A B-Spline Representation of Optical Surfaces and its Accuracy in a Ray Trace Algorithm

Philipp Jester,¹,² Christoph Menke,¹ and Karsten Urban²

¹Corporate Research and Technology, Carl Zeiss AG, Carl-Zeiss-Str. 22, 73447 Oberkochen, Germany
²Institute for Numerical Mathematics and Scientific Computing Center Ulm (UZWR), University of Ulm, 89081 Ulm, Germany

We introduce a new representation of aspheric surfaces that is based on a B-spline quasi-interpolation scheme. The scheme is implemented in a ray trace algorithm and bounds on the approximation error are established. Examples for the reproduction of aspheric surfaces in polynomial description and the ray tracing accuracy are presented. The proposed approach allows the specification of local and global structures and the inclusion of measured surface data. The new representation gives access to a wavelet analysis, offering extended possibilities for the tolerance analysis of optical systems containing aspheric elements. © 2010 Optical Society of America

OCIS codes: 000.4430, 080.0080, 080.1753, 220.1250, 240.6700.

1. Introduction

Advances in fabrication technology, metrology and optical design led to a regular use of aspheric elements in optical systems. However, in some cases standard representations of aspheric surfaces turn out to be impractical. Freeform surfaces without symmetries that are frequently used in illumination systems may fall into this category. Another example is the description of an aspheric surface as it is built including manufacturing errors. Trying to represent such an as-built surface by polynomials inevitably leads to a high number of terms and makes ray tracing expensive and time-consuming. In addition, if local structures are present on the surface, a large approximation error is introduced despite of the high number of polynomial terms in a global representation of the surface. Consequently there is a need for alternative surface descriptions that combine high approximation accuracy with fast evaluation.
Several authors considered alternative aspheric representations. Greynolds [1] gives a brief review of the so-called ‘superconic’ and ‘subconic’ surface descriptions. The motivation for the introduction of these types was to produce steep aspheres with less terms and smoother correction. Forbes [2] proposed a sum of Jacobi polynomials to represent rotationally symmetric aspheres. These polynomials constitute an orthonormal basis of the unit circle. Forbes’ representation facilitates the enforcement of fabrication constraints, e.g. the deviation of the slope from a best-fit sphere. This allows the consideration of the cost of an asphere at an early design stage [3].

Lerner and Sasian [4,5] present approaches that use implicitly and parametrically defined surfaces for the design of imaging and illumination systems. Because these surface definitions are more general than the standard aspheric definition, they can better describe optical surfaces with large departures from a spherical or conic shape.

For non-rotationally symmetric surfaces, splines have been studied in the literature [6–8]. Parametric curves such as non-uniform rational B-splines (NURBS) have been applied to the design of rotationally symmetric aspheres [9] as well as to the design of freeform mirrors [10]. Because NURBS are a standard surface type in CAD software, they offer an attractive surface description when data has to be exchanged between optical design and CAD tools. Another optical surface representation using 2D Gaussians as radial basis functions has been introduced by Cakmakci [11].

Recently, Morita and co-workers [12] adopted the so-called Nagata patch for the representation of optical surfaces including form errors due to fabrication processes. The Nagata patch approximates surfaces through piecewise local quadratic interpolation by using position and normal vectors at the nodes of a triangular mesh. Its representation differs from usual descriptions of optical surfaces because it is continuous but not differentiable. The authors suggest that the Nagata patch may be used to simulate mid-spatial frequencies of surface errors.

A representation of optical surfaces in terms of B-splines seems promising as they are well adapted to represent both local and global structures and allow the inclusion of measured data into the surface description. In this paper, we propose a representation of optical surfaces via a B-spline quasi-interpolant. The surfaces are assumed to be non-diffractive and sufficiently smooth, i.e. twice differentiable. We have implemented the new scheme in our custom optical design software. Numerical results on the approximation of surfaces and on the accuracy in a ray-tracing algorithm are given. Finally, we point out the connection of the new surface representation to wavelets.
2. B-Spline Representation

2.A. B-Splines

We start by reviewing basic facts on B-splines that will be relevant in the sequel and refer to [13] for details. There are several equivalent definitions of cardinal B-splines $N_d$. We use the following recursive scheme

$$N_1(x) := \begin{cases} 
1, & \text{if } x \in [0, 1), \\
0, & \text{else},
\end{cases} \quad N_d(x) := \int_0^1 N_{d-1}(x-t) \, dt = (N_{d-1} \ast N_1)(x),$$

for $d \geq 2$ ($d$ is the order of the B-spline). The following properties are well-known and will frequently be used in the sequel

$$\text{supp } N_d \subset [0, d] \quad \text{ (locality)}, \quad (1)$$

$$N_d(x) = \frac{x}{d-1}N_{d-1}(x) + \frac{d-x}{d-1}N_{d-1}(x-1) \quad \text{ (recursion)}, \quad (2)$$

$$N_d \in C^{d-2}(\mathbb{R}) \text{ with } N'_d(x) = N_{d-1}(x) - N_{d-1}(x-1) \quad \text{ (regularity)}. \quad (3)$$

Note that (2) offers a fast recursive evaluation procedure since the evaluation is trivial for $N_1$. In addition (3) also allows a fast evaluation of derivatives. Another important property of $N_d$ is the following refinement equation

$$N_d(x) = 2^{1-d} \sum_{k=0}^d \binom{d}{k}N_d(2x-k), \quad x \in \mathbb{R}. \quad (4)$$

Hence, $N_d$ is also called (primal) scaling function. It is known that shifts of $N_d$, i.e. $N_d(\cdot - k)$, $k \in \mathbb{Z}$, are linearly independent. These last two properties mainly enable B-splines to generate a Multiresolution Analysis, which is essential for the access to wavelets, see [14,15] for details. Let us abbreviate by $\varphi := N_d$ and $S_j := \text{clos}_{L_2(\mathbb{R})} \{ \varphi_{j,k} : k \in \mathbb{Z} \}$. We call $j \in \mathbb{Z}$ the level, where for a piecewise continuous function $g : \mathbb{R} \to \mathbb{R}$, $g_{j,k}(x) := 2^{j/2}g(2^jx-k)$ denotes its scaled and shifted variant. Two-dimensional basis functions are constructed by the bivariate tensor product.

2.B. Projection

Let us now introduce the description of an optical surface in terms of B-splines. Of course, such an approach is by far not new [6]. However, we will use a more recent quasi-interpolation scheme in order to ensure the accuracy of the representation. Let $f : \Omega \to \mathbb{R}$, $f \in L_2(\Omega)$, be a smooth function, describing the (possibly unknown) surface. We use the following biorthogonal projection $P_j : L_2(\Omega) \to S_j$,

$$P_j f := \sum_k c_{j,k} \varphi_{j,k}, \quad c_{j,k} := (f, \varphi_{j,k})_{L_2(\Omega)},$$

$$x \in \mathbb{R}.$$
to approximate $f$, where $(u,v)_{L^2(\Omega)}$ denotes the inner product on $L^2(\Omega)$ with induced norm $\|u\|_{L^2(\Omega)}$ and $\tilde{\varphi}_{j,k}$ are the dual scaling functions. These functions satisfy a refinement equation similar to (4) and are dual in the sense $(\varphi_{j,k}, \tilde{\varphi}_{j,m})_{L^2(\Omega)} = \delta_{k,m}$, see [16] for details.

Now it holds, if $f$ is sufficiently smooth, i.e. $f \in H^s(\Omega)$, the following conclusion about the error is well-known [17]

$$\|f - P_j f\|_{L^2(\Omega)} = O(2^{-js}), \quad 0 \leq s \leq d. \quad (5)$$

We refer to [18] for the precise definition of the Sobolev space $H^s(\Omega)$.

Note however, that we are facing a problem for a real application. The coefficients $c_{j,k}$ involve an integral which normally cannot be computed exactly. Of course, one can resort to quadrature formulae which involve point values of $f$ and $\tilde{\varphi}_{j,k}$, but in most cases, point values of $\tilde{\varphi}_{j,k}$ cannot be computed exactly, see [17]. An alternative is offered by the following quasi-interpolation scheme.

### 2. C. Quasi-Interpolation

A quasi-interpolation scheme produces an approximation of a given function $f$ by only using point values of the function. This is similar to a classical interpolation. The difference is that a quasi-interpolation does not need to match point values of $f$ at given nodes. This allows error estimates in $L^2$ also with high orders of accuracy.

We use a quasi-interpolation scheme of the form

$$c_{j,k} \approx \bar{c}_{j,k} := 2^{-j/2} \sum_{\ell=-m}^{m} \gamma_{d,\ell} f (2^{-j} (k + \ell))$$

with $m := \left\lfloor \frac{d-1}{2} \right\rfloor$ and weights $\gamma_{d,\ell}$. We refer to [19] for the construction of the weights and their values. For the arising quasi-interpolant defined by

$$P^qi_j f := \sum_k \bar{c}_{j,k} \varphi_{j,k}$$

we obtain similar to Eq. (5), if $f \in H^s(\Omega)$,

$$\|f - P^qi_j f\|_{L^2(\Omega)} = O(2^{-js}), \quad 0 \leq s \leq d, \quad (6)$$

i.e., the same order of approximation as $P_j$.

### 3. Numerical Results

In this section, we describe numerical results for the approximation of optical surfaces in a ray trace algorithm. The described schemes have been implemented within FLENS and
LAWA\textsuperscript{1}. In order to test the performance within a realistic framework, we compare our results with the custom optical design software OASE\textsuperscript{2}. Further examples are described in [20].

3.4. Approximation of Optical Surfaces

We start by describing numerical experiments for approximating given optical surfaces. Fig. 1 shows contour-color plots of test surfaces we use in the sequel. The first surface, seen in part (a) of Fig. 1, is a KXY asphere given by the parameterization

\[ f_{\text{KXY}}(x, y) := \frac{\rho_x x^2 + \rho_y y^2}{1 + [1 - (1 + \kappa_x)(\rho_x x)^2 - (1 + \kappa_y)(\rho_y y)^2]^{1/2}} + c_1 x + c_2 y + c_3 x^2 + c_4 xy + c_5 y^2 + c_6 x^3 + \ldots, \]

while the others are rotationally symmetric KSA aspheres defined by

\[ f_{\text{KSA}}(h) := \frac{\rho h^2}{1 + [1 - (1 + \kappa)(\rho h)^2]^{1/2}} + \sum_{k=2}^{\infty} c_{k-1} h^{2k}, \quad \text{where} \quad h^2 := x^2 + y^2. \]

The bounding box is defined by \( \Omega = [-b, b] \times [-b, b] \), where \( b \) and the remaining parameters are listed in Tab. 1.

For fixed B-spline orders \( d \) we compute the approximation by the above quasi-interpolation method. We vary levels \( j = 4, \ldots, 10 \), where level \( j \) corresponds to \( 2^j \) coefficients in the two-dimensional representation and a mesh size \( h = 2^{-j} \). The root mean square value of the approximation error is computed on \( n \times n \) discrete points forming an equidistant mesh in \( \Omega \).

\[ E_j := \left( e(x_i, y_j) \right)_{i,j=1,\ldots,n}, \quad \text{with} \quad e := \left| f - P_j^{1} f \right|. \]

The following results are obtained with a fixed \( n = 10^4 \).

For B-spline orders \( d = 1, \ldots, 6 \), the errors \( \text{RMS}(E_j) \) of the first test surface KXY are drawn as markers in Fig. 2. Additionally, the plot shows best fit lines for every order \( d \) with expected slope \( 2^{-jd} \). Obviously the analytic rate of convergence from Eq. (6) is reached. Moreover one observes for high orders \( d \geq 5 \) an accuracy better than \( 10^{-12} \) with a coarse mesh size \( h \geq 1e^{-3} \).

The second example concerns the asphere KSA\textsubscript{1} shown in part (b) of Fig. 1. We choose this simple example in order to investigate the approximation order of the first derivative, where we expect to lose one order of approximation, see [17] for details. Hence, we compute \( E_j \) as above, where now

\[ e := \left| \frac{\partial}{\partial x} \left( f - P_j^{1} f \right) \right|. \]

\textsuperscript{1}Flexible Library for Efficient Numerical Solutions and Library for Adaptive Wavelet Applications, Institute for Numerical Mathematics, University Ulm.

\textsuperscript{2}Optische Analyse und Synthese, Carl Zeiss AG.
We obtain the corresponding error in Fig. 3. As mentioned, we lose one order of approximation for the convergence rates, so our best fit lines now have slope $2^{-j(d-1)}$. Once more, we observe the expected behavior and while we need a high approximation accuracy to the normal in a ray trace algorithm we recommend using splines with orders $d \geq 2$.

The last example in this section shows the accuracy for the surface KSA2 shown in part (c) of Fig. 1. We restrict the definition of the surface and the computation of the RMS-error to values in the domain $\Theta := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 20^2\}$, so we get undefined values in the bounding box. In $\Theta$ the surface is sufficiently smooth to meet our condition $f \in H^d(\Omega)$. The corresponding error plot is drawn in Fig. 4. As this asphere has a steep gradient near the boundary, more coefficients are needed to reach a sufficient accuracy. This becomes evident in the deviation of the markers from the slope for spline orders $d \geq 4$.

3.B. Ray Tracing, Intersection and Refraction

The following procedure for ray tracing is based on [7] and is adapted to the use of our B-spline representation. We define a ray $r : \mathbb{R} \rightarrow \mathbb{R}^3$ through a point $p = (p_x, p_y, p_z)^T \in \mathbb{R}^3$ in a direction $d = (d_x, d_y, d_z)^T \in \mathbb{R}^3$ as

$$r(\alpha) := p + \alpha d, \quad \alpha \in \mathbb{R},$$

and describe the refraction of such a ray by an optical surface given in terms of a function $g$ (which could also be an approximation e.g. determined by a quasi-interpolation). We have to find the intersection point $x = x(x, y) = (x, y, g(x, y))^T \in \mathbb{R}^3$ of the ray with the surface. The normal vector of the surface at $x$ is given as

$$n = n(x, y) = e_1 \times e_2 = (-\partial_x g(x, y), -\partial_y g(x, y), 1)^T,$$

where $e_1$ and $e_2$ are the two tangential vectors of the surface in $x$, i.e., $e_1 = \partial_x x(x, y) = (1, 0, \partial_x g)^T$ and $e_2 = \partial_y x = (0, 1, \partial_y g)^T$. In order to obtain a unique intersection point, we have to assume that the ray is not parallel to the tangential plane. Then, the condition $p + \alpha d = x$ for the 3 unknowns $\alpha$, $x$ and $y$ can be rephrased as

$$F(\alpha) := p_x + \alpha d_x - g(p_x + \alpha d_x, p_y + \alpha d_y) = 0, \quad F : \mathbb{R} \rightarrow \mathbb{R}.$$  

This is obviously a one-dimensional nonlinear equation which can be numerically solved e.g. by Newton’s method which requires the derivative

$$F'(\alpha) = d_x - \partial_x g(p_x + \alpha d_x, p_y + \alpha d_y)d_x - \partial_y g(p_x + \alpha d_x, p_y + \alpha d_y)d_y = n(x(\alpha), y(\alpha))^T \cdot d,$$

where $x(\alpha) := p_x + \alpha d_x$ and $y(\alpha) := p_y + \alpha d_y$.

The solution of $F(\alpha) = 0$ yields the intersection point $x$ and the normal vector $n$ at this point. Let us denote by $\gamma$ the incidence angle of the ray and the surface, i.e. $\cos(\gamma) = n^T \cdot d,$
and by \( n, n' \) the refractive indices of the media in front of and behind the surface. We apply the Law of Refraction, i.e. in \( \mathbb{R}^3 \),

\[
\cos(\gamma') = \left[ 1 - \left( \frac{n}{n'} \right)^2 (1 - \cos^2(\gamma)) \right]^\frac{1}{2},
\]

to obtain the new incidence angle \( \gamma' \). Hence, we can calculate the direction \( \mathbf{d}' \) of the ray in the medium behind the actual surface and proceed ray tracing with the next optical surface.

3.C. Accuracy in a Ray Trace Algorithm

We make use of the above ray trace method to validate our B-spline approximation in a realistic manner. For a given set of \( n \) rays \( \mathbf{r}_i \), we compute the intersection points \( \mathbf{x}_i^{(j)} \) and the normals \( \mathbf{n}_i^{(j)} \) depending on the level \( j \). By \( \mathbf{x}_i^{(O)} \) and \( \mathbf{n}_i^{(O)} \) we indicate the results computed by OASE for the standard representation of the surface.

For \( n = 10^6 \) rays, we report the maximal deviation of all rays

\[
M_{x,j} := \max_{i=1,\ldots,n} \left\{ \left\| \mathbf{x}_i^{(O)} - \mathbf{x}_i^{(j)} \right\| \right\}, \quad M_{n,j} := \max_{i=1,\ldots,n} \left\{ \left\| \mathbf{n}_i^{(O)} - \mathbf{n}_i^{(j)} \right\| \right\},
\]

where \( \| \cdot \| \) denotes the Euclidean norm in \( \mathbb{R}^3 \). The error plots for the surface KXY are given in Fig. 5 and Fig. 6. The first one shows the maximal deviation of the intersection point \( M_{x,j} \), while the latter one shows the error \( M_{n,j} \) of the normal vector. For the computation we use \( \varepsilon = 10^{-12} \) as the stopping criterion of Newton’s method.

We observe a similar behavior as for the RMS-error in Sec. 3.A. The rate of convergence for \( M_{x,j} \) is \( O(2^{-jd}) \). Since the normal is based on derivatives, it is not surprising that we obtain \( O(2^{-jd(d-1)}) \). For B-spline orders \( d \geq 4 \), the accuracy is sufficient to allow the analysis of imaging optical systems. We conclude that the B-spline quasi- interpolant scheme is usable in a ray trace algorithm and competes with other local representations such as the Nagata patch [12].

4. Connection to Wavelets

The localization property of compactly supported wavelets offers a unique advantage over Fourier methods for detecting local spatial structures. In signal analysis and image processing wavelet methods are well-established [15]. Wavelet transforms have also been used in the analysis of surface data [21]. Tien and Lyu introduced an inspection method based on the discrete wavelet transform [22].

The Fast Wavelet Transform

\[
\text{FWT} : (c) \mapsto (c_0, d_1, \ldots d_{j-1})
\]
maps the coefficients $c := (c_{j,k})_{(j,k)}$ to a \textit{multiscale representation}

$$P_j f = \sum_j \sum_k d_{j,k} \psi_{j,k},$$

where $\psi_{j,k}$ are wavelets. Thus, the B-spline quasi-interpolant representation can be used directly for a wavelet analysis. Some examples on error detection and frequency separation by wavelets are described in [16].

5. Conclusion and Outlook

A highly accurate approximation of optical surfaces in terms of a B-spline quasi-interpolation scheme has been introduced. The error bounds on the approximation have been verified through the reproduction of aspheric surfaces given by polynomial expansions. The new surface description has been implemented within our custom optical design software. Our results confirm that the desired accuracy of $10^{-12}$ is reached in the ray tracing. This enables the use of the new representation in the analysis of optical systems for imaging applications. Because B-spline-Wavelets offer the possibility to specify local and global structures simultaneously, they can be used to describe as-built optical surfaces including metrological data. This facilitates the tolerance analysis of optical systems containing aspheric elements.

Furthermore, the new scheme gives access to wavelet analysis methods for aspheric surfaces. The localization in space and time of compactly supported wavelets offers a unique advantage over Fourier methods for the detection of local spatial structures. This opens a broad field of applications in tolerancing and manufacturing and will be the focus of future investigations.

References

List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Contour-color plot of the test surfaces KXY (a), KSA₁ (b) and KSA₂ (c). jester-f1.eps.</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>RMS-error of the approximation of surface KXY. jester-f2.eps.</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>RMS-error of the approximation of the partial derivative $\partial_x P_j f_{KSA_1}$. jester-f3.eps.</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>RMS-error of the approximation of surface KSA₂ with undefined values in the bounding box. jester-f4.eps.</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>Maximal deviation $M_{x,j}$ of the intersection point. jester-f5.eps.</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>Maximal deviation of the normal $M_{n,j}$ of the intersection point. jester-f6.eps.</td>
<td>16</td>
</tr>
</tbody>
</table>
Fig. 1. Contour-color plot of the test surfaces $\text{KXY}$ (a), $\text{KSA}_1$ (b) and $\text{KSA}_2$ (c). jester-f1.eps.
Fig. 2. RMS-error of the approximation of surface KXY. jester-f2.eps.
Fig. 3. RMS-error of the approximation of the partial derivative $\partial_x P_j f_{KSA_1}$.

jester-f3.eps.
Fig. 4. RMS-error of the approximation of surface KSA₂ with undefined values in the bounding box. jester-f4.eps.
Fig. 5. Maximal deviation $M_{x,j}$ of the intersection point. jester-f5.eps.
Fig. 6. Maximal deviation of the normal $M_{n,j}$ of the intersection point. jester-f6.eps.
Table 1. Parameters of the test surfaces KXY, KSA$_1$ and KSA$_2$.

<table>
<thead>
<tr>
<th></th>
<th>KXY</th>
<th>KSA$_1$</th>
<th>KSA$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>5.0</td>
<td>$b$ 1.0</td>
<td>$b$ 20.0</td>
</tr>
<tr>
<td>$\rho_x$</td>
<td>$-2.15e^{-1}$</td>
<td>$\rho$ 1.00</td>
<td>$\rho$ $-3.87e^{-2}$</td>
</tr>
<tr>
<td>$\rho_y$</td>
<td>1.22e$^{-1}$</td>
<td>$\kappa$ $-1.00$</td>
<td>$\kappa$ 0.00</td>
</tr>
<tr>
<td>$\kappa_x$</td>
<td>$-1.00$</td>
<td>$\kappa_y$ $-1.00$</td>
<td>$\kappa$ 0.00</td>
</tr>
<tr>
<td>$c_{10}$</td>
<td>$-4.05e^{-4}$</td>
<td>$c_1$ 1.50</td>
<td>$c_1$ $-4.17e^{-6}$</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>$-8.13e^{-4}$</td>
<td>$c_2$ $-7.00e^{-1}$</td>
<td>$c_2$ 4.71e$^{-9}$</td>
</tr>
<tr>
<td>$c_{14}$</td>
<td>5.73e$^{-4}$</td>
<td>$c_3$ 5.00e$^{-1}$</td>
<td>$c_3$ $-4.94e^{-12}$</td>
</tr>
<tr>
<td>$c_{21}$</td>
<td>$-4.59e^{-6}$</td>
<td>$c_4$ $-5.00e^{-1}$</td>
<td>$c_4$ $-5.42e^{-15}$</td>
</tr>
<tr>
<td>$c_{23}$</td>
<td>1.14e$^{-5}$</td>
<td>$c_5$</td>
<td>$c_5$ $-4.98e^{-18}$</td>
</tr>
<tr>
<td>$c_{25}$</td>
<td>9.64e$^{-6}$</td>
<td>$c_6$</td>
<td>$c_6$ $-1.22e^{-20}$</td>
</tr>
<tr>
<td>$c_{27}$</td>
<td>4.45e$^{-7}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_{36}$</td>
<td>$-2.69e^{-9}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_{38}$</td>
<td>$-7.96e^{-8}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_{40}$</td>
<td>$-8.79e^{-8}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_{42}$</td>
<td>$-9.16e^{-8}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_{44}$</td>
<td>2.43e$^{-8}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>