Superconic and subconic surfaces in optical design

Alan W. Greynolds
Breault Research Organization, Inc., 6400 E. Grant Rd., Tucson, AZ 85715
520-721-0500, fax 520-721-9630, agreynolds@breault.com

Abstract: The superconic surface description has been around since 1986 and more recently implemented in commercial design software. A simpler version dubbed the ‘subconic’ is proposed and appears to work well in applications requiring steep aspherics.

The original summary of this document was printed in the 2002 International Optical Design Conference Technical Digest, p. 2.
© 2002 Optical Society of America
OCIS codes: (220.1250) Aspherics; (080.3620) Lens Design

Review of basic conic surfaces

The standard symmetric conic surface (i.e. optical conicoid) with vertex radius of curvature $R$ and classical conic constant $\kappa$ can be written in both implicit and explicit forms, respectively [1]:

\[
az^2 - 2z + cr^2 = 0 \iff z = \frac{cr^2}{1 + \sqrt{1 - acr^2}} \quad (1)
\]

The explicit version is just a convenient formula for the ‘sag’ of the surface. On the other hand, it is easier to calculate a ray intersection and normal vector with the implicit form. A sometimes useful alternative to the implicit form is:

\[
c\kappa z^2 - 2z + cs^2 = 0 \quad s^2 = x^2 + y^2 + z^2 \quad (2)
\]

Likewise, there is the power-series expansion of the explicit [2]:

\[
z = \frac{1}{2} cr^2 [1 + \frac{1}{3} acr^2 + \frac{1}{8}(acr^2)^2 + \frac{5}{64}(acr^2)^3 + \ldots] \quad |acr^2| < 1 \quad (3)
\]

Although any negative value of $\epsilon$ is allowed in the original formula, this series representation can easily diverge for very negative ones, i.e. cone-like hyperbolas.

A slightly more general optical conicoid

For the standard optical conicoid, the symmetry axis of the conic curve coincides with the rotational/optical axis of the surface. This is not the most general optical conicoid that has a well-defined axial curvature $c$ required in the design of most imaging systems. One more term proportional to $rz$ can be added to the quadratic implicit definition without forming a sharp tip on axis (i.e. no ‘axiconics’ which have a term linear in just $r$).

\[
az^2 - 2(1 + br)z + cr^2 = 0 \iff z = \frac{cr^2}{(1 + br) + \sqrt{(1 + br)^2 - acr^2}} \quad (4)
\]

Here the definition of the extra inverse-length parameter $b$ in terms of a dimensionless quantity $\beta$ is chosen so that the cases $c = 0$ and $\kappa = 0$ still refer to simple planes and spheres, respectively. However when $b \neq 0$, $\kappa$ can no longer be identified as the conic constant of the ‘tilted’ curve. The series expansion will of course be more complicated.

\[
z = \frac{1}{2} cr^2 [1 - br + (b^2 + \frac{1}{4}ac)r^2 - b(b^2 + \frac{3}{4}ac)r^3 + b^2(2b^2 + \frac{3}{2}ac)r^4 + \ldots] \quad (5)
\]

In fact, there are now odd powers in $r$ present. Classical aberration expansions of systems with these surfaces would still be valid within a certain range, but should not be expected to converge as rapidly as a whole new expansion that also admits odd powers (except linear) in the pupil radial coordinate.
**Rational parametric (Bezier) representation**

The fundamental building block of virtually all modern CAD systems is the NURBS (Non-Uniform Rational B-Spline) parametric curve or surface. All NURBS can be exactly represented by a piece-wise set of mathematically simpler rational Bezier curves or surfaces [3]. Also, a finite segment of any conic curve can be exactly represented by a rational quadratic parametric curve. A point \( P \) (in 2 or 3 dimensions) on this curve is given in terms of its parametric coordinate \( u \), a positive weight factor \( w \), and three fixed points \( P_0, P_1, P_2 \) in the plane of the conic curve.

\[
P = \frac{(1-u)^2P_0 + 2w(1-u)uP_1 + u^2P_2}{(1-u)^2 + 2w(1-u)u + u^2} \quad \left\{ \begin{array}{l} 0 \leq u \leq 1 \\
0 < w < 1 \quad \text{Ellipse} \\
w = 1 \quad \text{Parabola} \\
w > 1 \quad \text{Hyperbola} \end{array} \right. \tag{6}
\]

\( P_0 \) and \( P_2 \) are obviously the end points of the curve while the intermediate control point \( P_1 \) (though not on the curve) defines the tangents to the curve at these end points. For a maximum radial value (aperture semi-diameter) \( h \) and by first defining the following quantity,

\[
d = (1 + bh) + \sqrt{(1 + bh)^2 - ach^2} \tag{7}
\]

the rational Bezier coordinates for the more general optical conicoid of the last section can be simply expressed as follows:

<table>
<thead>
<tr>
<th>Point</th>
<th>( r )</th>
<th>( z )</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axial ( P_0 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Control ( P_1 )</td>
<td>( \frac{h}{d-bh} )</td>
<td>( \frac{d-bh}{\sqrt{2d}} )</td>
<td>( \frac{c_0}{d} )</td>
</tr>
<tr>
<td>Edge ( P_2 )</td>
<td>( \frac{h}{d} )</td>
<td>( \frac{c_0^2}{d^2} )</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. Rational quadratic Bezier curve points for the more general optical conicoid.

**The standard aspheric polynomial extension**

Although the conic alone is known to solve certain axial imaging problems exactly, for any conic constant it approaches a plane as the curvature goes to zero, and therefore can’t be used to describe a Schmidt-like corrector plate. To this end, the earliest extension added higher-order even-polynomial terms:

\[
a(z - Dr^4 - Er^6 - \ldots)^2 - 2(z - Dr^4 - Er^6 - \ldots) + cr^2 = 0 \iff z = \frac{cr^2}{1 + \sqrt{1 - acr^2}} + Dr^4 + Er^6 + \ldots \tag{8}
\]

Note the conspicuous absence of an additional quadratic term so that the vertex curvature and thus the paraxial properties are unchanged. For steep aspheric surfaces, it is not uncommon to require additional terms that go to 20th-order or higher and produce oscillatory behavior (as terms of opposite sign delicately balance each other).

**Brief introduction to the superconic surface**

The superconic surface was first implemented by the author in late 1986 using the new user-defined surface capability of CodeV version 7 [4]. The prime motivation was to produce steep aspheres with less terms and smoother correction. The starting points were the differential equation for one surface in an optical system that leads to perfect axial imaging [5] and the general Cartesian Oval solution for a single refractive surface [6]. Unfortunately, this meant giving up a closed-form explicit representation for the following implicit form:

\[
Az^2 - 2Bz + Cs^2 = 0 \quad \left\{ \begin{array}{l} B = 1 + b_1s^2 + b_2s^4 + \ldots \\
C = c + c_1s^2 + c_2s^4 + \ldots \\
c = R^{-1} \end{array} \right. \tag{9}
\]
Note that the ‘expansions’ are in terms of $s$ and not simply $r$ (the first approximation usually made in deriving the surface contribution to spherical aberration). It is rare that an application requires any more than the actual terms shown.

<table>
<thead>
<tr>
<th>Special Case</th>
<th>Non-zero Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane</td>
<td>None</td>
</tr>
<tr>
<td>Sphere</td>
<td>$c$</td>
</tr>
<tr>
<td>Conic</td>
<td>$c, A = ck$</td>
</tr>
<tr>
<td>Cartesian Oval</td>
<td>$A, b_1, c, c_1$</td>
</tr>
<tr>
<td>'Rational Polynomial'</td>
<td>All but $A$</td>
</tr>
</tbody>
</table>

Table 2. Subsets of the superconic surface.

**Plano-convex collimating lens example**

Figure 1 is a practical example of a successful superconic design where attempts to use the classical base conic plus higher-order even-polynomial surface failed [7].

![Superconic laser diode collimator](U.S. Patent 5,745,519)

Although the superconic was flexible enough to solve this problem to within the required accuracy, it is not the perfect surface for a thick plano-convex collimating lens. Let $d$ be the axial distance from the source point to the plane surface, $n$ the refractive index, $t$ the axial lens thickness, and $z$ the sag of the convex surface. Then using Fermat’s principle for an arbitrary ray with a starting angle $u$ from the axis:

$$\frac{d}{\cos u} + \frac{n(t - z)}{\cos u'} + z = d + nt$$

\[u' = \sin^{-1}(n^{-1}\sin u) \quad \text{(Snell’s law)}
\]

\[f = d + t/n \quad \text{(effective focal length)}
\]

The coordinates of a point on the perfect convex asphere as a function of the parameter $u$ are therefore:

$$z = \frac{d(\sec u - 1) + nt(\sec u' - 1)}{n \sec u' - 1}$$

$$r = d \tan u + (t - z) \tan u'$$

(10)
What’s in a name?

The name ‘superconic’ was actually coined by a colleague [8]. At the urgence of the same colleague and with
due credit, the superconic eventually became an intrinsic surface type in ZEMAX [9]. Much later, it was
learned that the same name is also used in the computer graphics field to refer to a much different curve
[10].

\[
|1 - az|^p + \epsilon|cr|^p = 1
\]

\[
\begin{align*}
& p > 1 \quad \text{not necessarily an integer} \\
& \epsilon > 0 \quad \text{SuperEllipse} \\
& \epsilon < 0 \quad \text{SuperHyperbola}
\end{align*}
\]  \quad (12)

Note that in the case of \( p = 2 \) this becomes a standard optical conicoid and only then are \( c \) and \( \epsilon \) the
ture vertex curvature and conic parameter, respectively. Future research could be directed at determining
whether this ‘alternative’ surface description (with \( p \) another variable) might be useful in optical design.

The subconic surface: less is more?

In a Schmidt telescope, the corrector plate preconditions the wavefront so that the spherical primary focusses
like a parabola would in the plate’s absence. Therefore, the shape of one surface of the plate is very nearly
the difference (times a scale factor) between a sphere and a parabola with the same vertex curvatures. This
suggests a slightly more general surface that subtracts a parabola from an arbitrary conic:

\[
a(z + \frac{1}{2}cr)^2 - 2(z + \frac{1}{2}cr)r + r^2 = 0 \iff z = \frac{cr}{1 + \sqrt{1 - acr^2}} - \frac{1}{2}cr^2 \quad \Delta c = c - c_0 = R^{-1}
\]  \quad (13)

This simple 3-parameter surface is appropriately called a subconic. Fortunately, most of the available design
codes already implement this as a subset of a ‘special’ surface (although not all properly take into account the
additional quadratic term when calculating paraxial properties). Figure 2 is an example of using a subconic
for the corrector plate of a Wright telescope. The tick marks on the axes of the spot size and ray fan plots
are located at the first zero of the ideal diffraction image. Therefore, the design is diffraction-limited on axis.

Fig. 2. Diffraction-limited correction of spherical aberration with a subconic corrector plate.
Representation as rational quartic Bezier

Any curve given by the sag of a base conic plus polynomial terms up to order N can be exactly represented by a rational Bezier curve of degree 2N. Therefore, the subconic is a rational quartic Bezier curve with two additional intermediate control points.

\[
P = \frac{t^4P_0 + 4w_1t^3uP_1 + 6w_2t^2u^2P_2 + 4w_3tu^3P_3 + u^4P_4}{t^4 + 4w_1t^3u + 6w_2t^2u^2 + 4w_3tu^3 + u^4}\]  \quad t = 1 - u \quad (14)

Given the three points \((r_0, z_0), (r_1, z_1), (r_2, z_2)\) of the base conic and its intermediate weight \(w\), then the five points and weights for the subconic are given by:

<table>
<thead>
<tr>
<th>Point</th>
<th>(r)</th>
<th>(z)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>(r_0)</td>
<td>(z_0 - \frac{1}{2}c_0r_0^2)</td>
<td>1</td>
</tr>
<tr>
<td>Control</td>
<td>(\frac{r_0 + r_1}{2})</td>
<td>(\frac{2z_0 + z_1 - c_0r_0r_1}{2})</td>
<td>(w)</td>
</tr>
<tr>
<td>Control</td>
<td>(\frac{r_0 + r_2 + vr_1}{2})</td>
<td>(\frac{z_0 + z_2 + vz_1 - c_0(r_0r_2 + 2v^2r_1^2)}{2 + v})</td>
<td>(\frac{2 + v}{6})</td>
</tr>
<tr>
<td>Control</td>
<td>(\frac{r_1 + r_2}{2})</td>
<td>(\frac{2z_1 + 2z_2 - c_0r_1r_2}{w})</td>
<td>(w)</td>
</tr>
<tr>
<td>End</td>
<td>(r_2)</td>
<td>(z_2 - \frac{1}{2}c_0r_2^2)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. Rational quartic Bezier curve points for the subconic \((v = 4w^2)\).

Solution to a simple interpolation problem

The subconic can be used to solve a useful interpolation problem. For a given vertex curvature \(\Delta c\) plus the sag \(z\) and slope \(dz/dr\) at a particular height \(r\), there exists a unique subconic:

\[
c = \frac{\Delta z \Delta z'}{(\frac{1}{2}r \Delta z' - 2 \Delta z)r}
\]

\[
\begin{cases} 
\Delta z = z - \frac{1}{2} \Delta c r^2 \\
\Delta z' = \frac{dz}{dr} - \Delta c r
\end{cases}
\]  \quad (15)

\[
c_0 = c - \Delta c \quad Z = z + \frac{1}{2}c_0r^2 \quad \epsilon = \frac{2}{cZ} - \left(\frac{r}{Z}\right)^2
\]

(16)

By numerically solving a coupled pair of differential equations for the shape of two surfaces in an optical system, all orders of spherical aberration and linear coma can be completely eliminated. A practical example of this technique is provided as a demo file with the freely downloadable version of OSLO [11]. Replacing the two ‘numerical’ surfaces with subconics determined by the above interpolation formulas suprisingly results in near diffraction-limited performance. Fine-tuning the subconics a small amount leads to the axial performance shown in Figure 3 while introducing only a slight amount of linear coma.

Conclusions: much ado about nothing

Optical designs requiring steep aspheres can benefit from ‘unusual’, though well-motivated, extensions to the classic conic that require only a few (typically one or two) additional parameters. But with today’s fast processors, globally optimizing software, and computer-controlled manufacture, are these simple analytical surface descriptions of any value when the ‘perfect’ aspheric for any application is just a large enough number of spline-interpolated points and slopes? Outside of the warm fuzzy feeling some of us get from being able to write down a simple equation for the surface, maybe one role that these variations on the conic can play is in the initial design passes. Eventually, they can be replaced by the equivalent ‘numeric’ spline surfaces which have significantly more degrees of freedom that can then be tweaked to get the last ounce of performance from the design.
3 mm diameter, diffraction-limited, .9 NA, near-IR, Polycarbonate lens

Fig. 3. OSLO LT demo lens (wwcd.len) with subconics substituted for 100-zone splines.

References

9. ZEMAX optical design program, Focus Software, Tucson AZ.