Research Article

Anika Broemel*, Uwe Lippmann and Herbert Gross

Freeform surface descriptions. Part I: Mathematical representations

DOI 10.1515/aot-2017-0030
Received April 13, 2017; accepted June 23, 2017; previously published online July 21, 2017

Abstract: Optical systems can benefit strongly from freeform surfaces; however, the choice of the right surface representation is not trivial and many aspects must be considered. In this work, we discuss the general approach classical globally defined representations, as well as the basic mathematics and properties of the most commonly used descriptions and present a new description developed by us for describing freeform surfaces.

Keywords: correction; freeform surface; optical design; optimization; surface representation.

1 Introduction

In recent years, great efforts can be observed in the community of optical design to use the additional degrees of freedom afforded by freeform surfaces in optical systems. In particular, there are several benefits of using these currently available technologies for mirror systems without central obscuration and with multifunctional approaches for more compact systems. If commercial software tools are considered, only very limited support of freeform surface systems is currently found. In particular, one important question at the beginning of the concept and design of a freeform system is the decision how to mathematically describe the surface itself. If the development of an optical system with freeform surfaces is considered, not only the optical design, but also the mechanical design, the manufacturing, and the assembly of the component inside the whole system are important. Experience in practical work shows that a fabricated surface with special local structures on it and the characteristic often periodic perturbation due to the tooling needs for a separate description supporting these properties [1]. Furthermore, some proposals in literature are using locally described mathematical functions like wavelets [2], splines [3], or radial basis functions [4, 5] for the layout of the surface. In this work, we focus on the design phase. Therefore, only functions that are globally defined on the computational area inside the boundary are considered.

From the viewpoint of practical work and efficiency, there are several criteria for this selection. The surface representation should allow for a fast raytrace, the parametrization should be flexible with a small number of parameters, and the optimization of the parameters should converge quickly with a good result in the design process. Additionally, the availability in the commercial software and the access to tolerancing and manufacturability is of importance.

Two possibilities that are easy to implement are a simple two-dimensional (2D) Taylor expansion into monomials in the transverse coordinates x and y. This simple approach suffers from some drawbacks, namely, that these representations are non-orthogonal, and there is no easy link to the classical definitions of primary aberrations. Another opportunity, which is often found, is the use of Zernike polynomials [6]. These functions are in widespread use in optics for the characterization of wavefront aberrations for circular pupils [7] and also for very special pupil shapes like hexagonal, rectangular, circular, and ring sections and ellipses [8–11].

The Zernike polynomials have the advantage of a direct link to the aberration-correcting impact of the freeform surfaces. A drawback of this approach is the restriction onto circular symmetric boundaries. In contrast to the monomial representation, Zernike functions have the great benefit to be orthogonal on the area of definition. This improves the numerical algorithms and the performance of optimization in the design process [12].

***Corresponding author: Anika Broemel,** Institute of Applied Physics, Friedrich-Schiller University, Albert-Einstein-Straße 15, 07745 Jena, Germany, e-mail: anika.broemel@uni-jena.de
Uwe Lippmann: Fraunhofer Institute of Applied Optics and Precision Engineering IOF, Albert-Einstein-Straße 7, 07745 Jena, Germany
Herbert Gross: Institute of Applied Physics, Friedrich-Schiller University, Albert-Einstein-Straße 15, 07745 Jena, Germany; and Fraunhofer Institute of Applied Optics and Precision Engineering IOF, Albert-Einstein-Straße 7, 07745 Jena, Germany

As the fundamental work of Forbes [13], it is known that there are two different possibilities to construct orthogonal surfaces. The classical property of orthogonality in the spatial domain is comfortable if wave aberrations are the criterion to optimize the system. On the other hand, the expansion functions can be made orthogonal in slope. This is more adapted to a correction of the transverse aberrations as they are linear relative to the gradient of the wavefront. Furthermore, in gradient orthogonal polynomials, the expansion coefficients are directly related to the slope of the surface and, therefore, can be used for the tolerancing of freeform surfaces with great success. Until now, the Forbes approach is found in literature only for a circular symmetric geometry of the surface boundary and circular symmetric basic shapes. If the shape of the light footprint has a pronounced elongated geometry, it is, therefore, an advantage to reconstruct the basic Forbes polynomials to fit a rectangular shape of the boundary. Moreover, Forbes introduced the concept of the 'best-fit' shape, which gives direct access to the departure from the basic shape. This approach is specifically beneficial for tolerancing and manufacturing aspects.

In the mathematical theory of orthogonal functions, one important aspect is the selection of the weighting function in the overlap integral of two functional terms. The influence of the weighting function governs the accentuation of the regions inside the area of definition. Typically freeform surfaces are used to correct higher-order aberrations. These higher orders are mostly observed far from the optical axis or central ray. Therefore, it is logical to implement a weighting function that emphasizes the outer regions of the surface. From a physical optical point of view, the impact of correcting deviations of a surface on the point spread function have to be weighted with the intensity of the light distribution [1]. The best choice of a weighting function, therefore, is the exact analytical footprint of light on the surface under consideration.

As the surfaces are 2D, one more option to consider in selecting the correct expansion function is the question of the decomposition of the functions. They can be constructed as a full 2D function, orthogonal over the area of definition, or they can be built as a Cartesian product of two 1D functional systems in x and y separately. The special properties of the functional products in the second case are sometimes critical, due to the zero lines parallel to the x- and y-axis, which leads to a failure of every second order in case of double-plane symmetry and is not able to represent most of the lower-order aberrations like both coma terms and one of the two astigmatism terms. Therefore, the former method is preferred.

Another aspect is the normalization of the basic area according to the maximum dimensions. This creates expansion coefficients without changing units, and the convergence of the series can be observed directly in the descending absolute values of the coefficients. Furthermore, all algorithmic computations are more robust if this normalization is used.

In the present work, the main approaches of describing freeform surfaces are discussed from a more mathematical point of view. The possibilities known from literature are discussed as well as some new definitions according to the aforementioned opportunities. In Section 2, the fundamental approach to describe freeform surfaces in optics is explained, and the various options considering the basic shape, the boundary of the supported domain, and the projection direction are discussed. In the third section, the development and properties of the classical polynomial description are described. Section 4 gives a detailed overview of the specific description which is, in particular, used in optical system freeform surfaces with great benefit and introduces the newly defined A-polynomials. Finally, a conclusion summarizes the main findings and recommendations.

The consequences and recommendations for practical design work are discussed in a forthcoming paper by comparing the properties of the possible surface representations in a benchmark, applying the surfaces for different system types under various conditions.

2 General approach for freeform descriptions

A simple way to describe a freeform surface is a modular decomposition into two major parts: The so called basic shape, e.g. sphere, conic, or biconic incorporates mainly the quadratic contributions around the axis and are fixing the parabasal behavior of the surface. This part is complemented by a higher-order term, which describes additional deformations and is responsible for the aberration correction. The second term is typically much smaller than the basic shape and contains the freeform contributions (see Figure 1 for illustration).

The deformation terms can generally be described as a sum over a polynomial expansion set with a prefactor. The prefactor is defined as the boundary function $A(\overline{x}, \overline{y})$ and a projection factor $P(x, y)$. The function $A(\overline{x}, \overline{y})$ controls the values of the deformation terms on the boundary line of the surface. The prefactor allows, in particular, for steep surfaces to orient the additional sag correction

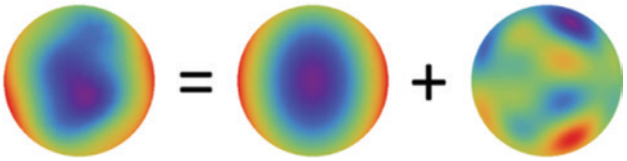


Figure 1: Decomposition of sag of a freeform surface (left) into the basic shape (middle) and higher-order deformations (right).

along the local normal of the surface and not simply along the z-axis. This decomposition is formally expressed

$$z(x, y) = Z_{\text{basic}} + \frac{A(\bar{x}, \bar{y})}{P(x, y)} \sum \mathrm{F}(\bar{x}, \bar{y}). \tag{1}$$

Here, the boundary function is formulated with the normalized coordinates according to

$$\bar{x} = x / x_{\max} \text{ and } \bar{y} = y / y_{\max}.$$

2.1 Basic shape

The basic shape contains the second-order contribution of the freeform surface and is the dominating part in the neighborhood of the z-axis. The simplest option is a sphere, with only one degree of freedom (radius):

$$\mathrm{Z}_{\text{basic}}^{(\text{sphere})}(r) = \frac{cr^2}{1+[1-c^2r^2]^{1/2}}. \tag{2}$$

In the case of the Q-polynomials of Forbes [14], the sphere is modified to a so-called 'best-fit-sphere'. Here, the basic shape is defined as a sphere by the center and the circular boundary of the surface

$$\mathrm{Z}_{\text{basic}}^{(\text{bestfitsphere})}(r) = \frac{c_{\text{bfs}}r^2}{1+[1-c_{\text{bfs}}{}^2r^2]^{1/2}}. \tag{3}$$

Later, this concept was further developed for a 'best-fit-conic' [15].

Most descriptions are used with a circular symmetric conic

$$\mathrm{Z}_{\text{basic}}^{(\text{conic})}(r) = \frac{cr^2}{1+[1-(1+\kappa)c^2r^2]^{1/2}}. \tag{4}$$

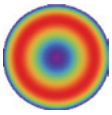
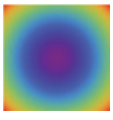
where c describes the curvature of the conic section in the vertex point and κ is the conical constant of the surface. In particular, in the case of mirror systems, one of the real problems is the occurrence of a large astigmatism for larger incidence angles of the axis ray. This can be taken into account in the second-order approximation by a non-spherical surface, and the corresponding large terms in the deviation expansion can be avoided. This is an advantage for the design and the convergence of the correction part of the freeform surface, from the viewpoint of manufacturing, where the non-circular contribution plays a major role. In the case of a biconic as a basic shape, an astigmatic behavior can be included. This is, in particular, an advantage if mirror systems are considered with large incidence angles of the axis ray

$$\mathrm{Z}_{\text{basic}}^{(\text{biconic})}(x, y) = \frac{c_x x^2 + c_y y^2}{1+[1-(1+\kappa_x)c_x{}^2(x^2+y^2)-(1+\kappa_y)c_y{}^2(x^2+y^2)]^{1/2}} \tag{5}$$

2.2 Boundary function

As already mentioned, the boundary function has the task to define or restrict the values of the additional deformation terms on the boundary of the surface. This is not a necessary condition, but in many practical cases, the behavior at the edge is then quite better controlled. In reality, the footprint of light on the surface is mostly not perfectly matching the ideal geometrical boundary. An additional overflow can help to get some flexibility for the mechanical design and the mountings. Furthermore, a clearly defined boundary value simplifies the task to relate the surface to the necessary marks and fiducials to locate and orient the surface inside the system. Extrapolating the polynomial expansions beyond the optically used diameter is an underestimated problem in practice. For simplicity, the boundary curve is selected by simple geometrical shapes. Table 1 gives some examples for boundary functions $A(\bar{x}, \bar{y})$ to fix the deformation values on the boundary and in the center, respectively.

Table 1: Typical boundary functions $A(\bar{x}, \bar{y})$ or $A(\bar{r})$ in freeform surface descriptions.

Boundary function		
1	Uniform, no special constraints (unit circle/square)	
$(1-\bar{r}^2)\,\bar{r}^2$	Center and boundary forced to be zero (unit circle)	
$(\bar{x}^2+\bar{y}^2)$	Center forced to be zero (unit square)	

2.3 Projection factor

The projection factor $P(x, y)$ has the task to define the direction of the additional correction of the deviation term onto the basic shape. Usually, the intention is to have a small correction contribution of higher orders. When the freeform surface is strongly bended, and the slope of the surface against the z-axis is large, the projection of a small change in the surface profile onto the axis direction generates a large difference. This is not really comfortable, and therefore, a second opportunity measures the deviation along the local surface normal vector of the basic shape surface. These two geometries are depicted in Figure 2. The same consideration is already well known in the case of aspheres [16]. If this additional prefactor is included in the representation of the surface, as is seen in equation (1), the expression is no longer a polynomial. In particular, this has the consequence that an exact conversion between different representations is no longer possible, and any conversion should take into account the desired accuracy and the necessary number of terms.

For surface descriptions without a projection factor, the deformation is independent of the basic shape. When a projection factor is included, it links the basic shape and the deformation.

If α is the angle between the local normal and the z-axis, the projection factor is given by $\cos(\alpha)$. It is now a question of the selected basic shape to formulate the corresponding projection functions in Cartesian coordinates. The general expression is given by equation (6), and the formulas for a sphere and a circular symmetric conic are given in equations (7) and (8), respectively:

$$P(x, y) = \cos(\alpha) = \left[1 + \left(\frac{d(Z(x, y)_{\text{basic}})}{dx}\right)^2 + \left(\frac{d(Z(x, y)_{\text{basic}})}{dy}\right)^2\right]^{-1/2}, \tag{6}$$

$$P^{(\text{sphere})}(x, y) = [1 - c^2(x^2 + y^2)]^{1/2}, \tag{7}$$

$$P^{(\text{conic})}(x, y) = \left[\frac{1 - c^2(x^2 + y^2)(1 + \kappa)}{1 - c^2\kappa(x^2 + y^2)}\right]^{1/2}. \tag{8}$$

This is only valid for slowly varying departures (i.e. slopes).

3 Structure of the polynomial set

3.1 Development

According to the discussion in the Introduction, there are several aspects, possibilities, and criteria to select and generate a function system for the deformation correction term. Mathematically, the geometry of the supported area, the weighting function, the selected orthogonality, as well as the choice of the initial shapes as a starting point of the set are of importance. As experience shows that an

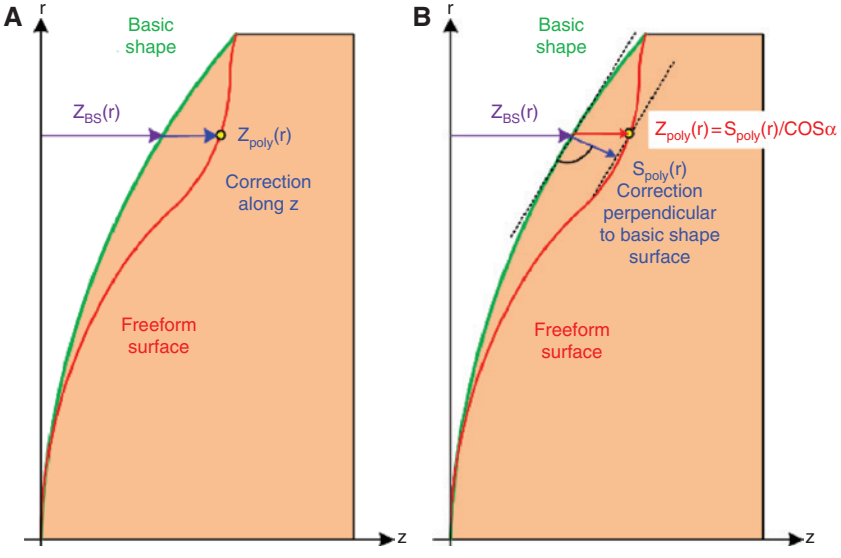


Figure 2: Direction of the correction term: (A) parallel to the z-axis, (B) perpendicular to the local surface orientation.

orthogonal set of function is quite an advantage[8], the algorithm for generating the series of functions is after the selection of the criteria above the most important step. The classical Gram-Schmidt method is usually used to guarantee the orthogonality of functions with a different index. For the detailed algorithm, the reader is referred to corresponding textbooks [6]. A special question is the selection of the basis function f_1 and f_2, which define the functional shape of the polynomial set. A second point that should be noticed is the special case of a slope orthogonality. In this case in the classical method, the calculation of the scalar product between two functions f_n and f_m must be replaced by

$$< f_n, f_m >= \iint \mathrm{w}(\bar{x},\bar{y}) \cdot \nabla[\mathrm{A}(\bar{x},\bar{y}) f_n(\bar{x},\bar{y})] \cdot \nabla[\mathrm{A}(\bar{x},\bar{y}) f_m(\bar{x},\bar{y})] d\bar{x}\, d\bar{y}. \tag{9}$$

Here, $\mathrm{w}(\bar{x},\bar{y})$ describes the weighting function, and the integration is performed over the area of support.

3.2 Properties

In this work, several approaches for the polynomial set are investigated. If a circular symmetric support area is selected, the Zernike approach is the most frequently used representation. Thus, it is included in the Fringe convention in this study. The famous Q-polynomials of Forbes are also defined on a circular area, but have the advantage of a slope orthogonality.

For Cartesian support areas, monomials are a very simple non-orthogonal opportunity. Another option is the selection of spatial orthogonal Chebyshev polynomials. The Chebyshev polynomials of the first kind have a weighting function, which grows toward the boundary. The second-kind Chebyshev polynomials instead have a weighting growing toward the center. Also, an option is the Legendre polynomials with a uniform weighting. The newly developed A-polynomials use a rectangular boundary, but follow the basic idea of the Q-polynomials.

3.3 Sorting

By definition, the functional systems have two ordering indices, describing their order and variation in x and y, respectively. Very often, a 2D matrix scheme is cumbersome, and therefore, a 1D vectorial scheme is preferred and easier from the viewpoint of the handling. If this mapping of the indices is done, a definition of the rule of conversion is necessary. From the optical point of view, it is most beneficial if with increasing index value the corresponding order or spatial frequency of the correcting function is growing. In this case, the decision of fixing the maximum index makes sense and fixes the needed largest order of correction. This scheme is realized in the Fringe convention of the Zernike polynomials. If the series is truncated at square numbers for Zernikes in Fringe convention (e.g. 1, 4, 9, 16, ...) and triangle numbers (e.g., 1, 3, 6, 10, 15, …) for monomials, Chebyshev, and Legendre polynomials, exactly a full higher order is included in the description. For Zernike and A-polynomials with nine terms, the primary aberrations of the third order in lateral deviations are covered; with 16 terms, the fifth order is included, and so on. Unfortunately, this rule is not easy to follow in the case of Cartesian polynomial sets. Table 2 shows as an example the mapping matrix between the 1D running index j between four different sets of polynomials with the 2D indices n and m. In the case of Cartesian representations, n and m are responsible for the x and y directions, respectively; in the case of the polar descriptions, n describes the radial and m the azimuthal behavior. Tables 3 and 4 visualized the corresponding shapes of the functions.

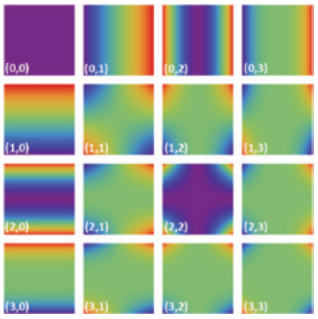
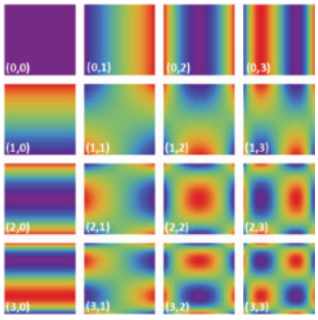
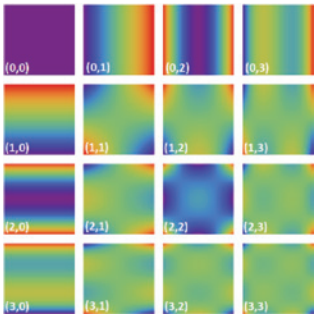
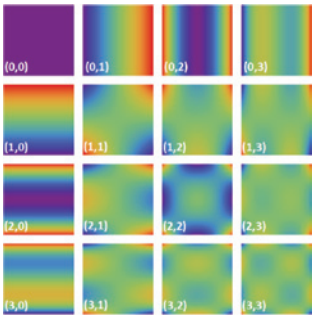
Table 2: Conversion of 2D sorting (m, n) to 1D (j) 'Fringe' sorting.

	Monomials/ Chebyshev/Legendre		Zernike	
j	m	n	m	n
1	0	0	0	0
2	1	0	1	1
3	0	1	−1	1
4	2	0	0	2
5	1	1	2	2
6	0	2	−2	2
7	3	0	1	3
8	2	1	−1	3
9	1	2	0	4
10	0	3	3	3
11	4	0	−3	3
12	3	1	2	4
13	2	2	−2	4
14	1	3	1	5
15	0	4	−1	5
16	5	0	0	6

4 Surface descriptions

In this section, the explicit mathematical terms as well as some examples and properties are summarized for the

Table 3: Overview of important properties of Cartesian-defined freeform surface representations.

Surface representation	Domain	Orthogonality	Boundary function	Weight function	Polynomial set
Monomials	Arbitrary	None	None	None	
Chebyhshev 2D (first kind)	Unit square	Spatial	1	$\frac{1}{\sqrt{[1-\bar{x}^2][1-\bar{y}^2]}}$	
Chebyhshev 2D (second kind)	Unit square	Spatial	1	$\sqrt{[1-\bar{x}^2][1-\bar{y}^2]}$	
Legendre 2D	Unit square	Spatial	1	1	

most important sets of polynomials. The descriptions can be divided into the Cartesian- and polar-defined descriptions. For all of the following descriptions, the coordinates of the polynomial set are normalized to the maximal dimensions.

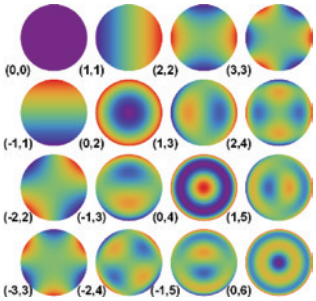
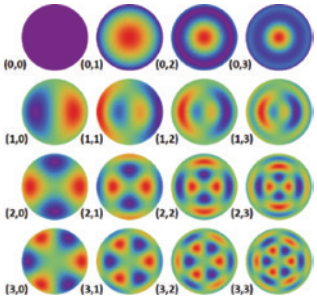
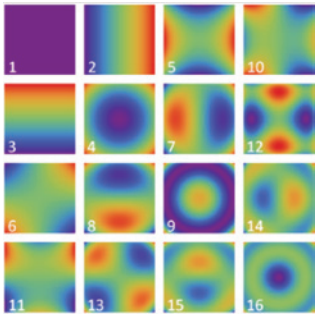
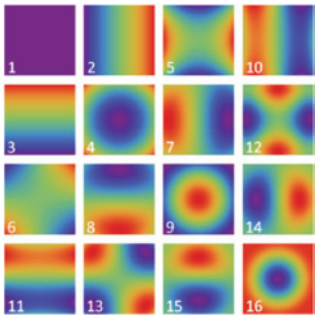
First, the Cartesian-defined descriptions are described. The properties and lower-order terms can be found in Table 3.

4.1 Monomials

Monomials, as mentioned in the introduction, are a simple Taylor expansion in x and y with no orthogonolity at all and a simple uniform weighting [6]:

$$z(x, y) = Z_{\text{basic}}(x, y) + \sum_{m,n} a_{mn} \bar{x}^m \bar{y}^n \quad \text{with } m,n = 0, 1 \ldots N. \tag{10}$$

Table 4: Overview of important properties of the polar-defined freeform surface representations.

Surface representation	Domain	Orthogonality	Boundary function	Weight function	Polynomial set
Zernike Fringe	Unit circle	Spatial	1	1	
Q-polynomials	Unit circle	Gradient	$(1-\bar{r}^2)\,\bar{r}^2$	$\frac{1}{\bar{r}\sqrt{1-\bar{r}^2}}$	
A-polynomials (first kind)	Unit square	Gradient	1	1	
A-polynomials (second kind)	Unit square	Gradient	$(\bar{x}^2+\bar{y}^2)$	$\frac{1}{\sqrt{1-\bar{x}^2}\sqrt{1-\bar{y}^2}}$	

4.2 Chebyshev 2D

The Chebyshev 2D polynomials are Cartesian products of 1D function Chebyshev polynomials. The resulting set is spatially orthogonal on a unit square [6].

The Chebyshev 2D are presented here in the first kind:

$$z(x,y)=Z_{\text{basic}}(x,y)+\sum_{n,m}a_{nm}T_n(\bar{x})T_m(\bar{y})$$
$$\text{with } m,n=0,1\ldots N. \tag{11}$$

The initial terms of the 1D functions are

$$T_0(x)=1$$
$$T_1(x)=x$$
$$T_{n+1}(x)=2xT_n(x)-T_{n-1}(x) \tag{12}$$

Moreover, the Chebyshev polynomials of the second kind are defined similarly as

$$z(x,y)=Z_{\text{basic}}(x,y)+\sum_{n,m}a_{nm}U_n(\bar{x})\mathrm{U}_m(\bar{y})$$
$$\text{with } m,n=0,1\ldots N. \tag{13}$$

The difference between these two is the weighting (Table 3), and therefore, the initial terms are

$$\begin{aligned} U_0(x) &= 1 \\ U_1(x) &= 2x \\ U_{n+1}(x) &= 2xU_n(x) - U_{n-1}(x) \end{aligned} \tag{14}$$

4.3 Legendre 2D

The Legendre polynomials are similar to the Chebyshev polynomials defined as Cartesian products of 1D functions. In contrary to the Chebyshev polynomials, the weighting function is uniform over the domain [6].

$$\begin{aligned} &z(x, y) = Z_{\text{basic}}(x, y) + \sum_{n,m} a_{nm} P_n(\bar{x}) P_m(\bar{y}) \\ &\text{with } m,n = 0, 1 \ldots N. \end{aligned} \tag{15}$$

The initial terms of the 1D functions are

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ (n+1)P_{n+1}(x) &= (2n+1)\mathrm{x}\mathrm{P}_n(x) - nP_{n-1}(x) \end{aligned} \tag{16}$$

Another option is to define the polynomial set in polar coordinates. The descriptions using this approach are described in the following, and the properties and lower-order terms can be found in Table 4.

4.4 Zernike

The Zernike polyomials are a spatially orthogonal set, which is well known for describing wavefront errors and aberrations. The terms are defined in polar coordinates and have a constant weighting function [6]. The Standard convention is:

$$z(\mathrm{r}, \varphi) = Z_{\text{basic}} + \sum_{n,m} a_{nm} Z_n^m(\bar{\mathrm{r}}, \varphi) \text{ with } m,n = 0, 1 \ldots N, \tag{17}$$

$$Z_n^m(\bar{r}, \varphi) = N(n, m) * R_n^m(\bar{r}) * \begin{Bmatrix} \sin(m\varphi), & m < 0 \\ \cos(m\varphi), & m > 0 \\ 1, & m = 1 \end{Bmatrix}$$

with the normalized radius $\bar{r}$, (18)

$$R_n^m(\bar{r}) = \sum_{k=0}^{\frac{n-|m|}{2}} (-1)^k \binom{n-m}{k} \binom{n-2k}{\frac{n-m}{2}-k} \bar{r}^{n-2k}, \tag{19}$$

$$N(n, m) = \sqrt{\frac{2(n-1)}{1-\delta_{m0}}}. \tag{20}$$

In Table 2, the conversion from the 2D Standard convention to the more convenient 1D Fringe convention can be found. The properties and lower-order Fringe terms are shown in Table 4.

4.5 Q-polynomials

The previously mentioned Q-polynomials developed by Forbes are based on his 'Mild-Asphere' approach [13, 14]. The gradient orthogonal description, defined for circular boundaries, incorporates the concept of the 'best fit' basic shape, an individual boundary function for the aspheric terms and freeform terms, as well as the projection factor and a weighting function, which emphasize the center and the boundary of the domain.

The Mild-Asphere set (m = 0) is restricted at the center and the boundary by the boundary function to achieve the 'best fit' basic shape where the surface sag is only defined by the basic shape in these two points. The further development of the freeform set (m > 0), called Q-polynomials, is only restricted at the center of the domain. Therefore, the surface sag at the center of the freeform is still only defined by the basic shape, but the boundary incorporates freeform contributions. The first terms and additional properties can be found in Table 4. The goal of this specific boundary function, the projection factor, and the weighting, is the limitation of the slopes, specifically on the boundary, to generate a manufacturing-friendly surface.

$$\begin{aligned} z(\mathrm{r},\theta) &= Z_{\text{basic}}^{(\text{bestfitsphere})} + Z_{\text{Mild-Asphere}}(\bar{r}) + Z_{\text{Q-Polynomials}}(\bar{r},\theta) \\ &= Z_{\text{basic}}^{(\text{bestfitsphere})} + \frac{1}{\sqrt{1-c^2r^2}} \cdot \left[\bar{r}^2 \cdot (1-\bar{r}^2) \cdot \sum_{n=0} a_n^0 Q_n^0(\bar{r}^2) \right] \\ &+ \frac{1}{\sqrt{1-c^2r^2}} \cdot \left[\sum_{m=1} (\bar{r})^m \sum_{n=0} \left[a_n^m \cos(m\theta) + b_n^m \sin(m\theta) \right] \cdot Q_n^m(\bar{r}^2) \right] \\ &\text{with } m,\ n = 0, 1 \ldots M,N \end{aligned} \tag{21}$$

The representations presented, so far, are showing a broad range of defined domain, geometry, orthogonality, and weighting. Nevertheless, a description incorporating the projection factor, as well as the gradient orthogonality on a rectangular domain for elongated geometry, is missing. Therefore, a new set is developed, called the A-polynomials with two versions: the first kind and the second kind, only differing in the weighting over the aperture and boundary function.

4.6 A-polynomials

The A-polynomials are a new set, specifically developed for rectangular domains. The surface description is based

on the approach by Forbes, combining a projection factor and a gradient orthogonal set for better access to manufacturing and tolerancing. Additionally, the basic shape was extended to a biconic, to include lower-order astigmatism. Because of the adjustment to a rectangular domain and a biconic basic shape, a 'best-fit-shape' is no longer meaningful. In general, the A-polynomials are described by:

$$z(x,y) = Z_{\text{basic}}^{(\text{biconic})} + \frac{A(\bar{x},\bar{y})}{P_{\text{Biconic}}(x,y)} \sum_i a_i A_i(\bar{x},\bar{y})$$
$$\text{with i} = 0, 1 \ldots N, \tag{22}$$

The previous polynomial sets for rectangular domains, like monomials or Chebyshev polynomials or the Q-Legendre [17], are based on Cartesian products of 1D functions depending on x and y. As discussed earlier, these descriptions have some drawbacks for design. Therefore, we followed the approach of Bray [18] and used the Zernike Fringe set. The polar description, defined on a unit circle, was hereby converted for a Cartesian grid, and the original domain is circumscribed by the definition area of the new set, a unit square. The Zernikes will be extrapolated to the corners of the unit square.

We developed two sets of A-polynomials: the first kind, with no restrictions for the boundary and uniform weighting and the second kind, emphasizing the boundary with the weighting function and restricting the center with the boundary function.

Both polynomial sets were developed with the Gram-Schmidt process and the modified relation to ensure slope orthogonality:

$$< f_n, f_m > = \frac{\int_{-1}^{1}\int_{-1}^{1} \text{w}(\bar{x},\bar{y}) \cdot \nabla[\text{A}(\bar{x},\bar{y}) f_n(\bar{x},\bar{y})] \cdot \nabla[\text{A}(\bar{x},\bar{y}) f_m(\bar{x},\bar{y})] d\bar{x}\, d\bar{y}}{\int_{-1}^{1}\int_{-1}^{1} \text{w}(\bar{x},\bar{y}) d\bar{x}\, d\bar{y}}$$
$$\text{with } \bar{x} = \bar{r}\cos(\theta) \text{ and } \bar{y} = \bar{r}\sin(\theta). \tag{23}$$

The functions f_n and f_m are hereby the basis function of the Zernike Fringe set in Cartesian coordinates.

For the A-polynomials of the first kind, equation (23) turns into:

$$< f_n, f_m > = \int_{-1}^{1}\int_{-1}^{1} \nabla[f_n(\bar{x},\bar{y})] \cdot \nabla[f_m(\bar{x},\bar{y})] d\bar{x}\, d\bar{y}. \tag{24}$$

In contrast, the second kind A-polynomials are developed with:

$$< f_n, f_m > = \frac{\int_{-1}^{1}\int_{-1}^{1} \frac{1}{\sqrt{1-\bar{x}^2}\sqrt{1-\bar{y}^2}} \cdot \nabla[(\bar{x}^2+\bar{y}^2) f_n(\bar{x},\bar{y})] \cdot \nabla[(\bar{x}^2+\bar{y}^2) f_m(\bar{x},\bar{y})] d\bar{x}\, d\bar{y}}{\int_{-1}^{1}\int_{-1}^{1} \frac{1}{\sqrt{1-\bar{x}^2}\sqrt{1-\bar{y}^2}}) d\bar{x}\, d\bar{y}}. \tag{25}$$

Table 5: First terms of the A-polynomials (in polar coordinates).

	First kind	Second kind
1	1	$\frac{1}{2\sqrt{\pi}}$
2	$\frac{1}{2}\bar{r}\cos(\theta)$	$\frac{2}{5\sqrt{\pi}}\bar{r}\cos(\theta)$
3	$\frac{1}{2}\bar{r}\sin(\theta)$	$\frac{2}{5\sqrt{\pi}}\bar{r}\sin(\theta)$
4	$\frac{1}{8}\sqrt{\frac{3}{2}}(2\bar{r}^2-1)$	$\frac{1}{2\sqrt{3\pi}}(2\bar{r}^2-5)$
5	$\frac{1}{4}\sqrt{\frac{3}{2}}\bar{r}^2\cos(2\theta)$	$\frac{1}{\sqrt{10\pi}}\bar{r}^2\cos(2\theta)$
6	$\frac{1}{4}\sqrt{\frac{3}{2}}\bar{r}^2\sin(2\theta)$	$\frac{1}{5\sqrt{\pi}}\bar{r}^2\sin(2\theta)$
7	$\frac{1}{4}\sqrt{\frac{1}{3}}(3\bar{r}^3-4\bar{r})\cos(\theta)$	$\frac{8}{5\sqrt{4749\pi}}(25\bar{r}^3-56\bar{r})\cos(\theta)$
8	$\frac{1}{4}\sqrt{\frac{1}{3}}(3\bar{r}^3-4\bar{r})\sin(\theta)$	$\frac{8}{5\sqrt{4749\pi}}(25\bar{r}^3-56\bar{r})\sin(\theta)$
9	$\frac{1}{32}\sqrt{\frac{21}{31}}(15\bar{r}^4-28\bar{r}^2+9)$	$\frac{1}{2\sqrt{105\pi}}(16\bar{r}^4-58\bar{r}^2+61)$

The resulting A-polynomials kept the main structure of the Zernike-Fringe set, as well as the polar character. Moreover, the order and sorting of the A-polynomials are equivalent to the Zernike Fringe set. The similarities can be seen in Tables 4 and 5, with the first nine terms of the A-polynomial first and second kind in polar expression. The advantages by initially using Zernikes, like the direct link to the aberrations are still given, but now defined on a unit square with the possibility of direct tolerancing. Because of the use of a projection factor, the direct conversion into a polynomial is no longer possible.

5 Conclusion

A comprehensive investigation of possible freeform surface representations is presented. The criteria of selection are

discussed, and the meaning of the basic shape term is explained. In particular, the description of the second-order astigmatic effects in the case of mirror systems is preferred to be included in this extra term. The meaning of the corresponding differences considering the basic shape, the projection factor, and a function controlling the boundary behavior as well as the various opportunities in the selection of the polynomial set is described. In particular, the spatial or slope orthogonality as well as the geometry of the supporting domain and the coordinate basis is distinguished. Moreover, a new set – the A-polynomials – is introduced.

In a forthcoming publication, the corresponding cases and opportunities are evaluated and compared from a practical viewpoint within a benchmark of different optical system types. Corresponding recommendations and experiences for the practical tasks of a designer are given there.

Acknowledgment: The authors would like to thank Johannes Hartung and Chonghuai Ma for valuable discussions throughout the work on this publication and the Federal Ministry of Education and Research (BMBF) (03WKCK1D) for funding this work, as part of the project 'fo+'.

References

[1] J. Stock, A. Broemel, J. Hartung, D. Ochse and H. Gross, Appl. Opt. 56, 391–396 (2017).
[2] P. Jesterm, C. Menke and K. Urban. IMA. J. Appl. Math. 77, 495–515 (2012).
[3] G. Gregory, E. Freniere and L. Gardner, Proc. SPIE Optical Design and Analysis Software II, 4769, 75 (2002).
[4] M. Maksimovic, Proc. SPIE 9626, 962613-1 (2015).
[5] I. Kaya and J. P. Rolland, Adv. Opt. Technol. 2, 81–88 (2013).
[6] M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions. NIST (1964).
[7] V. Mahajan and M. Aftab, Appl. Opt. 49, 6489–6501 (2010).
[8] A. Broemel, H. Gross, D. Ochse, U. Lippmann, C. Ma, et al., Proc. SPIE 9626, 96260W (2015).
[9] V. Mahajan and G. Dai, Opt. Lett. 31, 16 (2006).
[10] V. Mahajan, Appl. Opt. 49, 6924–6929 (2010).
[11] J. A. Díaz and V. Mahajan, Appl. Opt. 52, 1136–1147 (2013).
[12] J. A. Díaz and R. Navarro, Appl. Opt. 53, 2051–2057 (2014).
[13] G. W. Forbes, Opt. Exp. 15, 5218–5226 (2007).
[14] G. W. Forbes, Opt. Exp. 20, 2483–2499 (2012).
[15] G. W. Forbes, Opt. Exp. 19, 9923–9941 (2011).
[16] M. Krautter, Proc. DGaO 1982, Luzern, Mathematische Darstellung von steilen Rotationsfläche.
[17] M. Nikolic, P. Benítez, B. Narasimhan, D. Grabovickic, J. Liu et al., Opt. Eng. 55, 071204 (2016).
[18] M. Bray, Proc. SPIE Optical Fabrication, Testing, and Metrology 5252, 314 (2004).

Anika Broemel
Institute of Applied Physics
Friedrich-Schiller University
Albert-Einstein-Straße 15, 07745 Jena
Germany
anika.broemel@uni-jena.de

Anika Broemel studied Physics at the Friedrich-Schiller University Jena. She received her diploma in the field of Thin-Film Physics in 2010. She then joined the Institute of Photonic Technology to work on optical filters for THz applications. Since 2014, she is working in the Optical System Design group at the Institute of Applied Physics in Jena. Her main research interest is surface descriptions for freeform systems.

Uwe Lippmann
Fraunhofer Institute of Applied Optics and Precision Engineering IOF
Albert-Einstein-Straße 7, 07745 Jena
Germany

In 2002, Uwe Lippmann received his Diploma in Mechanical Engineering from the Technical University of Ilmenau. Since 2003, he has been working in the Optical Systems Department of the Fraunhofer Institute for Applied Optics and Precision Engineering in Jena. His fields of work include optical system design for a wide variety of applications as well as the simulation and analysis of optical systems with a focus on stray light analyses.

Herbert Gross
Institute of Applied Physics, Friedrich-Schiller University, Albert-Einstein-Straße 15, 07745 Jena, Germany; and Fraunhofer Institute of Applied Optics and Precision Engineering IOF, Albert-Einstein-Straße 7 07745 Jena, Germany

Herbert Gross studied Physics at the University of Stuttgart. He received his PhD on Laser Simulation in 1995. He joined Carl Zeiss in 1982 where he worked as a scientist in optical design, modeling, and simulation. From 1995 to 2010 he headed the central department of optical design and simulation. Since 2012, he has been a professor at the University of Jena in the Institute of Applied Physics and holds a chair of Optical System Design. His main working areas are physical optical simulations, beam propagation, partial coherence, classical optical design, aberration theory, system development, and metrology. He was editor and main author of the book series 'Handbook of Optical systems'.