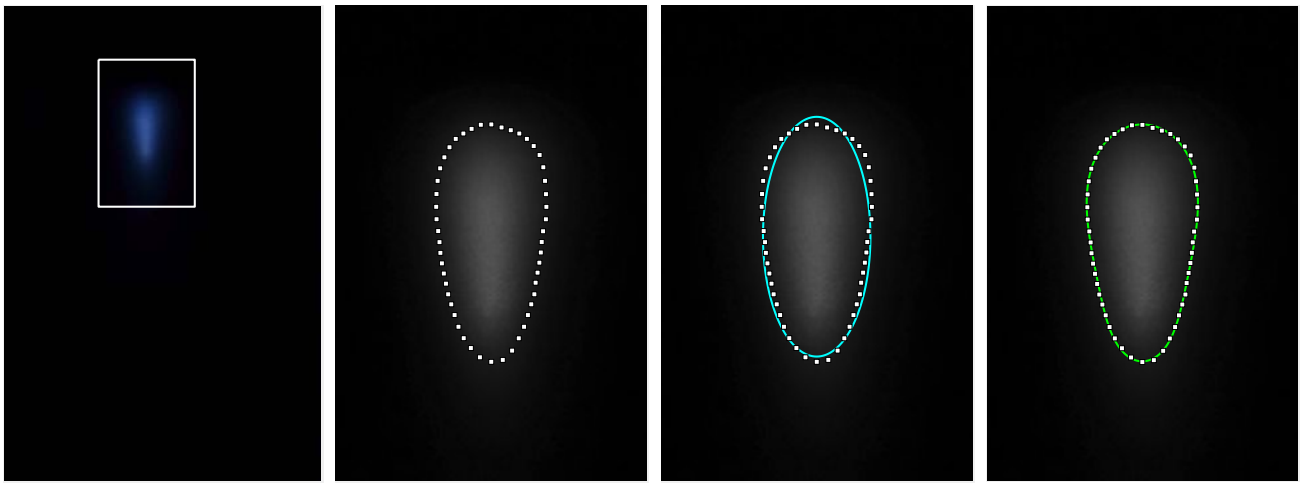


Figure 3: **Example of model fitting to a real image without saturation.** (from left to right) The image was taken by a digital camera in fast-shutter mode, allowing the sensor to image the specular highlight. The box around the specular highlight has size 404×616 pixels. The sample points were extracted from the image. The ellipse and algebraic highlight oval obtained fitting residuals of 3.26 px and 0.72 px respectively.

of specular highlights in natural photographs. The transition between the saturated and non-saturated parts of the image represents the shape of the specular highlight, simply called the *specular shape*. Changing the camera settings may allow one to avoid saturation. This may be achieved by setting a mode amongst fast-shutter, high dynamic range and low ISO sensitivity. In the absence of saturation, the specular shape is not observable but the image rather reveals a smooth transition from a brightest point located approximately at the centre of the specular highlight toward its boundary, as illustrated by the example of figure 3. In this case, the so-called *specular isocontours* form interesting closed curves to study. Specular highlights are more rarely observed this way in natural photographs. Interestingly, the specular shape and isocontours both correspond to a single model, as the specular shape is simply a specular isocontour. This can be understood by considering the camera response function as a simple truncated linear function. The truncation value represents saturation and the specular shape is thus simply the specular isocontour at the truncation value. Modeling the specular shape and isocontours requires one to model the formation of a specular highlight. The way light is reflected by a surface is described mathematically by a reflectance function, which predicts the outgoing light as a function of the incoming light and viewpoint (Hughes et al., 2014). Phong proposed a seminal reflectance model (Phong, 1975), which has been extensively used in image synthesis. Phong's model is phenomenological but models the shape of real specular highlight convincingly (Breuss and Ju, 2011). It has been replaced by other models such as Blinn-Phong's in modern rendering engines, mainly for reasons related to computation speed. Phong's model has three terms: ambient, diffuse and specular. The latter of which specifically models specular highlights and is the only one that activates for a purely specular material. Concretely, this term predicts the outgoing light at a surface point $p \in \mathbb{R}^2$ using the cosine of the angle between the vector joining point p to the viewpoint and the direction of perfect reflection of the incoming light at point p. A commonly admitted model of the specular shape and isocontours is the ellipse, for which a formal derivation may be found in (Forsyth and Ponce, 2003). Though this model works in some cases, it lacks accuracy as many specular highlights are ovals, a shape which cannot be generated by the elliptical model. In contrast, the highlight oval provides a much more faithful model of the specular shape and isocontours, as illustrated by the examples of figures 2 and 3.

We start by formalizing the algebraic highlight ovals and expliciting their relationship with specular highlights in §2. We then give a complete classification of the algebraic highlight ovals and study their topology, represented by the number of connected components, and their convexity, stating whether the connected components are convex or not, in §3. This forms the main theorem, whose proof is given in §§3 and 5. The latter part uses the cross-section space which we introduce in §4, where the properties of the

algebraic highlight oval quartic, which would be difficult to analyze directly, are revealed by the intersection of quadrics, a much easier to comprehend problem. We show in §6 how the main theorem specifically applies to the highlight ovals in Phong conditions. These results are used in §7 to create a fitting algorithm and evaluate the model on real images, by fitting specular shapes and isocontours acquired in a wide range of imaging conditions. We finally conclude in §8 and give some of the lemmas required in the proof of the main theorem in Appendix A.

2 Formalization

We first give a formal definition of the geometric and algebraic highlight ovals, show that the latter forms a quartic and give their so-called cosine form. We then establish the formal relationship between the highlight ovals and specular highlights following Phong's model. We finally give the so-called standard-form of the algebraic highlight ovals. The cosine-form and standard-form are both important to study the sought highlight oval properties.

2.1 The Highlight Ovals

We now provide a formal definition of the highlight ovals, using the distance function $d_F(P) : \mathbb{R}^2 \to \mathbb{R}_+$ with $d_F(P) = \|F - P\|_2$. We define $p^\top = [x\, y]$ and $P^\top = [p^\top\, 0] = [x\, y\, 0]$.

Definition 1 (Geometric highlight oval)**.** *A geometric highlight oval is defined by two space-foci $F, G \in \mathbb{R}^3$ and a scale $t \in \mathbb{R}$ as:*

$$\mathcal{H}(F, G, t) \stackrel{\text{def}}{=} \left\{ p \in \mathbb{R}^2 \mid t\, d_F(P)\, d_G(P) = \langle F - P, G - P \rangle \right\}. \tag{1}$$

Definition 2 (Algebraic highlight oval)**.** *The algebraic highlight oval is defined as:*

$$\bar{\mathcal{H}}(F, G, t) \stackrel{\text{def}}{=} \mathcal{H}(F, G, t) \cup \mathcal{H}(F, G, -t) = \left\{ p \in \mathbb{R}^2 \mid t\, d_F(P)\, d_G(P) = \pm \langle F - P, G - P \rangle \right\}. \tag{2}$$

Proposition 1 (Quarticity of the algebraic highlight ovals)**.** *The set $\bar{\mathcal{H}}(F, G, t)$ is defined by the roots of a bivariate quartic, obtained by squaring each side of the equality in equation (2):*

$$p \in \bar{\mathcal{H}}(F, G, t) \;\Leftrightarrow\; \Big((F - P)^\top (G - P) \Big)^2 - t^2 \|F - P\|_2^2 \|G - P\|_2^2 = 0. \tag{3}$$

The proof of proposition 1 is straightforward but understanding the topology and convexity of the locus of roots of a quartic curve may be extremely difficult (Basset, 1901). For the algebraic highlight oval we are more specifically interested in the number of connected components of the quartic's real roots, and their convexity. A number of special cases can be understood directly from the quartic (3), namely the cases with $|t| = 1$ or $t = 0$, or from the cosine form (5) given directly below, namely the cases with $|t| > 1$. The cases with $|t| < 1$ are however much harder and their study forms one of our main contributions. We give a theorem which is a complete classification of the algebraic highlight oval's topology and convexity for all possible combinations of parameters in the next section. Our proof is based on the fact that the algebraic highlight oval is the orthogonal projection of the intersection between a canonical paraboloid and a generalized cylinder, forming what we call the pseudo-3D form. This form may be analyzed in the plane perpendicular to the cylinder's axis and containing the origin. This is the cross-section space, where the problem boils down to analyzing the intersection points between two quadrics. Roughly speaking, we found that for $|t| > 1$ the algebraic highlight oval is almost always empty, that for $|t| = 1$ it is almost always a single singular point, that for $t = 0$ it is a circle in the general case with a sufficient condition that F and G be on different sides of the plane, and that for $0 < |t| < 1$ it takes a variety of shapes.

Finally, we observe that the geometric and algebraic highlight ovals can also be defined by equating t to the cosine of an angle subtending the two foci. This is the cosine form, given in the following proposition.

Proposition 2 (Cosine form of the highlight ovals)**.** *The cosine form of the geometric highlight oval is:*

$$p \in \mathcal{H}(F,G,t) \;\Leftrightarrow\; (P=F) \vee (P=G) \vee \left(\cos\left(\widehat{FPG}\right) = t\right). \tag{4}$$

The cosine form of the algebraic highlight oval is:

$$p \in \bar{\mathcal{H}}(F,G,t) \;\Leftrightarrow\; (P=F) \vee (P=G) \vee \left(\cos\left(\widehat{FPG}\right) = \pm t\right). \tag{5}$$

Proof. We first establish the cosine form of the geometric highlight oval. Substituting $P = F$ or $P = G$ in the definition (1) shows that both foci are part of the oval. For $P \neq F, G$, we simply rewrite the equality in the definition (1) as:

$$t \;=\; \frac{\langle F-P, G-P\rangle}{d_F(P)\, d_G(P)} \;=\; \cos\left(\widehat{FPG}\right).$$

Regarding the cosine form of the algebraic highlight oval, we use the definition (2) and the cosine form of the geometric highlight oval (4). □

2.2 Relationship to Specular Highlights

We show how the isocontours of Phong's specular term are related to the highlight ovals. We denote the light source as $L \in \mathbb{R}^3$ and the viewpoint as $D \in \mathbb{R}^3$. We assume they are both located on the positive side of plane $\mathcal{S}$, and so $L_z > 0$ and $D_z > 0$. This assumption forms what we have called Phong conditions. The perfect reflexion vector passes through a fixed point $R \in \mathbb{R}^3$ obtained by reflecting the light source about the plane $\mathcal{S}$, and so located on its negative side, with $R_x = L_x$, $R_y = L_y$ and $R_z = -L_z$. In essence, the specular term in Phong's model is defined as:

$$r(p) \;\stackrel{\text{def}}{=}\; j\left(\cos\left(\pi - \widehat{DPR}\right)\right)^{\alpha},$$

where $j \in \mathbb{R}_+$ is a fixed factor and $\alpha \in \mathbb{N}_+$ a fixed exponent. The isocontours of this model are found by setting $r(p) = \tau$, for some $\tau \in \mathbb{R}_+$, which we rewrite as:

$$\cos\left(\pi - \widehat{DPR}\right) \;=\; \sqrt[\alpha]{\frac{\tau}{j}}.$$

Expanding the cosine, we arrive at:

$$-\frac{\langle D-P, R-P\rangle}{\|D-P\|_2 \|R-P\|_2} \;=\; \sqrt[\alpha]{\frac{\tau}{j}},$$

which we identify with a geometric highlight oval by setting $F = D$, $G = R$ and $t = -\sqrt[\alpha]{\frac{\tau}{j}}$. We thus have the following correspondence between an isocontour and a geometric highlight oval:

$$r(p) = \tau \;\Leftrightarrow\; p \in \mathcal{H}\left(D, R, -\sqrt[\alpha]{\frac{\tau}{j}}\right).$$

We will prove the following proposition in §6.

Proposition 3 (Topology and convexity of the geometric highlight ovals in Phong conditions)**.** *The set* $\mathcal{H}\left(D, R, -\sqrt[\alpha]{\frac{\tau}{j}}\right)$ *is empty for* $\tau > j$*, a single point for* $\tau = j$*, a convex closed curve for* $j > \tau > 0$ *and a circle for* $\tau = 0$*. Defining* $w_+ \stackrel{\text{def}}{=} \frac{1}{2}[D_x + R_x\, D_y + R_y]^\top$ *and* $w_- \stackrel{\text{def}}{=} \frac{1}{2}[D_x - R_x\, D_y - R_y]$*, we have that the point is at* $w_+ + \frac{R_z + D_z}{R_z - D_z} w_-$*, that the circle is centred at* w_+ *and has radius* $\sqrt{\|w_-\|^2 - R_z D_z}$*. The convex closed curve always contains the point and is always contained by the circle.*

2.3 The Standard Form

We define the standard form of the highlight ovals by translating and rotating the coordinate frame in plane $\mathcal{S}$ so that, for $a \geq 0$, we have $F_x = a$, $G_x = -a$ and $F_y = G_y = 0$. We also define $f = F_z$ and $g = G_z$. We make the slight language abuse of using quartic to mean quartic curve. The standard form eliminates the three extrinsic parameters (formed by a plane rotation and translation) and reveals the four intrinsic parameters (f, g, a and t). We use 'iff' to mean 'if and only if'.

Proposition 4 (Standard form of the algebraic highlight oval)**.** *In standard form, the algebraic highlight ovals are defined by the real roots of the following bivariate quartic:*

$$\begin{aligned} h(x,y) \stackrel{\text{def}}{=} & \left(1-t^2\right)\left(x^2+y^2\right)^2 \\ & + \left(t^2\left(2a^2-f^2-g^2\right)+2\left(fg-a^2\right)\right)x^2 \\ & + \left(t^2\left(-2a^2-f^2-g^2\right)+2\left(fg-a^2\right)\right)y^2 \\ & + 2at^2\left(g^2-f^2\right)x + \left(1-t^2\right)\left(f^2g^2+a^4\right) - a^2\left(t^2\left(f^2+g^2\right)+2fg\right). \end{aligned} \tag{6}$$

Proposition 4 is proved trivially by substitution of the standard values in equation (3). It directly reveals that h is a bicircular quartic (Basset, 1901). It is also tempting to rescale the ambient space by a factor $\frac{1}{a}$ for $a \neq 0$ and then consider two cases, $a = 1$ and $a = 0$. We avoided this rescaling for three reasons. First, most of our results hold for both cases and keeping a does not clutter the equations. Second, we simply consider the two cases $a \neq 0$ and $a = 0$ as and when needed. Third, the experimental evaluation of the algebraic highlight ovals described in §7 involve images containing specular highlights for which a is unknown. Keeping a in the derivations thus allows us to use our results in real images directly.

Proposition 5 (Symmetries)**.** *The algebraic highlight oval in standard form is symmetric about the x axis. It is also symmetric about the y axis if $a = 0$, $t = 0$ or $|f| = |g|$.*

Proof. Regarding the x axis symmetry, we simply notice that $h(x,y) = h(x,-y)$. Regarding the y axis symmetry, we have that $h(x,y) - h(-x,y) = 4at^2\left(g^2-f^2\right)$, which vanishes iff $a = 0$, $t = 0$ or $|f| = |g|$. □

3 Topology and Convexity

We give our central result on the topology and convexity of the highlight ovals as theorem 1. The theorem is organized using the value of t, following an ansatz which we give directly below. The proof for three out of the four main cases is simple and given directly after the theorem. The proof for the last case is more complex and given in §5 following the cross-section space introduced in §4.

Ansatz 1 (Topology and convexity classification by $|t|$)**.** *The algebraic highlight oval is invariant to the sign of t, as it only involves t^2. We hypothesize that $|t|$ is a key to organize the results on its topology and convexity, especially using the following sets: $|t| > 1$, $|t| = 1$, $1 > |t| > 0$ and $t = 0$.*

Theorem 1 (Topology and convexity of the highlight ovals)**.** *The algebraic highlight oval, represented by the real roots of the quartic $h(x,y) = 0$, has the following number of connected component and convexity:*

1. *$|t| > 1$. The algebraic highlight oval is empty in the general case and made of one or two singular points in some cases. More specifically, the algebraic highlight oval is a conditional union of point $[a\,0]^\top$ if $f = 0$ with point $-[a\,0]^\top$ if $g = 0$.*
2. *$|t| = 1$. The algebraic highlight oval is a single singular point in the general case and the empty set, the whole plane or the x axis in some cases. More specifically, it is:*
   - *(a) Point $\left[a\frac{g+f}{g-f}\,0\right]^\top$ for $f \neq g$*
   - *(b) $\mathbb{R}^2$, for $f = g$ and $a = 0$*
   - *(c) $\emptyset$, for $f = g \neq 0$ and $a \neq 0$*

(d) The x axis, for $f = g = 0$ and $a \neq 0$

3. *$1 > |t| > 0$. The algebraic highlight oval may be empty, formed of a single singular point, a single closed curve with 0, 2 or 4 inflection points with possibly a self-intersection point or an isolated singular point, or be a convex closed curve nested in a closed curve with 0, 2 or 4 inflection points.*

4. *$t = 0$. The algebraic highlight oval may be empty, formed of the origin as a singular point, or be a circle centred at the origin. More specifically, it is:*

   (a) $\emptyset$, for $a^2 - fg < 0$

   (b) Point $[0\,0]^\top$, for $a^2 - fg = 0$

   (c) The circle of radius $\sqrt{a^2 - fg}$ with centre $[0\,0]^\top$, for $a^2 - fg > 0$

These are the only possibilities, including in part 3.

Our proof of theorem 1 is divided in four parts. Cases $|t| = 1$ and $t = 0$ are easily proved by examining the standard form directly. Case $|t| > 1$ is straighforward from the cosine form. They form the first three parts of our proof and are given directly below. The fourth part deals with case $1 > |t| > 0$. It requires the cross-section space, the height space and a number of lemmas. We use standard results in algebraic geometry of the conics (Semple and Kneebone, 1963). For a matrix $E \in \mathbb{R}^{3\times3}$, we use the notation $H_E(x, y) = [x\,y\,1]E[x\,y\,1]^\top$ and denote as $\bar{E} \in \mathbb{R}^{2\times2}$ the leading submatrix of E, formed by its first two rows and columns.

Proof of theorem 1, $|t| = 1$. We substitute $t = 1$ in the algebraic highlight oval in standard form (6). The quartic degenerates into a conic, defined by the following equation:

$$H_Q(x, y) \;=\; 0 \qquad \text{with} \qquad Q \;\stackrel{\text{def}}{=}\; \begin{bmatrix} -(f-g)^2 & 0 & a\left(g^2 - f^2\right) \\ 0 & -4a^2 - (f-g)^2 & 0 \\ a\left(g^2 - f^2\right) & 0 & -a^2(f+g)^2 \end{bmatrix}.$$

This forms a degenerate conic because $\det(Q) = 0$. The leading minor is $\det\left(\bar{Q}\right) = (f-g)^2\left(4a^2 + (f-g)^2\right)$. It is always non-negative. More specifically, it is positive for $f \neq g$ and vanishes otherwise. For $f \neq g$, the conic H_Q thus represents a single singular point, whose location at $\left[a\frac{g+f}{g-f}\,0\right]^\top$ can be easily verified from the conic's equation. For $f = g$, denoting proportionality up to a non-zero scalar as $\propto$, the conic matrix becomes $Q \propto \mathrm{diag}(0, a^2, a^2g^2)$, and so the quartic degenerates further into $a^2y^2 + a^2g^2 = 0$. For $a = 0$, it thus degenerates to the whole of $\mathbb{R}^2$. For $a \neq 0$ and $g \neq 0$, it is rewritten $y^2 = -g^2$ and is thus empty, while for $g = 0$ it is rewritten $y^2 = 0$ and is thus the x axis. □

Proof of theorem 1, $t = 0$. We substitute $t = 0$ in the algebraic highlight oval in standard form (6) which is then simplified in:

$$\left(x^2 + y^2 + fg - a^2\right)^2 \;=\; 0,$$

and equivalently in:

$$x^2 + y^2 \;=\; a^2 - fg,$$

which represents a circle centred at the origin for $a^2 - fg > 0$, the origin if $a^2 - fg = 0$ and the empty set otherwise. □

Proof of theorem 1, $|t| > 1$. We use the cosine form (5) of the algebraic highlight oval, defined as the union of three sets. The first two sets are $P = F$ and $P = G$. They only occur when F or G, respectively, lie on the oval support plane $\mathcal{S}$. In standard form, this means when $f = 0$, in which case the point $p = [a\,0]^\top$ is on the oval, or when $g = 0$, in which case $p = [-a\,0]^\top$ is on the oval. The third set defined with the cosine is empty because $|t| > 1$. □

4 The Cross-Section Space

We first construct the cross-section space from the pseudo-3D form, which itself rests on the standard form. We then give general results on the mapping of curves from the cross-section space to curves in plane $\mathcal{S}$. It turns out that the key properties to characterize this mapping are the intersection points of the curve in the cross-section space with the canonical parabola and the convexity and concavity of the graph of the curve in the cross-section space. We finally study properties of the curves representing the algebraic highlight ovals in the cross-section space, with a particular emphasis on the aforementioned key properties.

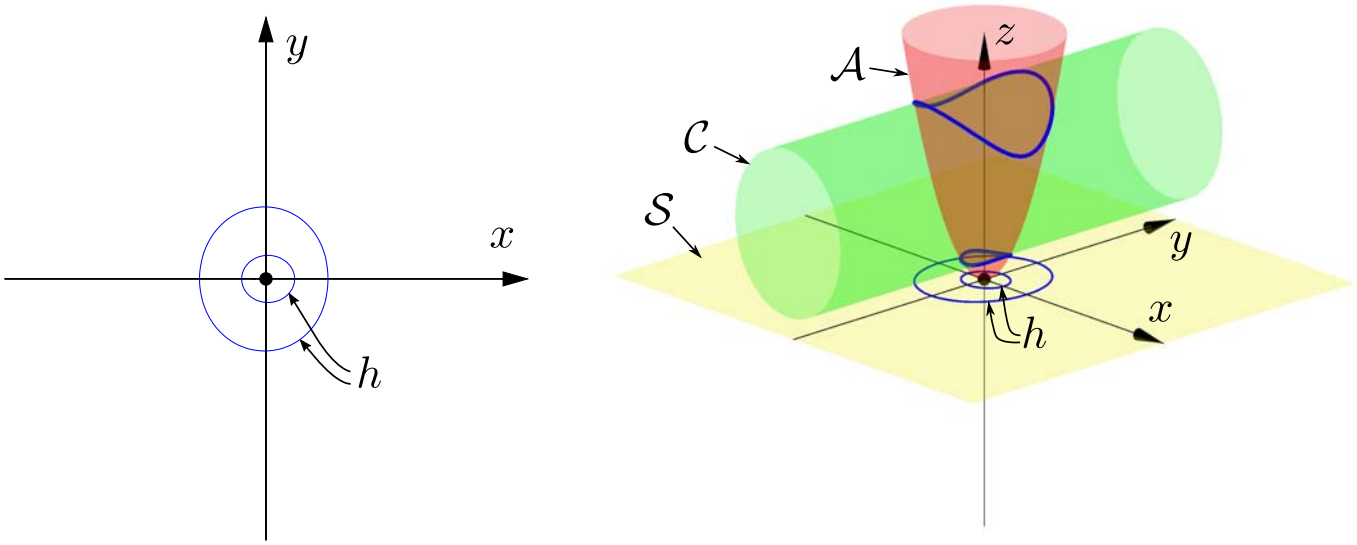


Figure 4: **The pseudo-3D form.** The numerical example is the algebraic highlight oval shown left (also top left in figure 1) in the $\mathcal{S}$ plane. In this example the cylinder's cross-section H_E is an ellipse and it intersects the canonical paraboloid $\mathcal{A}$ in four points, as shown in figure 5.

4.1 Construction using the Pseudo-3D Form

We first define the pseudo-3D form and then construct the cross-section space. The former is illustrated by figure 4 and formalized by the following proposition.

Proposition 6 (Pseudo-3D form)**.** *The algebraic highlight oval in standard form (6) is the orthographic projection on plane $\mathcal{S}$ of the intersection between the canonical paraboloid $\mathcal{A}$ and a cylinder $\mathcal{C}$, whose axis is parallel to the y axis. The cylinder's cross-section is a conic whose axes are parallel to the x and z axes.*

Proof. The canonical paraboloid is:

$$\mathcal{A} \stackrel{\text{def}}{=} \left\{P \in \mathbb{R}^3 \mid x^2 + y^2 - z = 0\right\}. \tag{7}$$

Substituting $y^2 = z - x^2$ in the algebraic highlight oval in standard form (6), we find the cylinder $\mathcal{C}$ defined as:

$$\mathcal{C} \stackrel{\text{def}}{=} \left\{P \in \mathbb{R}^3 \mid H_E(x, z) = 0\right\},$$

with:

$$E \stackrel{\text{def}}{=} \begin{bmatrix} 4a^2t^2 & 0 & at^2\left(g^2 - f^2\right) \\ 0 & 1 - t^2 & fg - a^2\left(1 + t^2\right) - \frac{t^2}{2}\left(f^2 + g^2\right) \\ at^2\left(g^2 - f^2\right) & fg - a^2\left(1 + t^2\right) - \frac{t^2}{2}\left(f^2 + g^2\right) & \left(1 - t^2\right)\left(f^2g^2 + a^4\right) - a^2\left(t^2\left(f^2 + g^2\right) + 2fg\right) \end{bmatrix}.$$

We have that H_E represents a conic in the (x, z) plane, whose axes align with the coordinate axes since $E_{1,2} = E_{2,1} = 0$. □

The pseudo-3D form is related to the algebraic highlight oval h via the cylinder's cross-section H_E. This relationship is inferred by simple substitution as:

$$h(x, y) \;=\; H_E(x, x^2 + y^2). \tag{8}$$

We define the cross-section space as the (x, z) plane. We chose this name because the intersection of the (x, z) plane with the cylinder $\mathcal{C}$ gives its cross-section conic, which forms a key representation to establish our proof of theorem 1. The cross-section space is illustrated by figure 5. Our motivation to use this space is that the intersection points of $\mathcal{A}$ and $\mathcal{C}$ in this space entirely reveal the topology and convexity of the algebraic highlight oval's quartic $h(x, y) = 0$ while being much easier to comprehend than $\mathcal{A}$ and $\mathcal{C}$ in the pseudo-3D form directly. Intuitively, this is because the cross-section space intersects $\mathcal{A}$ where it is the 'largest' and also captures the cross-section conic H_E of $\mathcal{C}$. The simplest example is when there are no intersections between $\mathcal{A}$ and $\mathcal{C}$ in the cross-section space, which directly implies that there are none in the pseudo-3D form either, and so that the algebraic highlight oval is empty. More specifically, in the cross-section space $\mathcal{A}$ becomes the canonical parabola $z = x^2$ and $\mathcal{C}$ becomes $H_E(x, z) = 0$. The canonical parabola is also written as:

$$H_A(x, z) \;=\; 0 \qquad \text{with} \qquad A \;\overset{\text{def}}{=}\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 0 \end{bmatrix}.$$

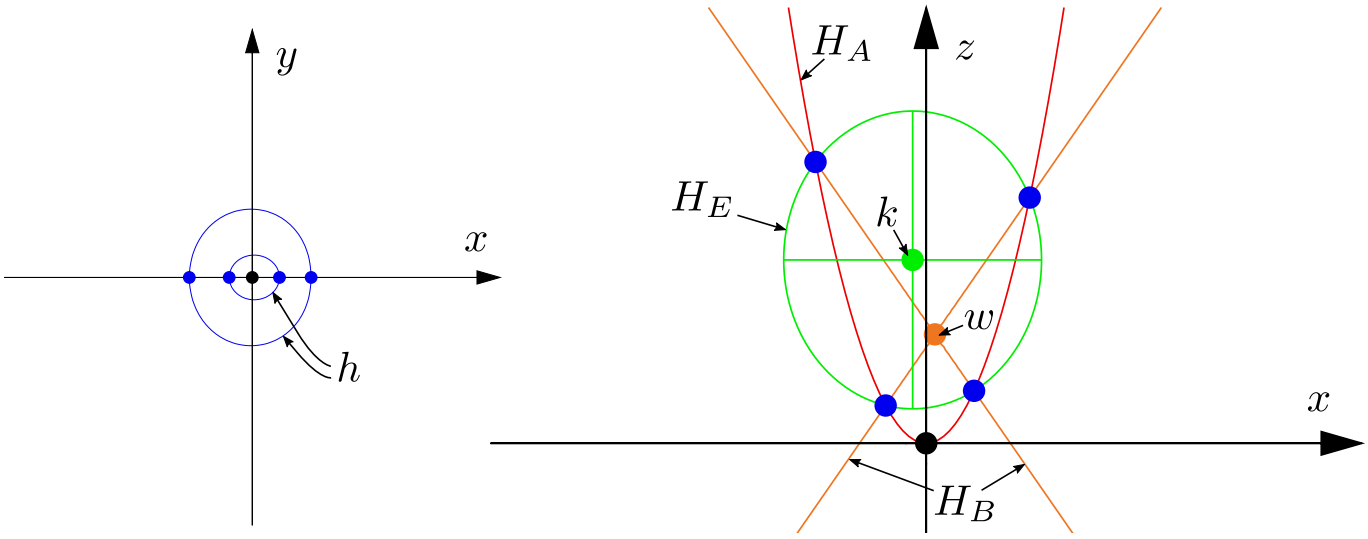


Figure 5: **The cross-section space.** The numerical example is the algebraic highlight oval shown left (also top left in figure 1 and in figure 4) in the $\mathcal{S}$ plane. In this example, we have $1 > |t| > 0$ and proposition 7 tells us that the cylinder's cross-section H_E is an ellipse. This ellipse has four intersections with the canonical parabola H_A, which are shown in blue in the cross-section space and the $\mathcal{S}$ plane.

4.2 General Results on Mapping Curves from the Cross-Section Space to Plane $\mathcal{S}$

We derive general results to characterize how a given curve in the cross-section space is mapped to plane $\mathcal{S}$. We start with equation (8). This yields the following partial derivatives:

$$\frac{\partial h}{\partial x}(x, y) \;=\; \frac{\partial H_E}{\partial x}(x, x^2 + y^2) + 2x\frac{\partial H_E}{\partial z}(x, x^2 + y^2) \qquad \text{and} \qquad \frac{\partial h}{\partial y}(x, y) \;=\; 2y\frac{\partial H_E}{\partial z}(x, x^2 + y^2). \tag{9}$$

We first give two lemmas relating properties of the cross-section H_E to properties of the algebraic highlight oval h.

Lemma 1 (Verticality of the algebraic highlight oval's tangent)**.** *The algebraic highlight oval's tangent is vertical iff the cylinder cross-section's tangent is vertical or the point is on the x axis.*

Proof. From equation (9) we have $\frac{\partial h}{\partial y}(x,y) = 2y\frac{\partial H_E}{\partial z}(x, x^2+y^2)$ and thus $\frac{\partial h}{\partial y} = 0 \Leftrightarrow \left(\frac{\partial H_E}{\partial z} = 0 \text{ or } y = 0\right)$. □

Lemma 2 (Singularities of the algebraic highlight oval)**.** *The algebraic highlight oval is singular at points where the cylinder cross-section is singular or has a contact point with the canonical parabola. The later type of singular points only occurs on the x axis of the $\mathcal{S}$ plane.*

Proof. A singularity occurs for $\frac{\partial h}{\partial x} = \frac{\partial h}{\partial y} = 0$ and thus, using equation (9), iff:

$$\begin{cases} \dfrac{\partial h}{\partial x} &= \dfrac{\partial H_E}{\partial x} + 2x\dfrac{\partial H_E}{\partial z} &= 0 \quad (10) \\ \dfrac{\partial h}{\partial y} &= 2y\dfrac{\partial H_E}{\partial z} &= 0. \quad (11) \end{cases}$$

We distinguish the two types of singular points from equation (11). The first type occurs when $\frac{\partial H_E}{\partial z} = 0$, implying from equation (10) that $\frac{\partial H_E}{\partial x} = 0$ and thus that the cylinder cross-section is also singular at this point. The second type occurs when $y = 0$. Noticing that the normal vector of the canonical parabola is $(2x, -1)$, equation (10) then means that the normal of the canonical parabola and of the cylinder cross-section must align. Because the cylinder cross-section is a conic it does not have inflection points. The second type of singularity is thus made of contact points. As a contact point is located on the canonical parabola it satisfies $z = x^2$ and thus $y = 0$ from equation (7). It thus projects to the x axis of the $\mathcal{S}$ plane. □

We now give three lemmas required to relate the results established for the cylinder cross-section to the algebraic highlight oval. These lemmas concern parts of the cross-section conic H_E represented by the 1D function $z = e(x)$ and relate it to parts of the algebraic highlight oval represented by the 1D function $y = b(x)$. Because the algebraic highlight oval is symmetric about the x axis, from proposition 5, we restrict function b to the positive side of the y axis. Because $x^2 + y^2 = z$ and $z = e(x)$ we have:

$$y = b(x) \stackrel{\text{def}}{=} \sqrt{e(x) - x^2}. \quad (12)$$

We call b the projected function. We characterize its shape from that of e. The projected function's derivative is:

$$\frac{\partial b}{\partial x}(x) = \frac{1}{2b(x)}\left(\frac{\partial e}{\partial x}(x) - 2x\right). \quad (13)$$

We first deal with the case of a concave function e, which turns out to be very simple. We then examine the more complex case of a convex function e. We use standard results on convex and concave functions which may be found in (Boyd and Vandenberghe, 2004). We use the sign function $\text{sign} : \mathbb{R} \to \{-1, 0, 1\}$.

Lemma 3 (The projected function for a concave cross-section function)**.** *The projected function b of a concave cross-section function e is concave. The opposite is not true.*

Proof. We have that $-x^2$ is concave and thus $e(x) - x^2$ is concave, as the sum of concave functions remains concave. Thus $b(x) = \sqrt{e(x) - x^2}$ is concave because the square root is concave and nondecreasing. The fact that the opposite is not true is a consequence of lemma 5. □

Lemma 4 (The sign function associated to a cross-section function)**.** *There exist a function γ called the sign function associated to a cross-section function e over the domain defined by $e(x) > x^2$ such that* $\text{sign}(\gamma) = \text{sign}\left(\frac{\partial^2 b}{\partial x^2}\right)$ *and* $\text{sign}\left(\frac{\partial \gamma}{\partial x}\right) = \text{sign}\left(\frac{\partial^3 e}{\partial x^3}\right)$.

Proof. We define $\hat{e}(x) \stackrel{\text{def}}{=} e(x) - x^2$ and so $\hat{e}(x) > 0$ and $b(x) = \sqrt{\hat{e}(x)}$. Differentiating we obtain:

$$\frac{\partial^2 b}{\partial x^2} = \frac{1}{2}\frac{1}{\sqrt{\hat{e}}}\left(\frac{\partial^2 \hat{e}}{\partial x^2} - \frac{1}{2}\left(\frac{\partial \hat{e}}{\partial x}\right)^2 \frac{1}{\hat{e}}\right).$$

We thus have:

$$\operatorname{sign}\left(\frac{\partial^2 b}{\partial x^2}\right) \;=\; \operatorname{sign}\left(\frac{\partial^2 \hat{e}}{\partial x^2} - \frac{1}{2}\left(\frac{\partial \hat{e}}{\partial x}\right)^2 \frac{1}{\hat{e}}\right) \;=\; \operatorname{sign}(\gamma) \qquad \text{with} \qquad \gamma \stackrel{\text{def}}{=} 2\hat{e}\frac{\partial^2 \hat{e}}{\partial x^2} - \left(\frac{\partial \hat{e}}{\partial x}\right)^2.$$

Differentiating, we arrive at:

$$\frac{\partial \gamma}{\partial x} \;=\; 2\hat{e}\frac{\partial^3 \hat{e}}{\partial x^3},$$

and so at $\operatorname{sign}\left(\frac{\partial \gamma}{\partial x}\right) = \operatorname{sign}\left(\frac{\partial^3 e}{\partial x^3}\right)$, because $\frac{\partial^3 \hat{e}}{\partial x^3} = \frac{\partial^3 e}{\partial x^3}$. □

Lemma 5 (The projected function for a convex cross-section function intersecting the canonical parabola twice)**.** *The projected function b of a convex cross-section function e that intersects the canonical parabola at x_1 and x_2, $x_1 < x_2$, and not anywhere else on $[x_1, x_2]$, which is within the parabola on $[x_1, x_2]$ and whose second derivative is convex, is concave on $[x_1, x_2]$. The opposite is not true.*

Proof. Using lemma 1 we have $\frac{\partial b}{\partial x}(x_1) = +\infty$ and $\frac{\partial b}{\partial x}(x_2) = -\infty$, from which we have $\frac{\partial^2 b}{\partial x^2}(x_1), \frac{\partial^2 b}{\partial x^2}(x_2) < 0$. We use lemma 4 to finish the proof. First, it tells us that we have $\gamma(x_1), \gamma(x_2) < 0$. Because $\frac{\partial^2 e}{\partial x^2}$ is convex, we have that $\frac{\partial^3 e}{\partial x^3}$ is increasing. If $\frac{\partial^2 e}{\partial x^2}$ does not have a minimum, $\frac{\partial^3 e}{\partial x^3}$ is *(i)* negative or *(ii)* positive on $[x_1, x_2]$. If $\frac{\partial^2 e}{\partial x^2}$ has a minimum at $x_0 \in]x_1, x_2[$, $\frac{\partial^3 e}{\partial x^3}$ is *(iii)* negative on $]x_1, x_0[$ and positive on $]x_0, x_2[$. Because $\operatorname{sign}\left(\frac{\partial \gamma}{\partial x}\right) = \operatorname{sign}\left(\frac{\partial^3 e}{\partial x^3}\right)$, γ is either *(i)* decreasing, *(ii)* increasing or *(iii)* decreasing and increasing with a minimum at x_0. Because $\gamma(x_1), \gamma(x_2) < 0$, γ is thus negative on $[x_1, x_2]$. We thus have that $\frac{\partial^2 b}{\partial x^2}$ is negative and that b is concave on $[x_1, x_2]$. The fact that the opposite is not true is a consequence of lemma 3. □

Lemma 6 (The projected function for a convex cross-section function intersecting the canonical parabola once)**.** *The projected function b of a convex cross-section function e whose derivative goes to negative infinity in x_1 and that intersects the canonical parabola at x_2, $x_1 < x_2$, and not anywhere between x_1 and x_2, and whose second derivative is convex, has a convex and a concave pieces.*

Proof. Using lemma 1 we have $\frac{\partial b}{\partial x}(x_1) = \frac{\partial b}{\partial x}(x_2) = -\infty$, from which we have $\frac{\partial^2 b}{\partial x^2}(x_1) = +\infty$ and $\frac{\partial^2 b}{\partial x^2}(x_2) = -\infty$. We use lemma 4 to finish the proof. First, it tells us that we have $\gamma(x_1) > 0$ and $\gamma(x_2) < 0$. Because $\frac{\partial^2 e}{\partial x^2}$ is convex, we have that $\frac{\partial^3 e}{\partial x^3}$ is increasing. If $\frac{\partial^2 e}{\partial x^2}$ does not have a minimum $\frac{\partial^3 e}{\partial x^3}$ is *(i)* negative or *(ii)* positive on $[x_1, x_2]$. If $\frac{\partial^2 e}{\partial x^2}$ has a minimum at $x_0 \in]x_1, x_2[$, $\frac{\partial^3 e}{\partial x^3}$ is *(iii)* negative on $]x_1, x_0[$ and positive on $]x_0, x_2[$. Because $\operatorname{sign}\left(\frac{\partial \gamma}{\partial x}\right) = \operatorname{sign}\left(\frac{\partial^3 e}{\partial x^3}\right)$, γ is either *(i)* decreasing, *(ii)* increasing or *(iii)* decreasing and increasing with a minimum at x_0. Because $\gamma(x_1) > 0$ and $\gamma(x_2) < 0$, case *(ii)* is impossible. In cases *(i)* and *(iii)*, γ is thus positive on $[x_1, x_3[$ and negative on $]x_3, x_2]$ with $x_3 \in]x_1, x_2[$ in case *(i)* and $x_3 \in]x_1, x_0]$ in case *(iii)*. We thus have that $\frac{\partial^2 b}{\partial x^2}$ is positive and b convex on $[x_1, x_3]$ and that $\frac{\partial^2 b}{\partial x^2}$ is negative and b concave on $[x_3, x_2]$. □

4.3 Properties of the Highlight Ovals in the Cross-Section Space

We first characterize the conic H_E in a proposition, with an illustration given by figures 5 and 6. We then give two lemmas which reveal the centre and axes of H_E when it is a central conic. In order to characterize the intersection points between H_A and H_E, we then find a degenerate member H_B of the pencil of conics (H_A, H_E). Excluding the cases $t \in \{-1, 0, 1\}$ which are solved in §3, we show that $\operatorname{rank}(B) \in \{1, 2\}$. For $\operatorname{rank}(B) = 1$ we show that H_A and H_E have only zero, one or two intersection points and give their coordinates. For $\operatorname{rank}(B) = 2$ however the problem is much more involved. We characterize the type of conic for H_B but leave the problem of characterizing the intersection points to the next section, dedicated to the case $1 > |t| > 0$.

Proposition 7 (The cylinder's cross-section H_E)**.** *For $a \neq 0$, we have that the cylinder's cross-section, represented by the conic H_E, is a hyperbola for $|t| > 1$, a parabola for $|t| = 1$, an ellipse for $1 > |t| > 0$ and degenerates to a horizontal line for $t = 0$. The ellipse is always real. For $a = 0$, the cylinder's cross-section always degenerates to a pair of real or imaginary lines.*

Proof. We have $\det(E) = -a^2t^4\left(4a^2+(f-g)^2\right)^2$ and $\det\left(\bar{E}\right) = -4a^2t^2\left(t^2-1\right)$. For $t \neq 0$, we thus have $\operatorname{sign}\left(\det\left(\bar{E}\right)\right) = \operatorname{sign}\left(1-t^2\right)$, from which the hyperbola, parabola and ellipse classification holds. The ellipse is always real because $(E_{1,1}+E_{2,2})\det(E) = -a^2t^4\left(4a^2t^2+1-t^2\right)\left(4a^2+(f-g)^2\right)^2 < 0$. For $t=0$ we have $E = ll^\top$ with $l^\top = [0\,1\,fg-a^2]$, which represents the line $z = a^2 - fg$. For $a = 0$ we have $\det(E) = \det\left(\bar{E}\right) = 0$ but $\operatorname{rank}(E) = 2$ in general. □

Lemma 7 (Centre of H_E for $|t| \neq 1$ and $a \neq 0$)**.** *For $|t| \neq 1$ and $a \neq 0$ we have that H_E is a central conic whose centre is:*

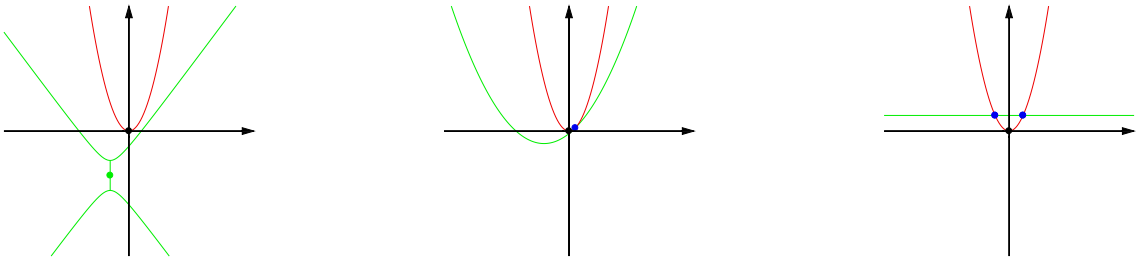
$$k \overset{\text{def}}{=} \begin{bmatrix} \frac{f^2-g^2}{4a} \\ \frac{2a^2\left(t^2+1\right)+t^2\left(f^2+g^2\right)-2fg}{2\left(1-t^2\right)} \end{bmatrix}.$$

Proof. Following proposition 7 we have $\det(E) \neq 0$ for $|t| \neq 1$ and $a \neq 0$, and thus that H_E is a central conic. The coordinates of the centre hold because $\frac{\partial H_E}{\partial x}(k) = \frac{\partial H_E}{\partial z}(k) = 0$. □

Lemma 8 (Axes of H_E for $|t| \neq 1$ and $a \neq 0$)**.** *For $|t| \neq 1$ and $a \neq 0$ we have that H_E is a central conic whose semi-axes along x and z are:*

$$\nu_x \overset{\text{def}}{=} \frac{4a^2+(f-g)^2}{4a\sqrt{1-t^2}} \qquad \text{and} \qquad \nu_z \overset{\text{def}}{=} \frac{t\left(4a^2+(f-g)^2\right)}{2\left(1-t^2\right)}.$$

Proof. Centrality for $|t| \neq 1$ and $a \neq 0$ hold from lemma 7. The semi-axes hold because from proposition 6 H_E is aligned with the coordinate axes and $H_E(k_x \pm \nu_x, k_z) = H_E(k_x, k_z \pm \nu_z) = 0$. □



Case $|t| > 1$ Case $|t| = 1$ Case $t = 0$

Figure 6: **The cylinder cross-section H_E in the cross-section space.** Together with figure 5 this figure shows the cylinder cross-section for the four main cases. From left to right, the cylinder's cross-section is a hyperbola, a parabola and a horizontal line, as shown by proposition 7. The algebraic highlight oval is respectively empty, a single singular point on the x axis and a circle centred at the origin, as shown in theorem 1. The ellipse case for $1 > |t| > 0$ is shown in figure 5.

Lemma 9 (A degenerate member of the pencil of conics (H_A, H_E))**.** *A degenerate member of the pencil of conics (H_A, H_E) is H_B with:*

$$B \overset{\text{def}}{=} \begin{bmatrix} -t^2(f-g)^2 & 0 & at^2\left(g^2-f^2\right) \\ 0 & 1-t^2 & \left(t^2-1\right)\left(a^2-fg\right) \\ at^2\left(g^2-f^2\right) & \left(t^2-1\right)\left(a^2-fg\right) & \left(1-t^2\right)\left(f^2g^2+a^4\right) - a^2\left(t^2\left(f^2+g^2\right)+2fg\right) \end{bmatrix}.$$

Proof. We define the pencil $B_\lambda = E + \lambda A$, and search its degenerate members such that $\det(B_\lambda) = 0$. This yields a cubic in λ. Amongst the three roots, one is always real, and is given by $\mu = -t^2\left(4a^2+(f-g)^2\right)$. It is easy to verify that we then arrive at $B = B_\mu$. □

The next two lemmas characterize the rank of matrix B for all possible combinations of the quartic's four parameters.

Lemma 10 (Rank of matrix B for $t \in \{-1,0,1\}$)**.** *For $t \in \{-1,0,1\}$ we have* $\operatorname{rank}(B) < 2$*. For $t = 0$ we have* $\operatorname{rank}(B) = 1$*. For $|t| = 1$ we have* $\operatorname{rank}(B) = 0$ *iff $f = g$ and ($a = 0$ or $f = g = 0$) and* $\operatorname{rank}(B) = 1$ *otherwise.*

Proof. Our proof uses simple substitutions in B. For $t = 0$ we have:

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & fg - a^2 \\ 0 & fg - a^2 & \left(fg - a^2\right)^2 \end{bmatrix},$$

whose rank is always 1. For $|t| = 1$ we have:

$$B = -\begin{bmatrix} (f-g)^2 & 0 & a\left(f^2 - g^2\right) \\ 0 & 0 & 0 \\ a\left(f^2 - g^2\right) & 0 & a^2(f+g)^2 \end{bmatrix},$$

whose rank is at most 1. More precisely, it drops to 0 iff all entries vanish. This implies $f = g$ from the top-left entry and $a = 0$ or $f = g = 0$ from the other non-zero entries. □

Lemma 11 (Rank of matrix B for $t \notin \{-1,0,1\}$)**.** *For $t \notin \{-1,0,1\}$ we have* $\operatorname{rank}(B) = 2$*, except if $f = g$ and ($a = 0$ or $f = g = 0$) for which* $\operatorname{rank}(B) = 1$*. In other words we have* $\operatorname{rank}(B) = 2$ *iff $f \neq g$ or ($a \neq 0$ and ($f \neq 0$ or $g \neq 0$)) and we have* $\operatorname{rank}(B) = 1$ *iff ($f = g$ and $a = 0$) or $f = g = 0$.*

Proof. We have $\operatorname{rank}(B) < 3$ by construction and $\operatorname{rank}(B) > 0$ because the central element of B never vanishes for $|t| \neq 1$. We then have $\operatorname{rank}(B) = 2$ iff at least one 2×2 minor does not vanish. The leading minor is $\det\left(\bar{B}\right) = t^2\left(t^2 - 1\right)(f - g)^2$. It can only vanish if $f = g$, which forms a necessary condition for $\operatorname{rank}(B) < 2$. Substituting f by g in B, we have:

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 - t^2 & \left(1 - t^2\right)\left(g^2 - a^2\right) \\ 0 & \left(1 - t^2\right)\left(g^2 - a^2\right) & \left(1 - t^2\right)\left(g^4 + a^4\right) - 2a^2g^2\left(1 + t^2\right) \end{bmatrix}, \tag{14}$$

whose single non-zero minor is $4a^2g^2t^2\left(t^2 - 1\right)$. It can only vanish if $a = 0$ or $g = 0$, which form a necessary condition for $\operatorname{rank}(B) < 2$. Combining the two necessary conditions concludes the proof. □

The next two lemmas analyze which conic type is H_B for $t \notin \{-1,0,1\}$.

Lemma 12 (Type of conic H_B for $t \notin \{-1,0,1\}$ and $\operatorname{rank}(B) = 1$)**.** *For $t \notin \{-1,0,1\}$ and* $\operatorname{rank}(B) = 1$ *we have that H_B represents a horizontal line.*

Proof. Lemma 11 tells us we have two cases. The first case is $f = g = 0$, for which we have:

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 - t^2 & -\left(1 - t^2\right)a^2 \\ 0 & -\left(1 - t^2\right)a^2 & a^4\left(1 - t^2\right) \end{bmatrix} \propto \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -a^2 \\ 0 & -a^2 & a^4 \end{bmatrix},$$

which represent the horizontal line $z = a^2$. The second case is $f = g$ and $a = 0$, for which we have:

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 - t^2 & \left(1 - t^2\right)g^2 \\ 0 & \left(1 - t^2\right)g^2 & g^4\left(1 - t^2\right) \end{bmatrix} \propto \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & g^2 \\ 0 & g^2 & g^4 \end{bmatrix},$$

and represents the horizontal line $z = -g^2$. □

Lemma 13 (Type of conic H_B for $\operatorname{rank}(B) = 2$)**.** *For* $\operatorname{rank}(B) = 2$ *we have $t \notin \{-1,0,1\}$ and H_B represents a single point for $|t| > 1$ and $f \neq g$, a pair of skew lines for $|t| < 1$ and $f \neq g$, the empty set for $|t| > 1$ and $f = g$ and a pair of horizontal lines for $|t| < 1$ and $f = g$.*

Proof. For $\text{rank}(B) = 2$ lemma 10 implies $t \notin \{-1, 0, 1\}$. The leading minor of B is $\det\big(\bar{B}\big) = t^2\left(t^2-1\right)(f-g)^2$. We then have two cases. The first case is when the leading minor is non-zero and thus $f \neq g$. We then have $\text{sign}\big(\det\big(\bar{B}\big)\big) = \text{sign}(|t|-1)$. It is thus positive if $|t| > 1$ and H_B is then a single point, and negative if $|t| < 1$ and H_B is then a pair of skew lines. The second case is when the leading minor vanishes and thus $f = g$. In this case we have B given by equation (14). We have $B_{1,3}^2 + B_{2,3}^2 - (B_{1,1} + B_{2,2})\, B_{3,3} = 4a^2g^2t^2(1-t^2)$, whose sign is given by $\text{sign}(1-|t|)$. It is thus negative if $|t| > 1$ and H_B is then imaginary, and negative if $|t| < 1$ and H_B is then a pair of horizontal lines. □

The next two lemmas respectively characterize and reveal the intersection points between H_E and H_A for $\text{rank}(B) = 1$ and complete lemma 13 by giving a further characterization of H_B for $\text{rank}(B) = 2$ and $f \neq g$.

Lemma 14 (Intersection points between H_E and H_A for $t \notin \{-1, 0, 1\}$, $\text{rank}(B) = 1$)**.** *For $t \notin \{-1, 0, 1\}$ and $\text{rank}(B) = 1$ we have at most two intersection points between H_E and H_A. These are located at $[\pm a\, a^2]^\top$ if $f = g = 0$.*

Proof. Lemma 11 tell us that we have two case. The first case is $f = g = 0$ and the second case is $a = 0$ and $f = g$. In the first case, we have from lemma 12 that H_E and H_A always intersect in points $[\pm a\, a^2]^\top$. In the second case, we have from lemma 12 that H_E and H_A intersect at the origin if $f = g = 0$. □

Lemma 15 (Further characterization of H_B for $\text{rank}(B) = 2$ and $f \neq g$)**.** *For $\text{rank}(B) = 2$ we have $t \notin \{-1, 0, 1\}$. For $f \neq g$ we have two cases. For $|t| > 1$, H_B is the point w and for $1 > |t| > 0$, H_B is a pair of lines meeting at point w, with $w^\top \stackrel{\text{def}}{=} \left[\frac{a(f+g)}{g-f}\ a^2 - fg\right]$.*

Proof. For $\text{rank}(B) = 2$ lemma 10 implies $t \notin \{-1, 0, 1\}$. Lemma 13 then shows that H_B is either a point for $|t| > 1$ or a pair of skew lines for $1 > |t| > 0$. The kernel of matrix B give the homogeneous coordinates of the point in the former case or the meeting point in the latter case. These coordinates are:

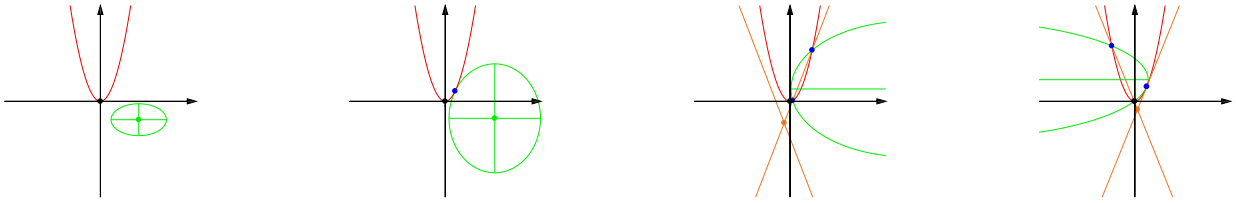
$$W \stackrel{\text{def}}{=} \begin{bmatrix} a(f+g) \\ (g-f)\left(a^2 - fg\right) \\ g-f \end{bmatrix}. \tag{15}$$

Because $f \neq g$, W can be dehomogenized and gives w. □

5 Proof of Theorem 1, $1 > |t| > 0$

This section provides our proof of theorem 1 for the case $1 > |t| > 0$. Our proof rests on the cross-section space representation and accompanying results of §4. We have from proposition 7 that H_E represents an ellipse for $a \neq 0$ and a pair of horizontal lines for $a = 0$. Our proof requires eleven new lemmas which we give directly below. An illustration is provided in figure 7. The first one, lemma 16, holds for $a \neq 0$ and characterizes the position of the centre of H_E with respect to H_A. It is used in the first and second parts of our proof, concerning respectively the cases $\text{rank}(B) = 1$ and $\text{rank}(B) = 2$. The next ten lemmas hold for $\text{rank}(B) = 2$ and concern the second part of our proof, which follows an analysis based on the number n of intersection points between H_E and H_A. Lemmas 17 and 18 establish a formula for $n \in [0, 4]$. Lemma 19 shows that H_E can never be entirely in H_A for any value of n. This lemma forms the key element of our proof for the cases $n \in [0, 2]$. The next three lemmas, lemmas 20, 21 and 22, show where the intersection points may be located for $n = 4$. The last four lemmas, lemmas 23, 24, 25 and 26, concern the case $n = 3$. These last seven lemmas form the key elements of our proof for the cases $n \in [3, 4]$.

Lemma 16 (Position of the centre of the ellipse H_E relative to the parabola H_A for $1 > |t| > 0$, $a \neq 0$)**.** *We define $r \stackrel{\text{def}}{=} 4a^2\left(t^2+1\right) + (f+g)^2\left(t^2-1\right)$. For $1 > |t| > 0$, $a \neq 0$, we have that the centre of H_E is inside the parabola H_A for $r > 0$, on the parabola H_A for $r = 0$ and outside the parabola H_A for $r < 0$.*



(a) – $n = 0$ (b) – $n = 1$ (c) – $n = 2$ (d) – $n = 3$

Figure 7: **The cross-section space for $1 > |t| > 0$.** For $a \neq 0$, H_E is an ellipse aligned with the coordinate axes. The ellipse is never entirely inside the canonical parabola H_A. The topology and convexity of the algebraic highlight oval depend on the number of intersection points n between H_A and H_E, with $0 \leq n \leq 4$. For $n = 0$, as in (a), the algebraic highlight oval is empty. For $n = 1$, as in (b), the algebraic highlight oval is a single point in the x axis. For $n = 2$, as in (c), the algebraic highlight oval is a closed curve with 0, 2 or 4 inflection points. In this example it is the curve with 2 inflection points shown on the left of the bottom row of figure 1. For $n = 3$, as in (d), the algebraic highlight oval has a singular point forming a self-intersection of the curve or being an isolated point and has 0, 2 or 4 inflection points. In this example it has a self-intersection point and no inflection points and is shown on the bottom row of figure 1. For $n = 4$, as illustrated in figure 5, the algebraic highlight oval is a pair of nested closed curves.

Proof. We test the sign of $z - x^2$ for the centre k, where positive means inside, null means on, and negative means outside the parabola H_A. This gives:

$$\frac{\left(4a^2 + (f-g)^2\right)\left(4a^2\left(t^2+1\right) + (f+g)^2\left(t^2-1\right)\right)}{16a^2\left(1-t^2\right)}$$

The denominator and the leading factor in the numerator are always positive. We thus have that only the value of r, the second factor in the numerator, matters. □

Lemma 17 (Factorization of B for $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *We have $B = \frac{1}{2}\left(lm^\top + ml^\top\right)$ for $l, m \in \mathbb{R}^3$ given by:*

$$l \stackrel{\text{def}}{=} \begin{bmatrix} t(f-g) \\ -\sqrt{1-t^2} \\ at(f+g) + \sqrt{1-t^2}\left(a^2 - fg\right) \end{bmatrix} \quad \text{and} \quad m \stackrel{\text{def}}{=} \begin{bmatrix} t(f-g) \\ \sqrt{1-t^2} \\ at(f+g) - \sqrt{1-t^2}\left(a^2 - fg\right) \end{bmatrix}$$

For $\text{rank}(B) = 2$ we have that l and m do not vanish and are linearly independent.

Proof. The expression of l and m can be simply verified by reforming B. For $\text{rank}(B) = 2$, their linear independence is proved as follows. We first note that $l_2 = -m_2 \neq 0$. From lemma 11, we then have two cases. The first case is $f \neq g$, from which we have that $l_1 = m_1 \neq 0$ and ensures linear independence. The second case is $a \neq 0$ and ($f \neq 0$ or $g \neq 0$). If $f \neq g$ we are back to the first case. If $f = g$ we have that $l_3 = u + u'$ and $m_3 = u - u'$ with $u = 2atf \neq 0$ and $u' = \sqrt{1-t^2}\left(a^2 - f^2\right)$. We then have $l_2 m_3 - l_3 m_2 = 2l_2 u \neq 0$, which ensures linear independence. □

Lemma 18 (Number n of intersection points between H_E and H_A for $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *There are $n \in [0, 4]$ intersection points between H_E and H_A. This number depends on the discriminants d_l and d_m:*

$$\frac{1}{2}d_{l,m} \stackrel{\text{def}}{=} 2\left(a^2 - fg\right) + t^2\left(\frac{1}{2}(f+g)^2 - 2a^2\right) \pm 2at\sqrt{1-t^2}(f+g).$$

More precisely, we have:

$$n \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } d_l < 0 \\ 1 & \text{if } d_l = 0 \\ 2 & \text{if } d_l > 0 \end{cases} + \begin{cases} 0 & \text{if } d_m < 0 \\ 1 & \text{if } d_m = 0 \\ 2 & \text{if } d_m > 0 \end{cases} - \begin{cases} 0 & \text{if } fg \neq 0 \\ 1 & \text{if } fg = 0. \end{cases}$$

Proof. We have that H_B is the pair of lines l, m. A point with homogeneous coordinates $V \in \mathbb{R}^3$ on one of the lines, say l, satisfies $l^\top V = l_1 V_x + l_2 V_z + l_3 = 0$. Substituting the constraint $H_A(V_x, V_z) = 0$ as $V_z = V_x^2$ we have $l_2 V_x^2 + l_1 V_x + l_3 = 0$. The discriminant of this quadratic is $d_l = l_1^2 - 4 l_2 l_3$, giving $d_{l,m}$ for l, m. Each line thus contributes between 0 and 2 intersection points, depending on the sign of its discriminant: $d_l < 0$ contributes no intersection point, $d_l = 0$ contributes one intersection point and $d_l > 0$ contributes two intersection points. The same rule applies to d_m. However, the two lines may share an intersection point, which must not be counted twice. This happens if their own intersection point lies on H_A. The homogeneous coordinates W of this intersection point are given by equation (15). It thus lies on H_A iff:

$$W^\top A W \;=\; fg\left(4a^2 + (f-g)^2\right) \;=\; 0.$$

Because the second term never vanishes as, from lemma 11, $\text{rank}(B) = 2$ implies $f \neq g$ or $a \neq 0$, it follows that there is a shared intersection point only if $f = 0$ or $g = 0$. □

Lemma 19 (Position of the ellipse H_E relative to the parabola H_A for $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$ and $\text{rank}(B) = 2$ we have that H_E is never entirely inside H_A. This holds for any number of contact points $\{0, 1, 2\}$ between H_E and H_A.*

Proof. Our proof has two parts. The first part deals with the case $a = 0$. From proposition 7 we have that H_E is a pair of horizontal lines. We thus have that H_E cannot be entirely inside H_A, whatever the number of contact points. The second part deals with the case $a \neq 0$. This part of the proof requires us to define the height space and is provided in Appendix A. □

Lemma 20 (Position of the n intersection points along axis x for $n = 4$, $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$, $\text{rank}(B) = 2$ and $n = 4$ intersection points between H_E and H_A we always have at least one intersection point on each side of axis z. In other words, we have at least one intersection point with $V_x > 0$ and one with $V_x < 0$.*

Proof. In order to have $n = 4$ we must have $d_l, d_m > 0$. Following the proof of lemma 18 we have $V_x = -\frac{1}{2l_2}\left(l_1 \pm \sqrt{d_l}\right)$ and $V_x = -\frac{1}{2m_2}\left(m_1 \pm \sqrt{d_m}\right)$. We have from lemma 17 that $l_1 = m_1$ and $l_2 = -m_2 < 0$. Therefore, we have $V_x = c \pm e_l$ and $V_x = -c \pm e_m$ with $c = -\frac{l_1}{2l_2}$, $e_l = \frac{\sqrt{d_l}}{2l_2} < 0$ and $e_m = \frac{\sqrt{d_m}}{2m_2} > 0$. We conclude with two cases: if $c \geq 0$ then $c - e_l > 0$ and $-c - e_m < 0$ and if $c \leq 0$ then $c + e_l < 0$ and $-c + e_m > 0$. □

Lemma 21 (Position of the n intersection points along axis z for $n = 4$, $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$, $\text{rank}(B) = 2$ and $n = 4$ intersection points between H_E and H_A the two points with largest magnitude in $z = x^2$ are always located on different sides of axis z. In other words, the outer points lie above the inner points along z.*

Proof. We follow the notation from the proof of lemma 20. We have three cases depending on c. For $c > 0$ we have three subcases. The first subcase is $-e_l > e_m$. The largest V_x in magnitude is then $c - e_l > 0$ and the second largest is $\min(-c - e_m, c + e_l) < 0$. The second subcase is $-e_l = e_m$, for which the largest and second largest have the same magnitude and are $c - e_l > 0$ and $-c - e_l < 0$. The third subcase is $-e_l < e_m$. The largest V_x in magnitude is then $-c - e_m < 0$ and the second largest is $\max(c - e_l, -c + e_m) > 0$. The cases $c = 0$ and $c < 0$ are solved very similarly. □

Lemma 22 (Position of the two inner intersection points along axis z for $n = 4$, $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$, $\text{rank}(B) = 2$ and $n = 4$ intersection points between H_E and H_A the two points with lowest magnitude in $z = x^2$ are always located below the horizontal axis of ellipse H_E.*

Proof. Our proof is based on showing that assuming otherwise leads to a contradiction. Imagine an inner intersection point V, say on the positive side of axis x, lies above the horizontal axis of H_E. This implies that the right-most tip $[k_x + \nu_x\, k_z]^\top$ of H_E lies outside, to the right of the parabola H_A. This further implies that there cannot possibly be another intersection point on the positive side of axis x which would be to the right of V. In other words, this means that there would not be an outer intersection point. So we must have that the intersection point V is an outer intersection point or that it lies below the horizontal axis of H_E. □

Lemma 23 (Position of the n intersection points along axis x for $n = 3$, $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$, $\text{rank}(B) = 2$ and $n = 3$ intersection points between H_E and H_A we always have at least one intersection point on each side of axis z or at the origin. In other words, we have at least one intersection point with $V_x \geq 0$ and one with $V_x \leq 0$.*

Proof. Our proof is based on the proof of lemma 20. The transition from $n = 4$ to $n = 3$ is by collapsing a pair of intersection points. The first way is with $d_l > 0$ becoming $d_l = 0$. We have two cases: if $c \geq 0$ then we have $V_x = c \geq 0$ and $V_x = -c + e_l < 0$ and if $c \leq 0$ then we have $V_x = c \leq 0$ and $V_x = -c - e_l > 0$. The second way is with $d_m > 0$ becoming $d_m = 0$, which leads to the same conclusion. The third way is with $d_l, d_m > 0$ and two points collapsing across the two pairs of solutions. Following the proof of lemma 20 this leads to one intersection point with $V_x > 0$ and one with $V_x < 0$. □

Lemma 24 (Position of H_E relative to H_A for $n = 3$ intersection points, $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$, $\text{rank}(B) = 2$ and $n = 3$ intersection points between H_E and H_A we have two possibilities. If the contact point is the leftmost or rightmost point, then H_E is inside H_A around the contact point. If the contact point is the middle point, then H_E is outside H_A around the contact point.*

Proof. In the first case, imagine H_E is outside H_A around the contact point, which we assume to be the leftmost one. Then H_E is oriented either opposite or towards H_A. In the former case it creates no other intersection point. In the latter case it creates an intersection point at the left of the contact point. This is contradictory in both cases, so H_E must be inside H_A around the contact point. In the second case, imagine H_E is inside H_A around the contact point. This means that H_E exits H_A to create an intersection point, say to the left of the contact point. However, this implies that H_E creates an extra, leftmost intersection point, which forms a contradiction. Thus, H_E must be outside H_A around the contact point. □

Lemma 25 (Type of the n intersection points for $n = 3$, $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$, $\text{rank}(B) = 2$ and $n = 3$ intersection points between H_E and H_A there is one contact point and two non-contact points.*

Proof. Imagine that we had three contact points. We have from lemma 24 that H_E would be inside H_A around the outer contact points and outside around the middle contact points. Because H_E is a continuous curve it means that this implies the existence of two additional intersection points, which is a contradiction. Consequently, for $n = 3$ intersection points we must have one contact and two non-contact intersection points. □

Lemma 26 (Position of the contact point for $n = 3$, $1 > |t| > 0$, $\text{rank}(B) = 2$)**.** *For $1 > |t| > 0$, $\text{rank}(B) = 2$ and $n = 3$ intersection points between H_E and H_A we have that the contact point is always located below the horizontal axis of H_E.*

Proof. We take without loss of generality the case of a contact point with a positive x coordinate. We have that the slope of the tangent to H_A at the contact point is positive too, and so does the slope of the tangent to H_E. Because H_E is an ellipse whose axes are aligned with the coordinate axes, its tangent has a positive slope only in its top-left or bottom-right quadrants. If the contact point were in the top-left quadrant, then this would contradict the hypothesis of $n = 3$ intersection points. Therefore, the contact point must be in the bottom-right quadrant, below the horizontal axis. □

Proof of theorem 1, $1 > |t| > 0$. We have from lemma 11 that $\text{rank}(B) \in \{1, 2\}$. For $\text{rank}(B) = 1$ we have two cases. For $a = 0$ lemma 14 tells us we have one intersection point at the origin if $f = g = 0$. Substituting in E we have $E = \text{diag}(0, 1 - t^2, 0)$, which represented the line $z = 0$. The algebraic highlight oval is thus formed of the origin. For $a \neq 0$ lemma 14 tells us we have two symmetric intersection points at $x = \pm a$ if $f = g = 0$. From proposition 7 H_E is an ellipse. Substituting $f = g = 0$ in r from lemma 16 we find $r = 4a^2\left(t^2 + 1\right) > 0$. This means that the ellipse H_E is inside H_A and consequently that the two intersection points are contact points. The algebraic highlight oval is thus a pair of circles. For $\text{rank}(B) = 2$ we follow the number $n \in [0, 4]$ of intersection points between H_E and H_A and use lemmas 1 to 6 to conclude on the topology and convexity of the algebraic highlight oval:

- For $n = 0$ intersection point, we have from lemma 19 that H_E is always outside H_A. Therefore, the algebraic highlight oval is always empty.

- For $n = 1$ intersection point, we have from lemma 19 that H_E is always outside and touching H_A at the intersection point, which is thus a contact point. Therefore, we have from lemma 2 that the algebraic highlight oval is made of a single singular point on the x axis. In the formula for n from lemma 18 there are only two ways to obtain $n = 1$, which are $d_l = 0$, $d_m < 0$ and $d_l < 0$, $d_m = 0$. This is because for $d_l = d_m = fg = 0$ we would have that l and m would represent the same tangent to H_A, contradicting the hypothesis $\text{rank}(B) = 2$ from lemma 17, and configurations $d_l > 0$, $d_m < 0$, $fg = 0$ and $d_l < 0$, $d_m > 0$, $fg = 0$ are impossible because $fg = 0$ implies $d_l \geq 0$ and $d_m \geq 0$. The contact point is $\left[\frac{t(f-g)}{\sqrt{1-t^2}}\, 0\right]^\top$ if $d_l = 0$ and $\left[-\frac{t(f-g)}{\sqrt{1-t^2}}\, 0\right]^\top$ if $d_m = 0$. This is because if $d_l = 0$, we can solve for $V_x = -\frac{l_1}{2l_2}$ following the proof of lemma 18, and substitute the expressions of l_1 and l_2 from lemma 17. The same reasoning holds for $d_m = 0$ with $V_x = -\frac{m_1}{2m_2}$.

- For $n = 2$ intersection points, we have that these are not contact points because that would imply H_E to be entirely inside H_A, which from lemma 19 is impossible. We then have from lemmas 3 and 6 that the algebraic highlight oval is a closed curve with 0, 2 or 4 inflection points.

- For $n = 3$ intersection points, lemma 25 tells us that we have one contact point. We then have two cases, depending on whether the contact point is between the two non-contact intersection points or not. In the first case, lemma 24 tells us that the ellipse is inside the parabola around the contact point. Therefore, the contact point creates a self-intersection of the algebraic highlight oval. Using lemmas 3, 5 and 6, we thus have that the algebraic highlight oval has 0 or 2 inflection points. In the second case, lemma 24 tells us that the ellipse is outside the parabola around the contact point. Therefore, the contact point creates an isolated singular point of the algebraic highlight oval. Using lemmas 3, 5 and 6, we thus have that the algebraic highlight oval has 0, 2 or 4 inflection points.

- For $n = 4$ intersection points, we have no contact points, and at least one outer intersection point on each side of the z axis. The lower and upper parts of the ellipse thus yield two nested closed curves for the algebraic highlight oval. The lower part yields the inner closed curve which is convex, thanks to lemma 22. The upper part yields the outer closed curve and is analyzed in three cases, depending on whether the outer intersection points are above or below the ellipse's horizontal axis. In the first case, they are both above, and the outer curve is convex, thanks to lemma 5. In the second case, one is above and the other below, and the outer curve has 2 inflection points, thanks to lemmas 3 and 6. In the third case, they are both below, and the outer curve has 4 inflection points, again thanks to lemmas 3 and 6.

□

6 The Highlight Ovals in Phong Conditions

We formally define Phong conditions and show that the highlight ovals then have specific properties, from which we can prove proposition 3 about the topology and convexity of the geometric highlight ovals in Phong conditions.

Definition 3 (Phong conditions)**.** *In Phong's model the viewpoint and light source lie on the same side of the plane $\mathcal{S}$ and off this plane. We therefore have $fg < 0$.*

Lemma 27 (The splitting circle)**.** *We define the splitting circle as centred at the origin and with radius $\sqrt{a^2 - fg}$. This radius is always positive in Phong conditions. The splitting circle is the geometric and algebraic highlight ovals obtained for $t = 0$. For $t > 0$ the geometric highlight oval is outside the splitting circle and for $t < 0$ it is inside the splitting circle.*

Proof. We rewrite the geometric highlight oval equation (1) in standard form as:

$$t\, d_F(P)\, d_G(P) \;=\; x^2 + y^2 + fg - a^2.$$

Clearly, for $t = 0$ the equation represents the splitting circle. We have $\text{sign}(t\, d_F(P)\, d_G(P)) = \text{sign}(t)$. Therefore, we have $t > 0$ iff $x^2 + y^2 > a^2 - fg$, meaning if point p is outside the splitting circle, and similarly we have $t < 0$ iff point p is inside the splitting circle. □

Lemma 28 (Position of the intersection point of l and m relative to the parabola H_A in Phong conditions)**.** *In Phong conditions, the intersection point of l and m is always inside H_A.*

Proof. The intersection point of l and m is given by the kernel W of B, whose expression is given by equation (15). We then have $W^\top A W \propto fg < 0$, where the inegality is simply due to Phong conditions. □

Lemma 29 (The algebraic highlight oval in Phong conditions for $1 > |t| > 0$)**.** *In Phong conditions, the algebraic highlight oval for $1 > |t| > 0$ is always made of a convex closed curve nested in a closed curve. The inner curve is within the splitting circle and the outer curve is outside the splitting circle.*

Proof. We have that the lines l and m are neither horizontal nor vertical, and distinct, which is easy to find out from their expressions given by lemma 17. From lemma 28, their intersection point is always inside H_A. In pseudo-3D form, the splitting circle is represented by a plane of height $z = a^2 - fg > 0$. This plane contains point W, which can be verified from its z-coordinate being $a^2 - fg$, from equation (15). Consequently, we have that H_A and H_E always have four intersection points. Two of them are located below and the two others above the plane of height $z = a^2 - fg$. From lemma 5, the algebraic highlight oval is thus formed of a convex closed curve, nested, from lemmas 3 and 6, in a closed curve with 0, 2 or 4 inflection points. From lemma 27 the inner curve is within the splitting circle and corresponds to the geometric highlight oval with $t < 0$ and the outer closed curve is outside the splitting circle and corresponds to the geometric highlight oval with $t > 0$. □

Lemma 30 (The centre point)**.** *We define the centre point as the point $\left[a\frac{g+f}{g-f}\ 0\right]^\top$. This point is always real and finite in Phong conditions. The centre point is the geometric highlight oval and quartic obtained for $t = 1$. For $1 > |t| > 0$ the centre point is always inside the inner convex closed curve part of the quartic.*

Proof. We have from equation (15) that $a\frac{g+f}{g-f}$ is the x coordinate of point W, the intersection point of lines l and m, which from lemma 28 is always inside H_A in Phong conditions. The slope of l and m is equal up to sign, and given by $\pm\frac{t(g-f)}{\sqrt{1-t^2}}$. Consequently, point W lies inbetween the pair of lower intersection points and inbetween the pair of upper intersection points, and so does the centre point, its projection on the x axis. □

Proof of proposition 3. Following the identification $t = -\sqrt[\alpha]{\frac{\tau}{j}}$ from §2.2, we have that $\tau > j$ means $t < -1$ and that $\tau = j$ means $t = -1$. We have, using Phong conditions and theorem 1, that the algebraic and thus also the geometric highlight ovals, are empty in the former case and reduce to the centre point in the latter case. On the other hand, we have that $j > \tau > 0$ means $-1 < t < 0$. We have from lemma 29 that the algebraic highlight oval is formed of a convex closed curve nested in a closed curve. The inner curve is within the splitting circle and the closed curve is outside, from lemma 27. They both contain the centre point, from lemma 30. The expressions for the centre point and the splitting circle are directly given by following the definition of the standard form (6). Finally, we have that $\tau = 0$ means $t = 0$, from which we obtain the splitting circle, from lemma 27. □

It is interesting to note that proposition 3 implies that under Phong conditions the geometric highlight oval is the boundary of a so-called spectrahedron, meaning that it can be described by a linear matrix inequality (Helton and Vinnikov, 2007).

7 Experimental Evaluation

The algebraic highlight ovals show that Phong's specular highlight model, albeit not algebraic, is very tighly connected to algebraic curves. Our results reveal the nature of this connection very precisely, showing that these algebraic curves form a special type of bicircular quartics given by equation (6). Beside their theoretical interest, the highlight ovals may be fitted to real images of specular highlight.

7.1 Image Acquisition

The image acquisition setup is illustrated by figure 8. Typical specular materials are plastic, metal and glossy paper. The latter is commonly found in magazines. Images taken in regular conditions usually show specular highlights as white spots caused by sensor saturation. Saturation causes a loss of information in the centre of the specular highlight. However, information is still present on the 'side' of the saturated area. Interestingly, saturation can be largely avoided by taking images using the fast shutter or the high dynamic range modes. We have thus taken images in two different regimes: regular and fast-shutter. In the *regular regime*, the specular highlights are white spots and the models are fitted to sample points taken at the boundary between the saturated and non-saturated zones. In the *fast-shutter regime*, the specular highlights follow a smooth decrease away from their brighest point and the models are fitted to sample points sharing the same intensity.

We have used objects made of metal, plastic and glossy paper. More specifically, we have used 10 images of a metal shelf, a plastic box and a book for each regime, making a total of 60 images. We used a DSLR camera Nikon D3100. The images were taken indoor under various types of ambient light from regular to dark and a flashlight meant to create specular highlights. In the fast-shutter regime, we specifically set the shutter speed between $\frac{1}{20}$ and $\frac{1}{200}$ seconds depending on the amount of ambient light. The image resolution is 1152×768 pixels. Strictly speaking, it would be required to 'rectify' the image to make the observation virtually fronto-parallel, but we found this was not necessary in practice.

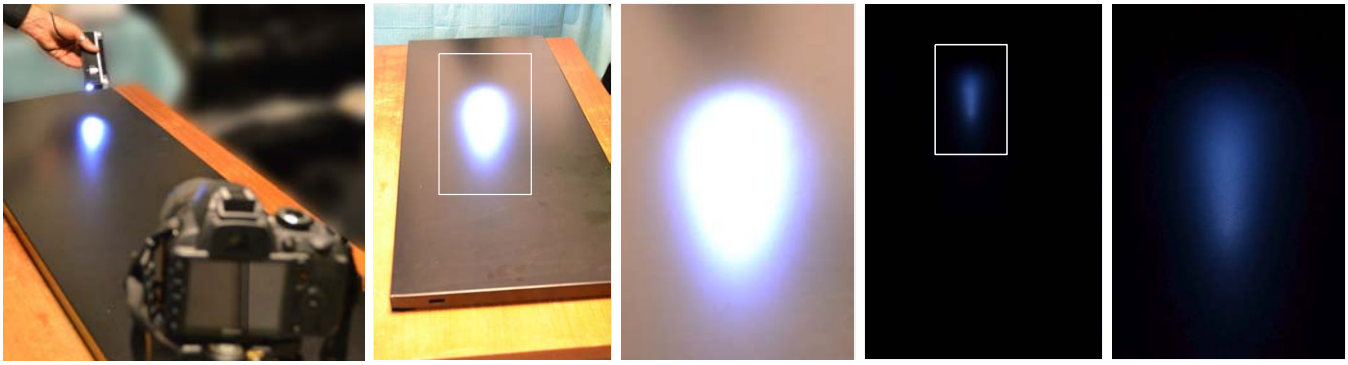


Figure 8: **Acquisition setup and example images.** From left to right: the image acquisition setup shown for the metal shelf, an example image in the regular and fast-shutter regimes with close-up on the detected specular highlight.

7.2 Evaluated Models

We evaluated the proposed algebraic highlight oval model of specular isocontours and compared it with the ellipse. The ellipse is a natural model to try as many specular highlights 'look' elliptical. A formal derivation was given in (Forsyth and Ponce, 2003). In addition, we use the ellipse as a means to initialize the algebraic highlight oval, as described in the next section. Concretely, the algebraic highlight oval model we fit is obtained in two steps. First, we rewrite the standard form (6) as a linear function of five unknown homogeneous coefficients $\beta_1, \ldots, \beta_5$ encapsulating the four intrinsic parameters:

$$h(x,y) \;=\; \beta_1\left(x^2+y^2\right)^2 + \beta_2 x^2 + \beta_3 y^2 + \beta_4 x + \beta_5.$$

Second, because the standard form assumes that the symmetry axis of the highlight oval is aligned with the x axis and that the highlight oval is centred as the origin, we introduce a simple Euclidean transformation of the plane, parameterized by a translation $t_x, t_y \in \mathbb{R}$ and a rotation angle $\theta \in \mathbb{R}$, representing the three extrinsic parameters, as:

$$h'(x, y) \quad = \quad h(\cos(\theta)x + \sin(\theta)y + t_x, -\sin(\theta)x + \cos(\theta)y + t_y).$$

7.3 Model Fitting Procedure

Our fitting procedure takes an image and an isocontour intensity level parameter t' as inputs. It produces an optimal estimate of the ellipse and algebraic highlight oval models, where optimal means maximum likelihood under the assumption of an Independently and Identically Distributed (IID) Gaussian noise on the input sample points. Our procedure has three main steps. The first step is to extract sample points on the isocontour. The second step is to optimally fit the ellipse to these sample points. The third step is to optimally fit the algebraic highlight oval to these sample points using the ellipse as initial solution.

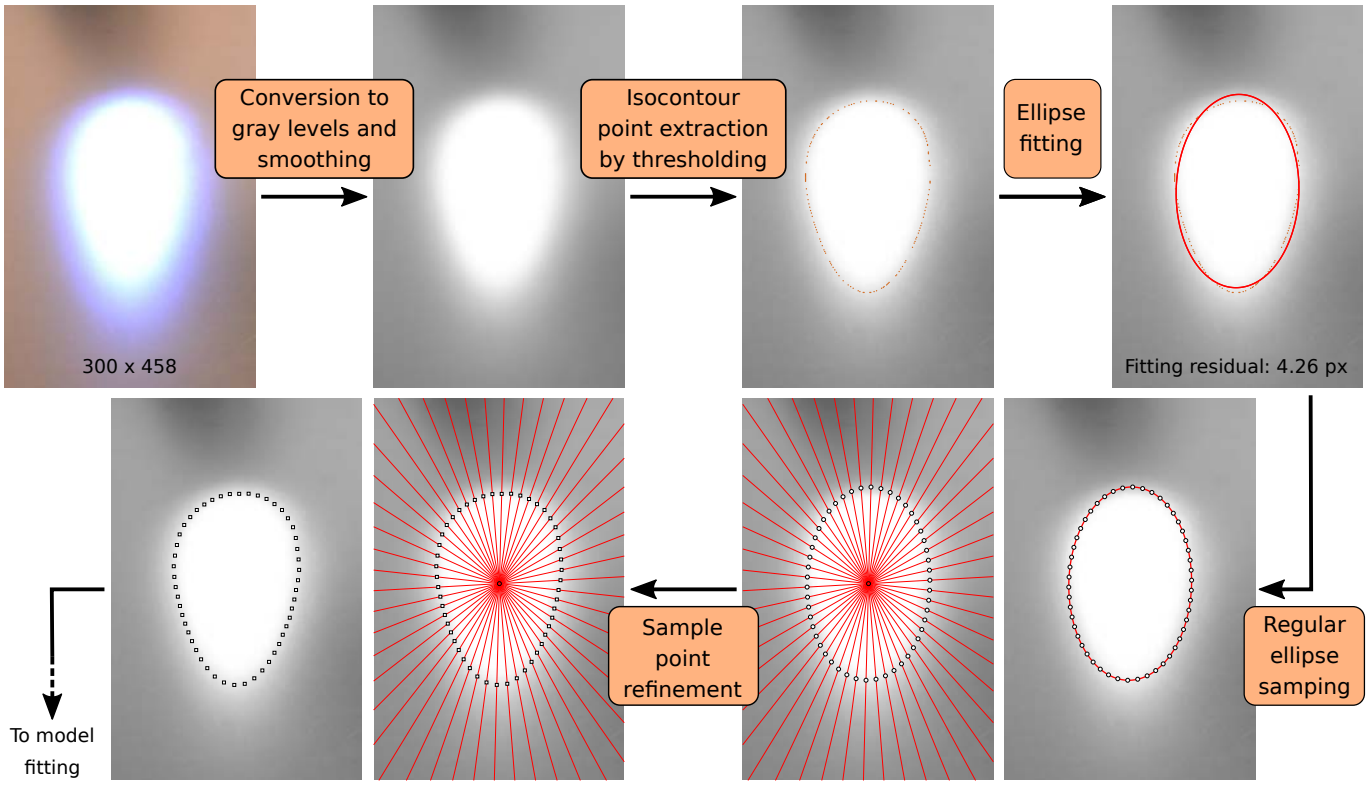


Figure 9: **Sample point extraction**. The procedure is illustrated on an image of a metal shelf taken in the regular regime as shown in figure 8. The model fitting result is shown in figure 2.

7.3.1 Sample Point Extraction

Our goal is to find m sample points (x_j, y_j), $j = 1, \ldots, m$, as close to the isocontour as possible and as well spread on the isocontour as possible. The process is illustrated in figures 9 and 10. We first convert the RGB image to gray levels and smooth it with a Gaussian filter with a standard deviation of 0.1% of the image diagonal. We then detect pixels whose value is close to t' with a tolerance of 0.1%. This gives a set of points representing the isocontour irregularly, potentially including parts suffering from redundancies or scarcities. Nonetheless, we can use them to fit a putative ellipse, as described in the next section, which forms the basis for a regular sampling of the isocontour, as follows. We start by sampling m putative points regularly on the putative ellipse. This is done by a regular angle-based sampling from the putative ellipse centre followed by

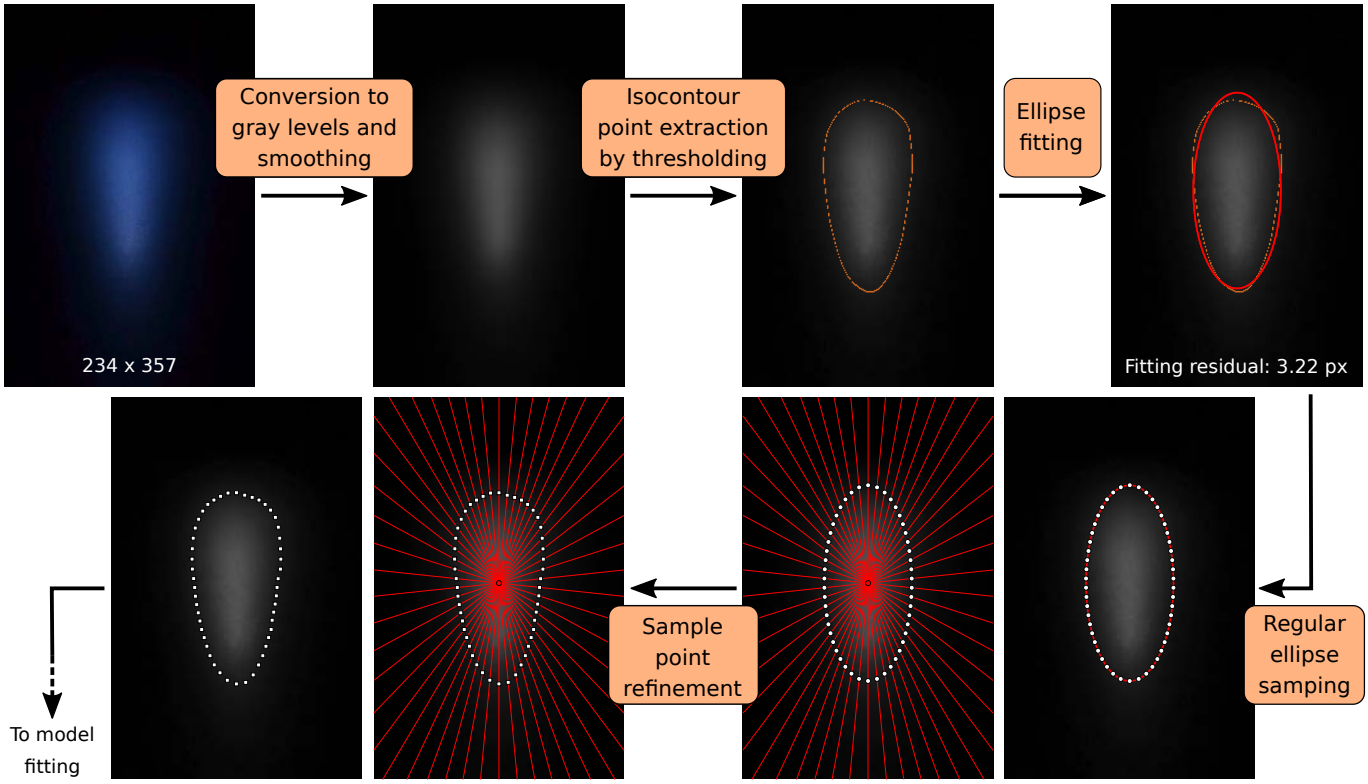


Figure 10: **Sample point extraction**. The procedure is illustrated on an image of a metal shelf taken in the fast-shutter regime as shown in figure 8. The model fitting result is shown in figure 3.

a nonlinear minimization of the variance of the distance between consecutive sample points. We then infer each final sample point by refining a putative sample point. This is done by searching along the line joining the putative point to the ellipse centre for the point whose gray level is closest to t'. The image is evaluated at non-integer coordinates using bilinear interpolation. This procedure guarantees that the sample points are well representative of the isocontour.

7.3.2 Ellipse Fitting

We follow a standard procedure to fit an ellipse optimally to m sample points. The ellipse is represented by the six homogeneous coefficients $e_1, \ldots, e_6 \in \mathbb{R}$ forming the symmetric matrix $E \in \mathbb{R}^{3\times 3}$ with unit norm. We minimize the sum of squared geometric distances between the sample points and the ellipse. This is done by introducing latent variables (x'_j, y'_j), $j = 1, \ldots, m$, representing the closest point to each sample point on the ellipse and solving:

$$\min_{\substack{e_1,\ldots,e_6 \\ (x'_j,y'_j),\ j=1,\ldots,m}} \sum_{j=1}^{m} (x'_j - x_j)^2 + (y'_j - y_j)^2 \qquad \text{s.t.} \qquad \sum_{i=1}^{6} e_i^2 = 1,\ H_E(x'_j, y'_j) = 0,\ j = 1, \ldots, m.$$

The minimization is conducted with the interior point method implemented in Matlab's `fmincon` function. The initialization is provided by (Fitzgibbon et al., 1999).

7.3.3 Algebraic Highlight Oval Fitting

We follow a similar procedure as for ellipse fitting. The algebraic highlight oval is represented by the five homogeneous coefficients $\beta_1, \ldots, \beta_5 \in \mathbb{R}$ and the three translation and rotation parameters $t_x, t_y, \theta \in \mathbb{R}$. We force the five homogeneous coefficients to have unit norm. Introducing the latent variables (x'_j, y'_j),

$j = 1, \ldots, m$, we have:

$$\min_{\substack{\beta_1,\ldots,\beta_5,t_x,t_y,\theta \\ (x'_j,y'_j),\ j=1,\ldots,m}} \sum_{j=1}^{m} (x'_j - x_j)^2 + (y'_j - y_j)^2 \qquad \text{s.t.} \qquad \sum_{i=1}^{5} \beta_i^2 = 1,\ h'(x'_j, y'_j) = 0,\ j = 1, \ldots, m.$$

The minimization is conducted similarly to the ellipse case using Matlab's `fmincon` function. The initialization is provided by the optimal ellipse. Concretely, we set t_x, t_y, θ from the ellipse's position and orientation, $\beta_1 = \beta_4 = 0$ and $\beta_2, \beta_3, \beta_5$ from the ellipse's semi-minor and semi-major axes, so that the initial algebraic highlight oval replicates the optimal ellipse exactly.

An alternative fitting procedure would be to use the so-called algebraic distance between the algebraic highlight oval and the sample points as a minimization criterion. The advantage is that no latent variables would be required:

$$\min_{\beta_1,\ldots,\beta_5,t_x,t_y,\theta} \sum_{j=1}^{m} h'(x_j, y_j)^2 \qquad \text{s.t.} \qquad \sum_{i=1}^{5} \beta_i^2 = 1.$$

The disadvantage is that this criterion has no direct physical or statistical interpretation. Our experiments showed that it was very unstable even in simple cases.

7.4 Results

The fitting residuals are measured in pixels (px). The original images all have the same size but the observed specular highlights do not, which causes discrepancies in the fitting residuals. We cope with these by using relative residuals with respect to the ellipse square root area, expressed in percents. More precisely, for e_p the fitting residual in pixels, we use the refined ellipse's area `area` to obtain the relative fitting residual as $e_r \stackrel{\text{def}}{=} 100\, e_p / \sqrt{\texttt{area}}$. The results in average and standard deviation are given in the following table:

	shelf		box		book	
	ellipse	oval	ellipse	oval	ellipse	oval
regular	$1.35\% \pm 0.77$	$0.61\% \pm 0.19$	$1.72\% \pm 0.78$	$1.22\% \pm 0.45$	$2.08\% \pm 1.24$	$1.42\% \pm 0.88$
fast-shutter	$1.91\% \pm 0.83$	$0.49\% \pm 0.14$	$1.69\% \pm 0.59$	$1.05\% \pm 0.47$	$1.50\% \pm 0.59$	$1.11\% \pm 0.44$

We observe that the algebraic highlight oval always has a lower fitting residual than the ellipse. The difference is particularly striking on the metal shelf in fast-shutter mode, which is probably where the images are the closest to the theoretical model. The difference is less important on the box and the book. This is explained by the fact that, though the algebraic highlight oval is a better explanation of the isocontour, the residual is dominated by measurement noise, as opposed to modeling error. The example from the plastic box in regular regime given in figure 11 illustrates this point.

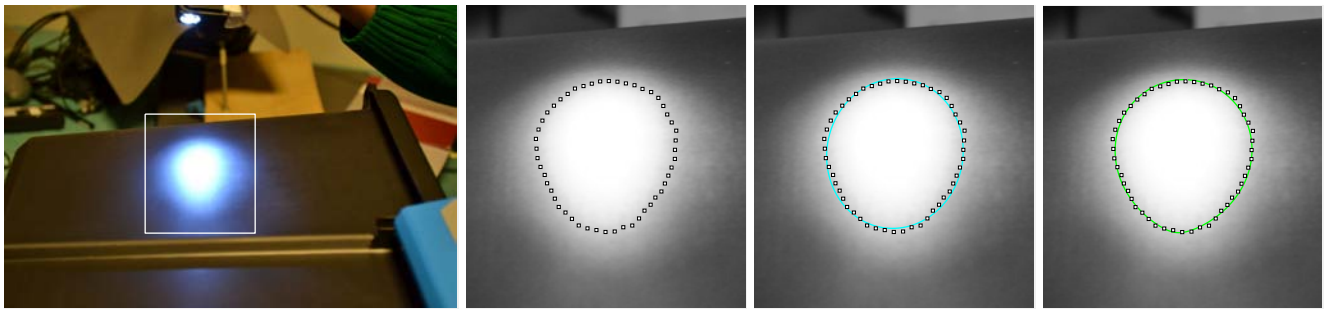


Figure 11: **Examples from the plastic box dataset in the regular regime.** The fitting residual is 2.55 px and 1.77 px for the ellipse and algebraic highlight oval respectively. This small difference indicates that, despite the perceptually strong asymmetry of the shape, measurement noise dominates the residuals.

7.5 Other Conditions

We have tested our fitting procedure in various other conditions.

Curved surfaces. Specular Highlights on curved surfaces tend to be smaller than on planar surfaces, because the surface normal changes more rapidly. Consequently, the surface supporting the specular highlight tends to be locally planar. Larger specular highlights on curved surfaces are created by rules surfaces, where the bending occurs mostly along one direction. It causes the specular highlight to shrink along this direction but preserves it orthogonally. An example of a ruled surface is the page of a book. Two examples are shown in figures 12 and 13, in the regular and fast-shutter regimes respectively. The first example, in regular regime, shows a specular highlight on a mildly curved part of the page, whose shrinking is thus very limited. We observe that the specular highlight has a strongly oval shape, which is not well fitted by the ellipse but very well by the algebraic highlight oval, as can be seen from the fitting residuals. The second example, in fast-shutter regime, shows a specular highlight on a highly curved part of the page, whose shrinking is thus strong. We observe that the specular highlight has an elliptical shape and is very well fitted by both the ellipse and algebraic highlight oval models.

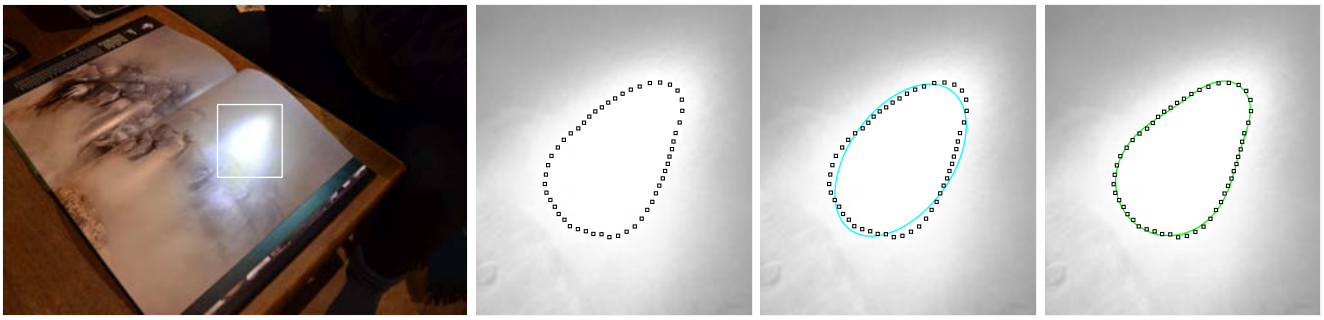


Figure 12: **Example with a curved glossy book in the regular regime.** The fitting residual is 2.89 px and 0.84 px for the ellipse and algebraic highlight oval respectively.

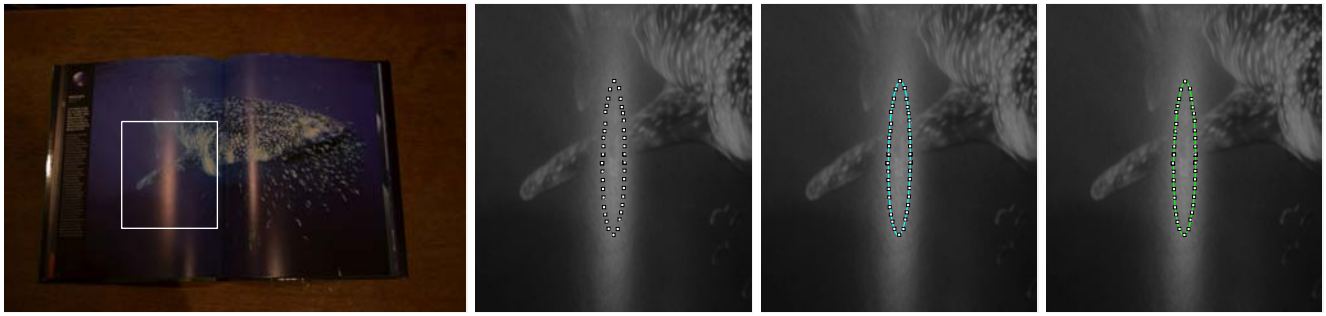


Figure 13: **Example with a curved glossy book in the fast-shutter regime.** The fitting residual is 0.34 px and 0.25 px for the ellipse and algebraic highlight oval respectively.

Smartphone camera. We used a smartphone to take images of the metal shelf in the same conditions as in our main setup. The conditions are equivalent to the regular regime as the shutter speed cannot be set on this camera. The image size is 1280×720 pixels. We observe in figure 14 that the specular highlight has a clear oval shape. It is much better modeled by the algebraic highlight oval than by the ellipse, as reflected by the fitting residuals.

Outdoor conditions. We used the same smartphone outdoor to photograph a car at night. The light was provided by streetlight in front of the car's hood. We observe in figure 15 that the specular highlight

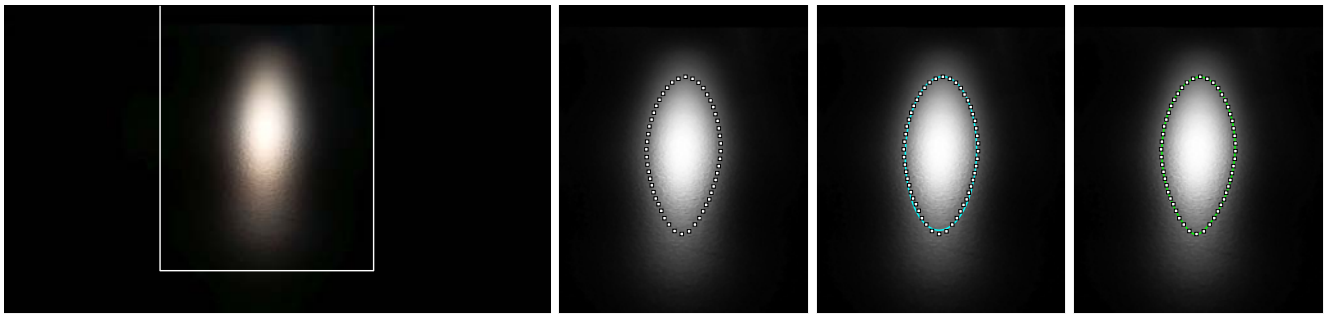


Figure 14: **Example with a metal shelf taken with a smartphone camera.** The fitting residual is 2.75 px and 0.67 px for the ellipse and algebraic highlight oval respectively.

has a clear oval shape. It is better modeled by the algebraic highlight oval than by the ellipse, as reflected by the fitting residuals.

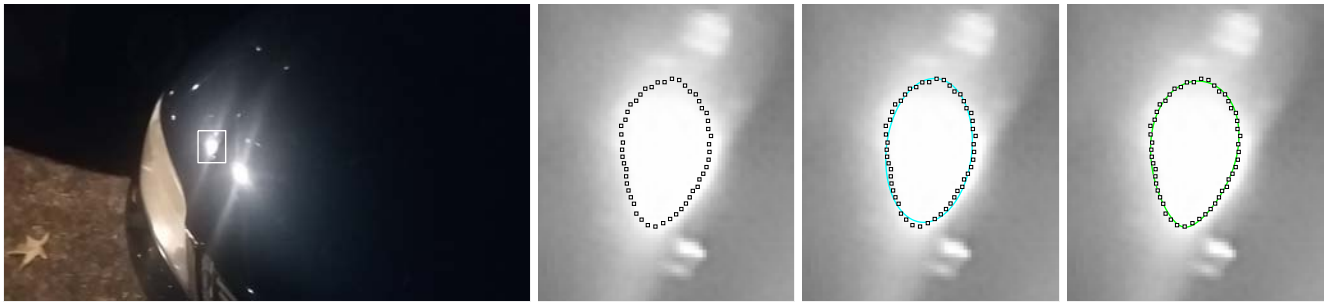


Figure 15: **Example with a car under streetlight taken at night with a smartphone camera.** The fitting residual is 0.65 px and 0.37 px for the ellipse and algebraic highlight oval respectively.

Low resolution. We took an image under the previous outdoor conditions but reducing the sensor sensitivity to ISO 100. This causes the image to darken and the specular highlights to shrink significantly. The resolution at which a specular highlight is observed is thus severely lowered. In the example shown in figure 16, the specular highlight's bounding box is 33×45 pixels. We observe that the sample points are very noisy but nonetheless reflect an oval shape for the specular highlight. This shape is then better modeled by the algebraic highlight oval than by the ellipse, as reflected by the fitting residuals.

7.6 Failure Cases

We present two failure cases due to modeling errors for our fitting procedure.

Overlapping specular highlights. This first failure case is due to the merging of two specular highlights caused by two distinct light sources. The image and resulting isocontour points are shown in figure 17. We observe that the shape of the merged specular highlights does not resemble an ellipse or an oval. While the ellipse produces a coarse fit of the isocontour points, this causes the algebraic highlight oval to fit the isocontour points with its outer rather than its inner closed curve.

Complex surface geometry. This second failure case is due to a complex surface geometry forming a locally concave shape. Because the main light source is the sun in this example, this causes strong interreflections (Nayar et al., 1991) and results in a complex specular highlight shape, as observed in figure 18. Again, we observe that the ellipse produces a coarse fit of the isocontour points. The algebraic highlight oval fits the isocontour points with a very similar fitting residual, indicating that the model's specific degrees of

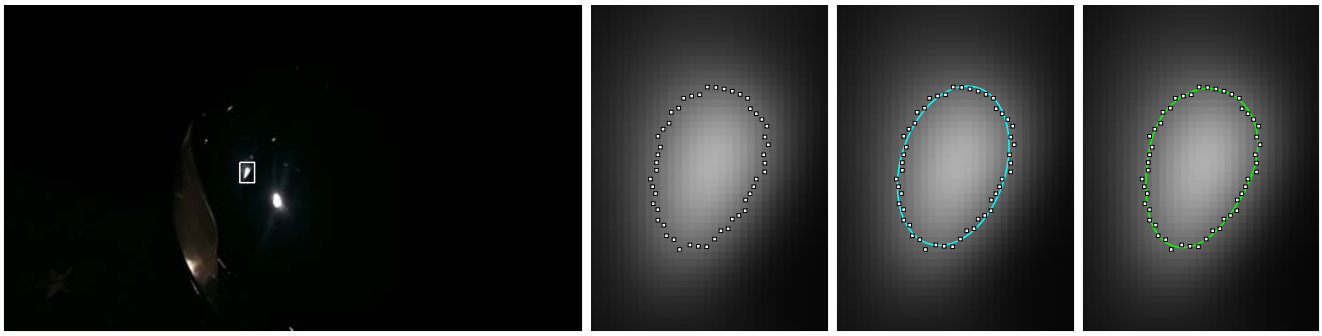


Figure 16: **Example with a car under streetlight taken at night with a smartphone camera in low-iso mode.** The fitting residual is 0.39 px and 0.28 px for the ellipse and algebraic highlight oval respectively.

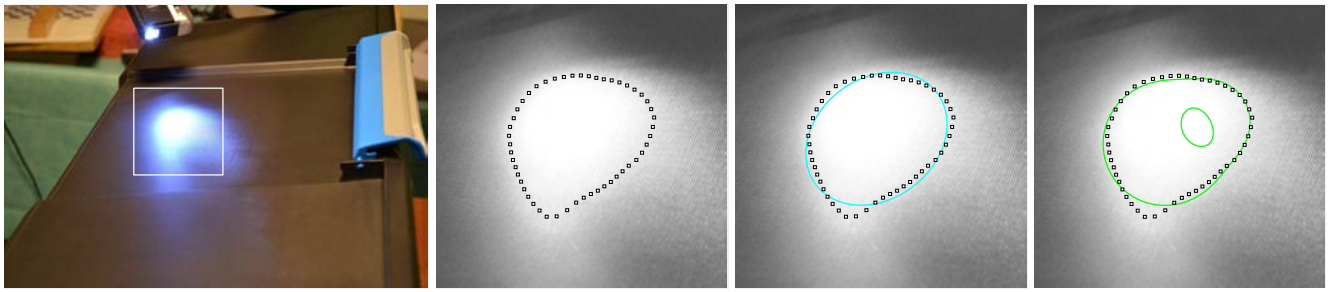


Figure 17: **Example with a plastic box under two flashlights.** The fitting residual is 4.18 px and 3.72 px for the ellipse and algebraic highlight oval respectively.

freedom are not well constrained by the data. We observe that the outer oval comes very close to the inner one, suggesting that the model approaches a singular configuration where they would merge into a single component.

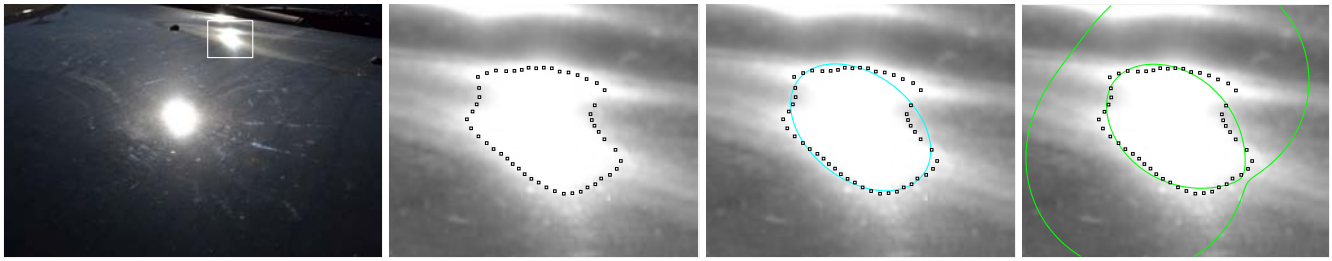


Figure 18: **Example with a car hood with complex geometry under sun light.** The fitting residual is 2.80 px and 2.70 px for the ellipse and algebraic highlight oval respectively.

8 Conclusion

We have proposed a novel model of ovals based on a bivariate quartic. We have thoroughly characterized the topology and convexity of this quartic. Interestingly, we have shown that a subset of these ovals form a pair of curves matching the isocontours of Phong's specular highlight model. Specular highlights are created by lighting a glossy surface with a point light source. In this respect, it would be interesting in future work to study oval models inspired by other reflection models. One possibility is Blinn-Phong's model, which we have found to lead to an octic. The proposed pseudo-3D and cross-section space framework can also

be applied to any bicircular quartic with symmetry about the x axis, such as the Cassinian and Cartesian ovals.

A potential application of the highlight ovals is diminished reality, where one deletes real objects from an image. In this application, it is very important for visual realism that the specular highlights be synthesized in the deletion zone. Because the illumination and reflectance parameters are generally unknown, this has to be done by propagating the specular highlight's visible part (Herling and Broll, 2014; Kawai et al., 2016). It was recently showed that taking the specular highlight's specific shape properties into account improved the results of propagation (Hadj Said et al., 2018). This was obtained using an empirical approximation of the specular highlight's isocontours by ellipses. The key idea of the method is to fit ellipses to pixels lying in the specular highlight's visible part and grouped by isocontour, and to constrain the filling-in of the deletion zone with these ellipses. Our result that the isocontours can be modeled by quartics may further improve this propagation method, by swapping the ellipse by the algebraic highlight oval. The algebraic highlight oval may also be useful in augmented reality, more precisely in surface retexturing. In this application, the texture of an object observed in an image is replaced by a user-supplied texture. Propagating specular highlights in the retextured area is crucial to ensure visual realism.

Acknowledgements. This research has received funding from the EU's FP7 through the ERC research grant 307483 FLEXABLE. We thank François Chadebecq for his kind help in taking the images for the experiments.

A The Height Space

We define the height space, which takes its name from the fact that its two dimensions are f and g, the heights of the highlight oval's 3D foci. The height space is essentially useful to prove a result on the impossibility of some configurations of H_E with respect to H_A in the cross-section space for $1 > |t| > 0$ and $a \neq 0$ in order to complete the proof of lemma 19. The above conditions imply from proposition 7 that H_E is an ellipse. The height space is illustrated in figure 19.

Definition 4 (The height space)**.** *We define the height space as the space $\mathbb{R}^2$ of the f and g parameters of the algebraic highlight oval in standard form (6).*

We start by expressing in the height space the variables r, d_l and d_m defined in the cross-section space. We recall that r is defined by lemma 16 for $1 > |t| > 0$ and $a \neq 0$ and reveals the position of the centre of the ellipse H_E with respect to the parabola H_A. We recall that d_l and d_m are defined by lemma 18 for $1 > |t| > 0$ and $\text{rank}(B) = 2$ and enter in the counting formula for the number of intersection points between H_E and H_A.

Lemma 31 (Representation of r in the height space)**.** *In the height space we have $r > 0$ in a band of half-width $u \stackrel{\text{def}}{=} \frac{a\sqrt{2}\sqrt{1+t^2}}{\sqrt{1-t^2}}$ centred on the line $f = -g$, $r = 0$ on the two lines defining the band, and $r < 0$ otherwise.*

Proof. We rewrite $r = 0$ from lemma 16 as:

$$(f+g)^2 = \frac{4a^2\left(1+t^2\right)}{1-t^2},$$

and then as:

$$f+g = \pm\frac{2a\sqrt{1+t^2}}{\sqrt{1-t^2}}.$$

This represents two lines parallel to $f = -g$, and at the same distance u of the origin of the height space. Finally, we have that $r = 4a^2\left(t^2+1\right) > 0$ at the origin, implying from lemma 16 that the centre of H_E is inside H_A in the band defined by the two above-derived parallel lines. □

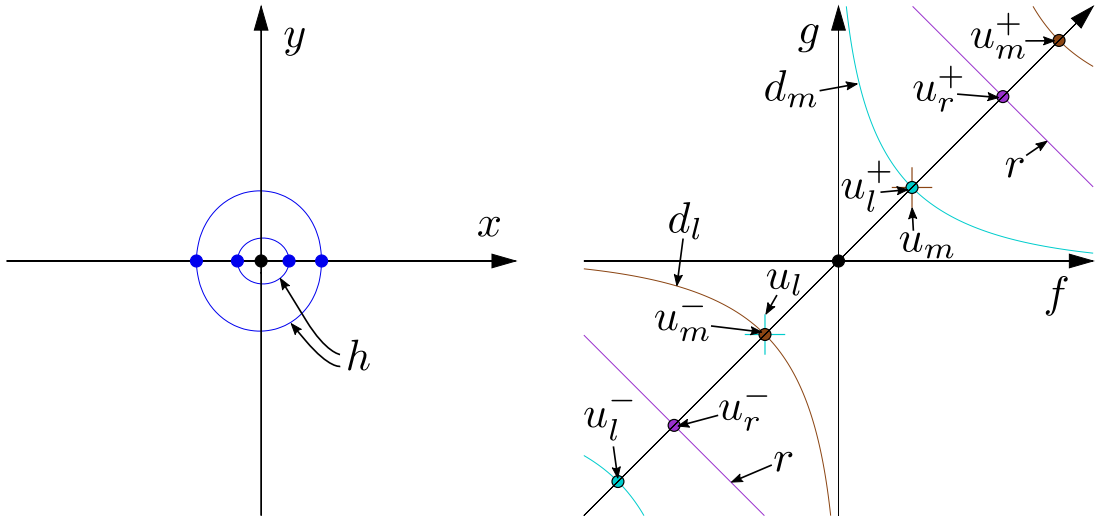


Figure 19: **The height space.** The numerical example is the algebraic highlight oval shown left (also top left in figure 1 and in figures 4 and 5) in the $\mathcal{S}$ plane. We have $1 > |t| > 0$ and proposition 7 tells us that the cylinder's cross-section H_E is an ellipse. In this example the ellipse has four intersections with the canonical parabola H_A. We can see that a wide range of values for f and g preserves the four intersection properties. The diagonal height space is indicated by the arrow showing the main diagonal.

Lemma 32 (Representation of d_l in the height space)**.** *In the height space we have $d_l < 0$ within a hyperbola of centre $\frac{at}{\sqrt{1-t^2}}[1\,1]^\top$, axis $f = g$ and semi-major axis $\frac{a\sqrt{2}}{\sqrt{1-t^2}}$. The hyperbola has a 'positive' and a 'negative' vertex at $\frac{a}{\sqrt{1-t^2}}(t \pm 1)[1\,1]^\top$.*

Proof. We rewrite $d_l = 0$ from lemma 18 as:

$$H_{Q_l}(f,g) \;=\; 0 \qquad \text{with} \qquad Q_l \;\stackrel{\text{def}}{=}\; \begin{bmatrix} t^2 & t^2-2 & 2at\sqrt{1-t^2} \\ t^2-2 & t^2 & 2at\sqrt{1-t^2} \\ 2at\sqrt{1-t^2} & 2at\sqrt{1-t^2} & -4a^2\left(t^2-1\right) \end{bmatrix}.$$

We have $\det(Q_l) = 16a^2\left(t^2-1\right) < 0$ and $\det\left(\bar{Q}_l\right) = 4\left(t^2-1\right) < 0$. Therefore, $d_l = 0$ is a hyperbola in the height space. The hyperbola's centre is:

$$q_l \;\stackrel{\text{def}}{=}\; \frac{at}{\sqrt{1-t^2}}[1\,1]^\top.$$

The value of $H_{Q_l}(q_l) = 4a^2$ is positive. The axis of the hyperbola is $f = g$. Its semi-major axis is $\frac{a\sqrt{2}}{\sqrt{1-t^2}}$. □

Lemma 33 (Representation of d_m in the height space)**.** *In the height space we have $d_m < 0$ within a hyperbola of centre $-\frac{at}{\sqrt{1-t^2}}[1\,1]^\top$, axis $f = g$ and semi-major axis $\frac{a\sqrt{2}}{\sqrt{1-t^2}}$. The hyperbola has a 'positive' and a 'negative' vertex at $\frac{a}{\sqrt{1-t^2}}(-t \pm 1)[1\,1]^\top$.*

Proof. We rewrite $d_m = 0$ as:

$$H_{Q_m}(f,g) \;=\; 0 \qquad \text{with} \qquad Q_m \;\stackrel{\text{def}}{=}\; \begin{bmatrix} t^2 & t^2-2 & -2at\sqrt{1-t^2} \\ t^2-2 & t^2 & -2at\sqrt{1-t^2} \\ -2at\sqrt{1-t^2} & -2at\sqrt{1-t^2} & -4a^2\left(t^2-1\right) \end{bmatrix}.$$

We have $\det(Q_m) = 16a^2\left(t^2-1\right) < 0$ and $\det\left(\bar{Q}_m\right) = 4\left(t^2-1\right) < 0$. Therefore, $d_m = 0$ is a hyperbola in the height space. The hyperbola's centre is:

$$q_m \;\stackrel{\text{def}}{=}\; -\frac{at}{\sqrt{1-t^2}}[1\,1]^\top.$$

The value of $H_{Q_m}(q_m) = 4a^2$ is positive. The axis of the hyperbola is $f = g$. Its semi-major axis is $\frac{a\sqrt{2}}{\sqrt{1-t^2}}$. □

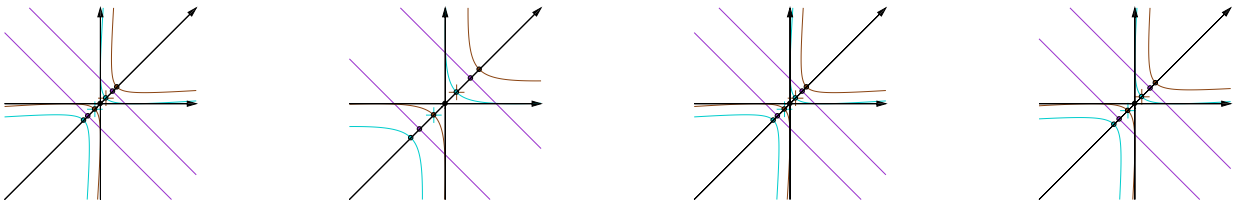
We now define the diagonal height space as a 1D restriction of the height space, and show how r, d_l and d_m are represented in this space.

Definition 5 (The diagonal height space)**.** *We define the diagonal height space as the line $f = g$. We use a single-coordinate system centred at the origin of the height space. We have from lemmas 32 and 33 that the line $f = g$ is the axis of the hyperbolas representing $d_l = 0$ and $d_m = 0$ and from lemma 31 that the line $f = g$ is orthogonal to the band representing $r > 0$.*

Lemma 34 (Representation and position of $r = 0$, $d_l = 0$ and $d_m = 0$ in the diagonal height space)**.** *We define $s \stackrel{\text{def}}{=} \frac{a\sqrt{2}}{\sqrt{1-t^2}}$. In the diagonal height space the points with $r = 0$ are located at $u_r^+ \stackrel{\text{def}}{=} s\sqrt{1+t^2}$ and $u_r^- \stackrel{\text{def}}{=} -s\sqrt{1+t^2}$. The centre and vertices of the hyperbola $d_l = 0$ lie at $u_m \stackrel{\text{def}}{=} st$, $u_m^+ \stackrel{\text{def}}{=} s(t+1)$ and $u_m^- \stackrel{\text{def}}{=} s(t-1)$. The centre and vertices of the hyperbola $d_m = 0$ lie at $u_l \stackrel{\text{def}}{=} -st$, $u_l^+ \stackrel{\text{def}}{=} -s(t-1)$ and $u_l^- \stackrel{\text{def}}{=} -s(t+1)$. Both hyperbolas' semi-major axis is equal to s.*

Proof. The intersection points between the diagonal height space and the two lines defining $r = 0$ are located at a distance u defined by lemma 31 from the origin, and we have $u_r^+ = -u_r^- = u$. The centre and vertices of the hyperbola $d_l = 0$ lie in the diagonal height space because, as given by lemma 32, they are located on the line $f = g$. We immediately see that the distance of the centre from the origin is $\left\| \frac{at}{\sqrt{1-t^2}} [1\,1]^\top \right\| = s|t|$ and thus its coordinate is simply $u_m = st$. The same reasoning applies to the vertices, and then to the hyperbola $d_m = 0$, following lemma 33. □

We finally return to the height space to give the following lemma, which forms the result we need to conclude our proof of lemma 19. Figure 20 illustrates the proof with four cases.



(a) – $n = 0$ (b) – $n = 1$ (c) – $n = 2$ (d) – $n = 3$

Figure 20: **The height space for** $1 > |t| > 0$**.** We observe that d_l and d_m are hyperbolas which always have one vertex within and one vertex off the $r > 0$ band. It means that for any value of t, $1 > |t| > 0$, and a one cannot find f and g such that the ellipse's centre is in the canonical parabola and they have zero, one or two contact points. This means that the ellipse H_E cannot be entirely within the canonical parabola H_A.

Lemma 35 (Position of the vertices of $d_l = 0$ and $d_m = 0$ with respect to $r > 0$ in the height space)**.** *The hyperbola $d_l = 0$ always has one vertex in the band $r > 0$ and one vertex outside. The same statement holds for the hyperbola $d_m = 0$. The vertices in the band $r > 0$ for the two hyperbolas are always on different sides.*

Proof. We always have $u_l^+, u_m^+ > 0$ and $u_l^-, u_m^- < 0$. For the case $t > 0$ we can easily verify that $u_l^+ > u_r^+$, $u_l^- > u_r^-$, $u_m^+ < u_r^+$ and $u_m^- < u_r^-$. In other words, the positive vertex of $d_l = 0$ is outside the band and its negative vertex is inside, while $d_m = 0$ is in the opposite configuration. For the case $t < 0$ we can easily verify that the opposite configuration holds for $d_l = 0$ and $d_m = 0$. □

Proof of lemma 19, $a \neq 0$. Recall that $d_l \leq 0$ occurs inside the hyperbola, meaning for $u \geq u_l^+$ or $u \leq u_l^-$, and similarly for $d_m \leq 0$. According to lemma 35 we cannot find a point in the diagonal height space for which ($u \geq u_l^+$ or $u \leq u_l^-$) and ($u \geq u_m^+$ or $u \leq u_m^-$) and $u_r^- \leq u \leq u_r^+$. Therefore, we cannot find a point in the height space for which $d_l \leq 0$, $d_m \leq 0$ and $r > 0$. This means that we cannot not find an ellipse H_E which lies inside H_A and has only contact points with H_A. □

References

A. B. Basset. *An Elementary Treatise on Cubic and Quartic Curves*. Cambridge University Press, 1901.

A. Blake and H. Bülthoff. Does the brain know the physics of specular reflection? *Nature*, 343(6254): 165–168, January 1990.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

M. Breuss and Y. C. Ju. Shape from shading with specular highlights: Analysis of the Phong model. In *International Conference on Image Processing*, 2011.

R. A. Dixon. *Mathographics*. Dover Publications, Inc, 1987.

A. Fitzgibbon, M. Pilu, and R. Fisher. Direct least square fitting of ellipses. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, 1999.

D. Forsyth and J. Ponce. *Computer Vision – A Modern Approach*. Prentice Hall, 2003. Second Edition.

S. Hadj Said, M. Tamaazousti, and A. Bartoli. Image-based models for specularity propagation in diminished reality. IEEE *Transactions on Visualization and Computer Graphics*, 24(7):2140–2152, July 2018.

J. W. Helton and V. Vinnikov. Linear matrix inequality representation of sets. *Communications on Pure and Applied Mathematics*, 60(5):654–674, May 2007.

J. Herling and W. Broll. High-quality real-time video inpainting with pixmix. IEEE *Transactions on Visualization and Computer Graphics*, 20(6):866–879, June 2014.

J. F. Hughes, A. van Dam, M. McGuire, D. F. Sklar, J. D. Foley, S. K. Feiner, and K. Akeley. *Computer Graphics: Principles and Practice*. Addison-Wesley, 2014. Third Edition.

N. Kawai, T. Sato, and N. Yokoya. Diminished reality based on image inpainting considering background geometry. IEEE *Transactions on Visualization and Computer Graphics*, 22(3):1236–1247, March 2016.

J. D. Lawrence. *A Catalog of Special Plane Curves*. Dover Books on Advanced Mathematics. Dover Publications, 1972.

J. S. Milne. *Elliptic Curves*. BookSurge Publishers, 2006.

S. K. Nayar, K. Ikeuchi, and T. Kanade. Shape from interreflections. *International Journal of Computer Vision*, 6(3):173–195, August 1991.

B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, June 1975.

J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1963.