Orthogonal basis with a conicoid first mode for shape specification of optical surfaces

Chelo Ferreira,[1] **José L. López,**[2,*] **Rafael Navarro,**[3] **and Ester Pérez Sinusía**[1]

[1]*Dpto. de Matemática Aplicada & IUMA, Universidad de Zaragoza, 50009 Zaragoza, Spain*
[2]*Dpto. de Ingeniería Matemática e Informática, Universidad Pública de Navarra, 31006 Pamplona, Spain*
[3]*ICMA, Consejo Superior de Investigaciones Científicas & Universidad de Zaragoza, 50009 Zaragoza, Spain*

[*]*jl.lopez@unavarra.es*

Abstract: A rigorous and powerful theoretical framework is proposed to obtain systems of orthogonal functions (or shape modes) to represent optical surfaces. The method is general so it can be applied to different initial shapes and different polynomials. Here we present results for surfaces with circular apertures when the first basis function (mode) is a conicoid. The system for aspheres with rotational symmetry is obtained applying an appropriate change of variables to Legendre polynomials, whereas the system for general freeform case is obtained applying a similar procedure to spherical harmonics. Numerical comparisons with standard systems, such as Forbes and Zernike polynomials, are performed and discussed.

References and links

1. R. N. Wilson, *Reflecting Telescope Optics I*, (Springer, 2004).
2. Y. Ding, X. Liu, Z.-R. Zheng, and P.-F. Gu, "Freeform LED lens for uniform illumination," Opt. Express **16**, 12958–12966 (2008).
3. M. Victoria, C. Domínguez, I. Antón, and G. Sala, "Comparative analysis of different secondary optical elements for aspheric primary lenses," Opt. Express **17**, 6487–6492 (2009).
4. Y. Wu, J. Xi, M. J. Cobb, and X. Li, "Scanning fiber-optic nonlinear endomicroscopy with miniature aspherical compound lens and multimode fiber collector," Opt. Lett. **34**, 953–955 (2009).
5. D. J. Meister and S. W. Fisher, "Progress in the spectacle correction of presbyopia. Part 2: Modern progressive lens technologies," Clin Exp Optom. **91**, 251–264 (2008).
6. R. Navarro, "Adaptive model of the aging emmetropic eye and its changes with accommodation," J. Vis. **14**(21), 1–17 (2014).
7. F. Chen, G. M. Brown, M. Song, "Overview of three-dimensional shape measurement using optical methods," Opt. Engineering **39**, 10–22 (2000).
8. D. Michaelis, P. Schreiber, and A. Bräuer, "Cartesian oval representation of freeform optics in illumination systems," Opt. Lett. **36**, 918–920 (2011).
9. H. Ries and J. Muschaweck, "Tailored freeform optical surfaces," J. Opt. Soc. Am. A **19**, 590–595 (2002).
10. G. W. Forbes, "Shape specification for axially symmetric optical surfaces," Opt. Express **15**(8), 5218–5226 (2007).
11. A. B. Bhatia, E. Wolf, and M. Born, "On the circle polynomials of Zernike and related orthogonal sets," Proc. Camb. Philos. Soc. **50**(1), 40–48 (1954).
12. G. W. Forbes, "Robust, efficient computational methods for axially symmetric optical aspheres," Opt. Express **18**(19), 19700–19712 (2010).

13. G. W. Forbes, "Manufacturability estimates for optical aspheres," Opt. Express **19**(10), 9923–9941 (2011).
14. G. W. Forbes, "Characterizing the shape of freeform optics," Opt. Express **20**(3), 2483–2499 (2012).
15. G. W. Forbes, "Fitting freeform shapes with orthogonal bases," Opt. Express **21**(16), 19061–19081 (2013).
16. R. Navarro, J. L. López, J. A. Díaz, and E. Pérez Sinusía, "Generalization of Zernike polynomials for regular portions of circles and ellipses," Opt. Express **22**, 21263–21279 (2014).
17. C. Ferreira, J. L. López, R. Navarro , and E. Pérez Sinusía, "Zernike-like systems in polygons and polygonal facets," Appl. Opt. **54**, 6575–6583 (2015).
18. E. W. Weisstein, "Legendre Polynomial", From MathWorld. A Wolfram Web Resource. http://mathworld.wolfram.com/LegendrePolynomial.html.
19. E. W. Weisstein, "Spherical Harmonic", From MathWorld. A Wolfram Web Resource. http://mathworld.wolfram.com/SphericalHarmonic.html.
20. E. W. Weisstein, "Associated Legendre Polynomial", From MathWorld. A Wolfram Web Resource. http://mathworld.wolfram.com/AssociatedLegendrePolynomial.html.

1. Introduction

The number and relevance of applications of aspheric and freeform optics is continuously increasing, ranging from astronomy [1], industry [2], solar energy [3], biomedical optics [4], or physiological optics [5], among others. The high complexity of optical surfaces found in biological systems such as the human eye [6], or the new advances in fabrication and testing of freeform surfaces [7], are demanding precise, robust and efficient methods of specifying these surfaces. Ideally, the shape specification should be physically meaningful and invariant for the different stages of design, fabrication, testing or application. In optics, it is common to represent the surface sag z as a function of the coordinates $z = \widetilde{f}(x,y)$. Often the configuration is nearly rotationally symmetric and hence it is better to work in cylindrical coordinates $z = f(r,\theta)$. The variety of systems of representations range from the pure sampling grid of points, localized splines, or global or modal representations given by combinations of functions such as spheres, conicoids, monomials, polynomials, etc. Specific methods of representation tailored for specific applications were also proposed, such as generalized Cartesian ovals [8] or solutions of specific differential equations [9], among others. Due to the high relevance of spheres, the most widely used characterizations of optical surfaces is the sum of a sphere (or conicoid C) plus an aspherical part A usually given as a linear combination of terms $z = C + A$. In what follows we will talk of conicoids C which include the sphere as a particular case.

Typically the terms specifying A are either monomials or polynomials. Historically, monomials were used first due to their apparent simplicity, but as Forbes pointed out referring to monomials [10], "the most widely used characterization of surface shape is numerically deficient", mainly due to their lack of orthogonality. Orthogonal systems of polynomials, such as Zernike polynomials [11], Forbes polynomials [10, 12–15], etc., permit to overcome a series of key issues ranging from numerical instabilities to effective tolerance specifications. Furthermore, basic linear algebra tells us that the two crucial properties of a good system of representation (sets of basis functions) are orthogonality and completeness. An additional, but less crucial property is normality (norm unity for all basis functions). Orthogonality, that implies the linear independence between the basis functions, implies also good numerical behavior, avoids redundancy and ensures uniqueness of the representation, among other highly important properties. In this context completeness is even more crucial as it means that the system can represent all possible surface shapes, that is to insure that we have a real freeform system.

In this sense the most widely used methods of optical surface representation in the form of $z = C + A$ are essentially non orthogonal, even when they use Zernike polynomials, Forbes polynomials or orthogonal systems to represent the departure from the sphere (or conicoid), simply because the conicoid itself is not orthogonal to A. Here our goal was to solve this problem to obtain a system of representation in which C is orthogonal to A (of course the basis functions of A are orthogonal as well). This implies that C is one of the basis functions of the system.

To this end, we divided the main goal into the following specific objectives: (1) to develop a general theoretical framework to obtain this type of systems, initially restricted to rotationally symmetric surfaces $z = f(r)$; (2) to obtain and implement a complete orthogonal system, in which the first basis function is a conicoid; (3) to generalize the above results to non-rotationally symmetric surfaces which is essential to obtain true freeform systems; and (4) to test numerically our new systems and perform direct comparisons with previous methods. The next Sections are organized accordingly.

2. Basis for rotationally symmetric surfaces

In this section we introduce the general framework of our theory, restricted to rotationally symmetric surfaces defined by an equation of the form $z = f(r)$, $r \in [0,1]$, where r and z are cylindrical coordinates. We design an orthogonal system for $L_v^2[0,1]$ with measure $dv = rdr$, in which the first element of the system is a specified conicoid C that, in the following, we denote by $q_0(r)$ for convenience. The remaining elements of the system, the elements of the set A, are denoted by $q_n(r)$, $n = 1, 2, 3, \ldots$ The functions $q_n(r)$, $n = 1, 2, 3, \ldots$, are constructed using three essential ingredients: (i) an arbitrary orthonormal system $\{p_n(x)\}_{n=0,1,2,\ldots}$, with respect to a certain measure $d\mu = \rho(x)dx$ in an interval $[c,d]$, (ii) the selected conicoid $q_0(r)$ and (iii) a convenient change of variable $x = \varphi(r)$, $\varphi : [0,1] \to [c,d]$. A similar method based on a change of variables was successfully applied before to obtain orthogonal Zernike-like sytems on non-circular apertures [16], polygons and polygonal facets [17]. The resulting orthogonal system consists of functions $\{q_0(r), q_1(r), q_2(r), \ldots\}$ defined in the interval $[0,1]$ that are orthogonal with respect to the measure $dv = rdr$. Moreover, the functions $q_1(r), q_2(r), \ldots$ have also norm unity. Roughly speaking, the idea is the following: we use $\varphi(r)$ to replace the first element p_0 of the system $\{p_n(x)\}_{n=0,1,2,\ldots}$ by $q_0(r)$. To preserve the orthogonality of the new system, we must choose $\varphi(r)$ appropriately. In the remaining of this section we develop this idea and give an important example.

Let $\{p_n(x)\}_{n=0,1,2,\ldots}$ be an orthonormal basis of $L_\mu^2[c,d]$ with $p_0(x) = p_0$ constant and measure $d\mu = \rho(x)dx$. This means that

$$\delta_{m,n} = \int_c^d p_n(x)p_m(x)\rho(x)dx, \quad n,m = 0,1,2,\ldots \tag{1}$$

After a (at this moment unknown) change of variable $\varphi : [0,1] \to [c,d]$; $x = \varphi(r)$, with $\varphi(0) = c$, $\varphi(1) = d$ and $\varphi'(r) > 0$ in $(0,1)$, we find

$$\delta_{m,n} = \int_0^1 p_n(\varphi(r))p_m(\varphi(r))\rho(\varphi(r))\varphi'(r)dr, \quad n,m = 0,1,2,\ldots \tag{2}$$

We observe that, when we define

$$q_n(r) := C_n p_n(\varphi(r))w(r), \quad w(r) := \sqrt{\frac{\rho(\varphi(r))\varphi'(r)}{r}}, \tag{3}$$

with $C_n = 1$ for $n = 1, 2, 3, \ldots$, and C_0 arbitrary at this moment; we find that the functions $q_n(r)$, $n = 1, 2, 3, \ldots$ are orthonormal in $[0,1]$ with respect to the measure $dv = rdr$. It is worth to note that the variable r in the denominator of $w(r)$ is not dangerous as the numerator behaves as r when $r \to 0$, as we see in the first line of (4). To complete our basis, we need to introduce a first element $q_0(r)$, the arbitrary conicoid, assuring that it is orhogonal to all the elements $q_n(r)$, $n = 1, 2, 3, \ldots$ This is achieved by choosing the function $\varphi(r)$ as the unique solution of

the boundary value problem:

$$\begin{cases} \rho(\varphi(r))\varphi'(r) = \dfrac{1}{C_0^2 p_0^2} r q_0^2(r), \\ \varphi(0) = c, \quad \varphi(1) = d. \end{cases} \tag{4}$$

Since the first-order differential equation in Eq. (4) can be directly solved by integration, the unique solution $x = \varphi(r)$ of Eq. (4) is implicitly defined by the equation

$$\int_c^x \rho(t)dt = \frac{1}{C_0^2 p_0^2} \int_0^r s q_0^2(s)ds, \tag{5}$$

when we take

$$C_0^2 = \int_0^1 s q_0^2(s)ds \Big/ \left[p_0^2 \int_c^d \rho(t)dt \right]. \tag{6}$$

It is obvious that the left hand side of Eq. (5) is an increasing function of x and the right hand side of Eq. (5) is an increasing function of r. Then, $\varphi(r)$ is a monotonic function with $\varphi'(r) > 0$.

Thus, we have that the set $\{q_n(r)\}_{n=0,1,2,\dots}$ is a quasi-orthonormal basis of $L_v^2[0,1]$ (orthonormal except for the fact that $||q_0||_r^2 = C_0^2$) with

$$q_n(r) = \frac{q_0(r)}{C_0 p_0} C_n p_n(\varphi(r)), \quad n = 0,1,2,\dots \tag{7}$$

Moreover, $\{q_n(r)\}_{n=0,1,2,\dots}$ is complete in $L_v^2[0,1]$. For any function $F(r) \in L_v^2[0,1]$, we define

$$f(x) := \frac{F(\varphi^{-1}(x))}{w(\varphi^{-1}(x))}, \tag{8}$$

that belongs to $L_\mu^2[c,d]$:

$$||f||_\rho^2 = \int_a^b |f(x)|^2 \rho(x)dx = \int_0^1 |f(\varphi(r))|^2 \rho(\varphi(r))\varphi'(r)dr = \int_0^1 |F(r)|^2 r dr = ||F||_r^2. \tag{9}$$

As $\{p_n(x)\}_{n=0,1,2,\dots}$ is complete in $L_\mu^2[c,d]$, we have that

$$f(x) = \sum_{n=0}^\infty c_n p_n(x), \quad \text{almost everywhere in } [c,d], \tag{10}$$

with

$$c_n := \int_c^d p_n(x)f(x)\rho(x)dx = \frac{1}{C_n} \int_0^1 q_n(r)F(r)rdr, \quad n = 0,1,2,3,\dots \tag{11}$$

Then, $\{q_n(r)\}_{n=0,1,2,\dots}$ is a complete system in $L_v^2[0,1]$, since for any function $F(r) \in L_v^2[0,1]$, we have

$$F(r) = w(r)f(\varphi(r)) = \sum_{n=0}^\infty c_n w(r) p_n(\varphi(r)) = \frac{c_0}{C_0} q_0(r) + \sum_{n=1}^\infty c_n q_n(r). \tag{12}$$

Solution for normalized Legendre polynomials

The normalized Legendre polynomials [18]

$$p_n(x) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k}^2 (x-1)^{n-k}(x+1)^k, \quad p_0(x) = \frac{1}{\sqrt{2}}, \tag{13}$$

are an orthonormal basis of $L^2[-1,1]$ with respect to the weight function $\rho(x) = 1$. We take the following conicoid:

$$q_0(r) = b\sqrt{1 - s\frac{L^2 r^2}{a^2}}, \tag{14}$$

with $s = \pm 1$. If $s = 1$ and $b > 0$, q_0 is the portion of an ellipse with semiaxes a and b in the upper half-plane between the angles $\arctan(b\sqrt{a^2-L^2}/a)$ and $\pi/2$, and with $0 < L < a$. If $s = 1$ and $b < 0$, q_0 is the portion of an ellipse with semiaxes a and $-b$ in the lower half-plane between the angles $-\pi/2$ and $-\arctan(b\sqrt{a^2-L^2}/a)$, and with $0 < L < a$. If $s = -1$ and $b > 0$, q_0 is the portion of a hyperbola with semiaxes a and b in the upper half-plane between the angles $\arctan(b\sqrt{a^2+L^2}/a)$ and $\pi/2$, and if $s = -1$ and $b < 0$, q_0 is the portion of a hyperbola with semiaxes a and $-b$ in the lower half-plane between the angles $-\pi/2$ and $-\arctan(b\sqrt{a^2+L^2}/a)$. In any case, from Eq. (5),

$$x + 1 = \int_{-1}^{x} dt = \frac{2}{C_0^2} \int_0^r t\left(b^2 - s\frac{b^2}{a^2}L^2 t^2\right) dt = \frac{b^2 r^2}{C_0^2}\left[1 - s\frac{L^2}{a^2}\frac{r^2}{2}\right]. \tag{15}$$

We obtain the value of C_0 from Eq. (6),

$$C_0 = \frac{1}{2}\sqrt{2b^2 - s\frac{b^2}{a^2}L^2}. \tag{16}$$

Thus,

$$x = \varphi(r) = \frac{2r^2}{2a^2 - sL^2}\left[2a^2 - sL^2 r^2\right] - 1. \tag{17}$$

Therefore, we have that the set

$$q_n(r) = 2C_n \operatorname{sign}(b)\sqrt{\frac{2a^2 - 2sL^2 r^2}{2a^2 - sL^2}} p_n\left(\frac{2r^2}{2a^2 - sL^2}\left[2a^2 - sL^2 r^2\right] - 1\right), \quad n = 0, 1, 2, \dots, \tag{18}$$

is a quasi-orthonormal basis of $L_\nu^2[0,1]$ with $d\nu = rdr$ and any function $F(r) \in L_\nu^2[0,1]$ can be written as

$$F(r) = \frac{c_0}{C_0}b\sqrt{1 - s\frac{L^2 r^2}{a^2}} + \sum_{n=1}^{\infty} c_n q_n(r), \tag{19}$$

with c_n, C_0 and $q_n(r)$ given in Eqs. (11), (16) and (18) respectively.

In Table 1 we can find the first five functions $q_0(r), q_1(r), \dots, q_4(r)$. The graphs of these functions (up to $n = 4$) are illustrated in Fig. 1 for a elliptical cap, $r \in [0,1]$, $s = 1$, $b = 1$ and $a = L = 3/4$.

Observation 1. In optics, it is standard to express rotationally symmetric surfaces in terms of their deviation from the sagittal representation. If we consider as first term

$$q_0(r) = \frac{cL^2 r^2}{1 + \sqrt{1 - \varepsilon c^2 L^2 r^2}}, \tag{20}$$

Table 1. First five functions of the new quasi-orthonormal basis $\{q_n(r)\}_{n=0,1,2,\dots}$ (see Eq. (18)) obtained from the normalized Legendre polynomials and $q_0(r) = b\sqrt{1 - s\frac{L^2 r^2}{a^2}}$.

n	$q_n(r)$
0	$b\sqrt{1 - s\dfrac{L^2 r^2}{a^2}}$
1	$-2\sqrt{3}\,\mathrm{sign}(b)\dfrac{\sqrt{a^2 - sL^2 r^2}}{(2a^2 - sL^2)^{3/2}}\left(2sL^2 r^4 - 4a^2 r^2 + 2a^2 - sL^2\right)$
2	$2\sqrt{5}\,\mathrm{sign}(b)\dfrac{\sqrt{a^2 - sL^2 r^2}}{(2a^2 - sL^2)^{5/2}}$ $\times\left(6s^2 L^4 r^8 - 24sL^2 a^2 r^6 - 6\left(s^2 L^4 - 2sL^2 a^2 - 4a^4\right)r^4 + 12a^2\left(sL^2 - 2a^2\right)r^2 + \left(sL^2 - 2a^2\right)^2\right)$
3	$-2\sqrt{7}\,\mathrm{sign}(b)\dfrac{\sqrt{a^2 - sL^2 r^2}}{(2a^2 - sL^2)^{7/2}}$ $\times\left(20s^3 L^6 r^{12} - 120s^2 L^4 a^2 r^{10} - 30sL^2\left(s^2 L^4 - 2sL^2 a^2 - 8a^4\right)r^8 + 40a^2\left(3s^2 L^4 - 6sL^2 a^2 - 4a^4\right)r^6\right.$ $\left.+12\left(s^3 L^6 - 4s^2 L^4 a^2 - 6sL^2 a^4 + 20a^6\right)r^4 - 24a^2\left(sL^2 - 2a^2\right)^2 r^2 - \left(sL^2 - 2a^2\right)^3\right)$
4	$6\,\mathrm{sign}(b)\dfrac{\sqrt{a^2 - sL^2 r^2}}{(2a^2 - sL^2)^{9/2}}$ $\times\left(70s^4 L^8 r^{16} - 560s^3 L^6 a^2 r^{14} - 140s^2 L^4(s^2 L^4 - 2sL^2 a^2 - 12a^4)r^{12} + 280sL^2 a^2(3s^2 L^4 - 6sL^2 a^2 - 8a^4)r^{10}\right.$ $+10(9s^4 L^8 - 36s^3 L^6 a^2 - 132s^2 L^4 a^4 + 336sL^2 a^6 + 112a^8)r^8 - 40a^2(9s^3 L^6 - 36s^2 L^4 a^2 + 8sL^2 a^4 + 56a^6)r^6$ $\left.-20(sL^2 - 2a^2)^2(s^2 L^4 - 2sL^2 a^2 - 18a^4)r^4 + 40a^2(sL^2 - 2a^2)^3 r^2 + (sL^2 - 2a^2)^4\right)$

with $0 < L^2 \le 1/(\varepsilon c^2)$ if $\varepsilon > 0$, and as initial orthonormal basis the normalized Legendre polynomials given in Eq. (13), we obtain a new quasi-orthonormal basis $\{q_n(r)\}_{n=0,1,2,\dots}$ of $L^2_v[0,1]$ with $dv = r\,dr$ given by

$$q_n(r) = \frac{\sqrt{2}q_0(r)}{C_0}C_n p_n(\varphi(r)), \quad n = 0, 1, 2, \dots, \tag{21}$$

with

$$C_0^2 = \frac{8\left[(1 - \varepsilon c^2 L^2)^{3/2} - 1\right] + 3\varepsilon c^2 L^2(4 - \varepsilon c^2 L^2)}{12\varepsilon^3 c^4 L^2}, \quad C_n = 1, \quad n = 1, 2, 3, \dots \tag{22}$$

and

$$\varphi(r) = 2\frac{8\left[(1 - \varepsilon c^2 L^2 r^2)^{3/2} - 1\right] + 3\varepsilon c^2 L^2 r^2(4 - \varepsilon c^2 L^2 r^2)}{8\left[(1 - \varepsilon c^2 L^2)^{3/2} - 1\right] + 3\varepsilon c^2 L^2(4 - \varepsilon c^2 L^2)} - 1, \tag{23}$$
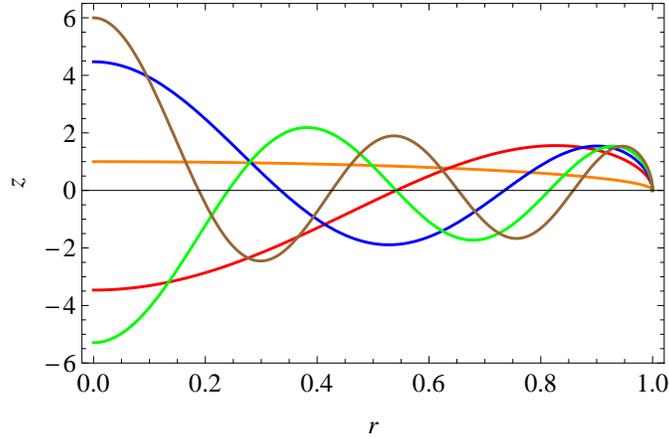
Fig. 1. Graphs of the first five functions of the new quasi-orthonormal basis $\{q_n(r)\}_{n=0,1,2,\ldots}$ (see Eq. (18)) obtained from the normalized Legendre polynomials and $q_0(r) = b\sqrt{1 - s\frac{L^2 r^2}{a^2}}$, with $s = 1$, $b = 1$ and $a = L = 3/4$ using different scale: q_0 (orange), q_1 (red), q_2 (blue), q_3 (green), q_4 (brown).

that is,

$$q_n(r) = 2\sqrt{6} C_n \frac{c^3 L^3 r^2}{1 + \sqrt{1 - \varepsilon c^2 L^2 r^2}} \sqrt{\frac{\varepsilon^3}{8\left[(1 - \varepsilon c^2 L^2)^{3/2} - 1\right] + 3\varepsilon c^2 L^2 (4 - \varepsilon c^2 L^2)}}$$

$$\times p_n\left(2\frac{8\left[(1 - \varepsilon c^2 L^2 r^2)^{3/2} - 1\right] + 3\varepsilon c^2 L^2 r^2 (4 - \varepsilon c^2 L^2 r^2)}{8\left[(1 - \varepsilon c^2 L^2)^{3/2} - 1\right] + 3\varepsilon c^2 L^2 (4 - \varepsilon c^2 L^2)} - 1\right), \quad n = 0, 1, 2, \ldots$$

(24)

This is a significantly less compact expression as compared to Eq. (18).

The theory developed in this section only applies to rotationally symmetric surfaces specified by an equation of the form $z = f(r)$, with $r \in [0, 1]$. In the following section we generalize the theory to arbitrary surfaces defined in cylindrical coordinates (r, θ, z), specified by an equation of the form $z = f(r, \theta)$.

3. Basis for freeform surfaces

In this section we formulate a more general theory to approximate arbitrary surfaces $z = f(r, \theta)$ defined over the unit disk $(r\cos\theta, r\sin\theta) \in D$. We design an orthogonal system for $L^2_v(D)$ with measure $dv = r\,dr\,d\theta$ in which the first element of the system is a specified rotationally symmetric surface $q_0^0(r)$. The remaining elements of the system $q_n^m(r, \theta)$, $n = 1, 2, 3, \ldots$, are constructed using again the three essential ingredients used in the previous section: (i) an arbitrary orthogonal system $p_n^m(x, \theta)$, $n, m(n) = 0, 1, 2, \ldots$, of $L^2_\mu([c, d] \times [0, 2\pi])$ with measure $d\mu = \rho(x)dx\,d\theta$, (ii) the surface $q_0^0(r)$ and (iii) a convenient change of variable $x = \varphi(r)$, $\varphi : [0, 1] \to [c, d]$. Then, the resulting orthogonal system consists of functions $\{q_0^0(r), q_m^n(r, \theta), \ldots\}$, $n, m(n) = 1, 2, \ldots$, defined in the unit disk D that are orthogonal with respect to the measure $dv = r\,dr\,d\theta$. Moreover, the functions $q_m^n(r, \theta), \ldots, n, m(n) = 1, 2, \ldots$ are also orthonormal. In the remaining of this section we accomplish this task and give a particular solution.

Let $\{p_n^m(x,\theta)\}_{n,m(n)=0,1,2,\ldots}$ be an orthonormal basis of $L_\mu^2([c,d]\times[0,2\pi])$ with $p_0^0(x,\theta) = p_0^0$ constant and measure $d\mu = \rho(x)dxd\theta$. We want to find a quasi-orthonormal basis $\{q_n^m(r,\theta)\}_{n,m(n)=0,1,2,\ldots}$ of $L_\nu^2(D)$, $d\nu = rdrd\theta$, with $q_0^0(r)$ predetermined. We have

$$\delta_{n,n'}\delta_{m,m'} = \int_c^d \rho(x)dx \int_0^{2\pi} d\theta\, p_n^m(x,\theta)p_{n'}^{m'}(x,\theta), \quad n,m(n)=0,1,2,\ldots \tag{25}$$

After a (at this moment unknown) change of variable $x = \varphi(r)$ with $\varphi(0) = c$, $\varphi(1) = d$ and $\varphi'(r) > 0$ in $(0,1)$, we find

$$\delta_{n,n'}\delta_{m,m'} = \int_0^1 \rho(\varphi(r))\varphi'(r)dr \int_0^{2\pi} d\theta\, p_n^m(\varphi(r),\theta)p_{n'}^{m'}(\varphi(r),\theta), \quad n,m(n)=0,1,2,\ldots \tag{26}$$

We observe that, when we define

$$q_n^m(r,\theta) := C_n^m p_n^m(\varphi(r),\theta)w(r), \quad w(r) := \sqrt{\frac{\rho(\varphi(r))\varphi'(r)}{r}}, \tag{27}$$

with $C_n^m = 1$ for $(n,m) \neq (0,0)$ and C_0^0 arbitrary at this moment, we find that the functions $q_n^m(r,\theta)$, $n,m(n)=0,1,2,\ldots$, are orthonormal in D with respect to the measure $d\nu = rdrd\theta$. To complete our basis, we need to introduce a first element $q_0^0(r)$ assuring that it is orthogonal to all the other elements $q_n^m(r,\theta)$, $n,m(n)=0,1,2,\ldots$, $(n,m) \neq (0,0)$. This is achieved by choosing the function $\varphi(r)$ the unique solution of the boundary value problem Eq. (4) given by Eq. (5).

Then, we have that the set $\{q_n^m(r,\theta)\}_{n,m(n)=0,1,2,\ldots}$ is a quasi-orthonormal system of $L_\nu^2(D)$ with

$$q_n^m(r,\theta) = C_n^m w(r)p_n^m(\varphi(r),\theta), \tag{28}$$

$q_0^0(r)$ predetermined and

$$(C_0^0)^2 = \int_0^1 s(q_0^0(s))^2 ds \left/ \left[(p_0^0)^2 \int_c^d \rho(t)dt\right]\right. . \tag{29}$$

The system $\{q_n^m(r,\theta)\}_{n,m(n)=0,1,2,\ldots}$ is complete in $L_\nu^2(D)$. The proof is similar to the proof of the one dimensional case given in Section 2. Therefore, for any $F(r,\theta) \in L_\nu^2(D)$, we have that

$$F(r,\theta) = \frac{c_0^0}{C_0^0}q_0^0(r) + \sum_{\substack{n,m(n)=0 \\ (n,m)\neq(0,0)}}^{\infty} c_n^m q_n^m(r,\theta), \tag{30}$$

with

$$c_n^m := \frac{1}{C_n^m} \int\!\!\int_D q_n^m(r,\theta)F(r,\theta)rdrd\theta, \quad n,m(n)=0,1,2,3,\ldots \tag{31}$$

Solution for spherical harmonics

For $n = 0,1,2,\ldots$, consider the spherical harmonic functions [19]

$$p_n^m(x,\theta) = \sqrt{\frac{(2-\delta_{m,0})(2n+1)(n-m)!}{4\pi(n+m)!}} \begin{cases} P_n^m(x)\cos(m\theta), & 0 \le m \le n, \\ P_n^m(x)\sin(m\theta), & -n \le m < 0, \end{cases} \tag{32}$$

where $P_n^n(x)$ are the associated Legendre polynomials [20]

$$P_n^m(x) = \frac{(-1)^m}{2^n n!}(1-x^2)^{m/2}\frac{d^{n+m}}{dx^{n+m}}(x^2-1)^n, \quad p_0^0(x,\theta) = \frac{1}{2\sqrt{\pi}}. \tag{33}$$

They are an orthonormal basis of $L^2_\mu([-1,1] \times [0,2\pi])$ with measure $d\mu = dxd\theta$. We choose the conicoid $q_0^0(r)$ given in Eq. (14) as first approximation of the surface. Then, from equation Eq. (5), and using the value of C_0^0 given in Eq. (29), we have $\varphi(r)$ given in Eq. (17). Therefore, the set $\{q_n^m(r,\theta)\}_{n=0,1,2,\ldots}^{m=-n,\ldots,n}$ is a quasi-orthonormal basis of $L^2_\nu(D)$ with $d\nu = rdrd\theta$,

$$q_n^m(r,\theta) = 2C_n \operatorname{sign}(b) \sqrt{\frac{2a^2 - 2sL^2r^2}{2a^2 - sL^2}} p_n^m \left(\frac{2r^2}{2a^2 - sL^2} \left[2a^2 - sL^2r^2\right] - 1, \theta \right) \qquad (34)$$

and, in particular, $q_0^0(r) = b\sqrt{1 - sL^2r^2/a^2}$.

Figure 2 shows the first functions (up to $n = 4$, and $m \geq 0$) for $r \in [0,1]$, $\theta \in [0,2\pi]$, $s = 1$, $a = b = 1$ and $L = 1/2$.
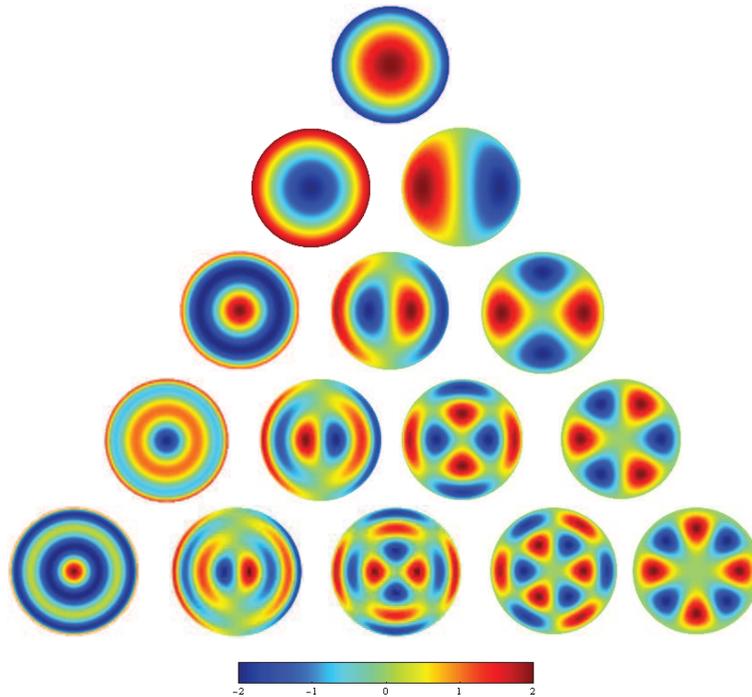

Fig. 2. First functions $q_n^m(r,\theta)$ (see Eq. (34)) for the case of the spherical harmonics and $q_0^0(r) = \sqrt{1 - r^2/4}$. The rows represent the ascending order from $n = 0$ to $n = 4$, the columns are the positive values of m from $m = 0$ to $m = n$.

4. Implementation and examples

For the numerical implementation and testing of the new basis proposed in Sections 2 and 3, we consider a Gaussian surface, as it is expected to require higher order expansions to obtain reasonable accuracies. In the first test, we use a rotationally symmetric Gaussian surface to compare with Forbes polynomials for the 1D case. The second test considers the same symmetric Gaussian surface to establish a comparison with our new basis in 2D, and the basis in 1D and Zernike polynomials. Finally, in the third test we use a non-symmetric elliptic Gaussian surface to compare our 2D system with Forbes and Zernike polynomials. In all the examples, we use the least square approximation for computing the coefficients.

Example 1. In order to check the accuracy of the approximation supplied by the basis $\{q_n(r)\}$ $_{n=0,1,2,...}$ given in Eq. (18) and compare with Forbes' approximations, we consider the following Gaussian surface of revolution (see Fig. 3),

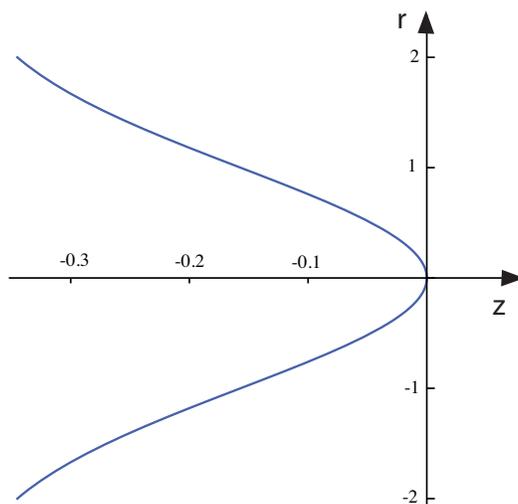$$f(r) = \frac{1}{\sqrt{2\pi}}\left(e^{-\frac{r^2}{2}} - 1\right), \quad r \in [0,1]. \tag{35}$$

Fig. 3. Gaussian surface of revolution Eq. (35) given in Example 1.

For Forbes' approximation, we implement both, Q_m^{con} and Q_m^{bfs} polynomials [10, 12]. On the one hand, Forbes' approximation using Q_m^{con} polynomials [10] reads

$$F(r) \simeq \frac{cr^2}{1+\sqrt{1-\varepsilon c^2 r^2}} + u^4 \sum_{m=0}^{M} a_m Q_m^{\text{con}}(u^2), \tag{36}$$

where c and ε are, at this moment, free parameters, u is the normalized radial coordinate given by $u = r/r_{max}$, $r_{max} = L$ is the aperture size, and $Q_m^{\text{con}}(x) = P_m^{(0,4)}(2x-1)$ with $P_m^{(\alpha,\beta)}(x)$ the Jacobi polynomials of parameters $(\alpha, \beta) = (0, 4)$. Following Forbes' algorithm, the values of c and ε in the first term on the right hand side of Eq. (36) can be approximated by $c \simeq -0.398942$ and $\varepsilon \simeq -8.646945$. Then, the first term in Eq. (36) is a hyperbola. On the other hand, we consider Q_m^{bfs} polynomials [10, 12]. In this case

$$F(r) \simeq \frac{c_{\text{bfs}} r^2}{1+\sqrt{1-c_{\text{bfs}}^2 r^2}} + \frac{u^2(1-u^2)}{\sqrt{1-c_{\text{bfs}}^2 r^2}} \sum_{m=0}^{M} a_m Q_m^{\text{bfs}}(u^2), \tag{37}$$

where c_{bfs} is the curvature of the best-fit sphere, $r_{max} = L$ and $u = r/r_{max}$. Polynomials $Q_m^{\text{bfs}}(x)$ can be generated using a non-standard recurrence relation [12, 15] that involves a set of orthogonal polynomials. For this example, the best-fit curvature c_{bfs} can be approximated by -0.306394.

Finally, we consider our expansion defined by Eq. (19), with $\{q_n(r)\}_{n=0,1,...}$ the new quasi-orthonormal basis given in Eq. (18), $C_0 = \frac{1}{2}[2b^2 - s\frac{b^2}{a^2}L^2]^{1/2}$ and c_n defined in Eq. (11). As the

first order approximation is given by Eq. (14), in order to get a first close-fitting approximation, we must translate the original function, that is, $f(r) + b$, in such a way that both functions are coincident at $r = 0$. For this function, we consider $s = 1$, $a = b = 1/c_{\text{bfs}}$ in Eq. (14).

The fit error obtained with the three approximations Eq. (19) (proposed here), Eq. (36) (Forbes Q_m^{con} polynomials) and Eq. (37) (Forbes Q_m^{bfs} polynomials) for the same number of terms ($N = 7$, $M = 6$ and $M = 6$ respectively, that is, 8 terms for the three cases) and for $L = r_{max} = 1$, are compared in Fig. 4.

In Fig. 5 we compare the relative error provided by equations Eq. (19) (new basis proposed in Section 2) and Eqs. (36) and (37) (Forbes polynomials) in the L^2-norm for the same number of terms. Our basis functions provide lower errors than Forbes Q_m^{con} polynomials, but not with respect to Forbes Q_m^{bfs} polynomials.

Example 2. In order to check the accuracy of the approximation supplied by the basis $\{q_n^m(r, \theta)\}_{n,m(n)=0,1,2,...}$ for the freeform case given in Eq. (34), we implement the same example as in the 1D case, but now we consider the approximation problem in 2D,

$$f(r, \theta) = \frac{1}{\sqrt{2\pi}} \left(e^{-\frac{r^2}{2}} - 1 \right), \quad r \in [0, 1], \ \theta \in [0, 2\pi). \tag{38}$$

We approximate this function by using the new quasi-orthonomal basis $\{q_n^m(r, \theta)\}_{n,m(n)=0,1,...}$ given in Eq. (34) but choosing $q_0^0(r)$ in Eq. (14) the sphere with $s = 1$, $a = b = 1/c_{\text{bfs}}$ and $L = 1$, $q_0^0(r) = \sqrt{(c_{\text{bfs}})^{-2} - r^2}$:

$$f(r, \theta) \simeq \frac{c_0}{C_0} \sqrt{(c_{\text{bfs}})^{-2} - r^2} + \sum_{\substack{n,m(n)=0 \\ (n,m)\neq(0,0)}}^{N} c_n^m q_n^m(r, \theta), \tag{39}$$

with $C_0 = \left[\frac{\pi}{2} \left(2c_{\text{bfs}}^{-2} + 1 \right) \right]^{1/2}$ and c_n^m defined in Eq. (31).

Here we compare with the approximation obtained with Zernike polynomials (ZPs) [11]

$$f(r, \theta) \simeq \sum_{n=0}^{N} \sum_{m=-n}^{n} \tilde{c}_n^m Z_n^m(r, \theta), \quad \tilde{c}_n^m = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 Z_n^m(r, \theta) f(r, \theta) r dr d\theta, \tag{40}$$

and with the approximation obtained with Forbes polynomials (37). The results are similar to the ones obtained in Fig. 5 for the 1D case in comparison with Forbes polynomials. Our method compares favorably with ZPs for 2D.

We consider now the non-symmetric two-dimensional Gaussian type surface (see Fig. 6),

$$f(r, \theta) = 2 - e^{-r^2(\cos^2\theta + 2\sin^2\theta)}, \quad r \in [0, 1], \ \theta \in [0, 2\pi). \tag{41}$$

In Fig. 7 we compare the root mean square errors (RMSE) provided by Forbes in [14, Eq. (2.2)], [15, Eq. (1.1)],

$$f(r, \theta) \simeq \frac{cr^2}{1 + \sqrt{1 - c^2 r^2}} + \frac{1}{\sqrt{1 - c^2 r^2}} \left\{ u^2(1 - u^2) \sum_{n=0}^{N} a_n^0 Q_n^0(u^2) \right.$$
$$\left. + \sum_{m=1}^{M} u^m \sum_{n=0}^{N} [a_n^m \cos m\theta + b_n^m \sin m\theta] Q_n^m(u^2) \right\}, \quad u = \frac{r}{r_{max}}, \tag{42}$$
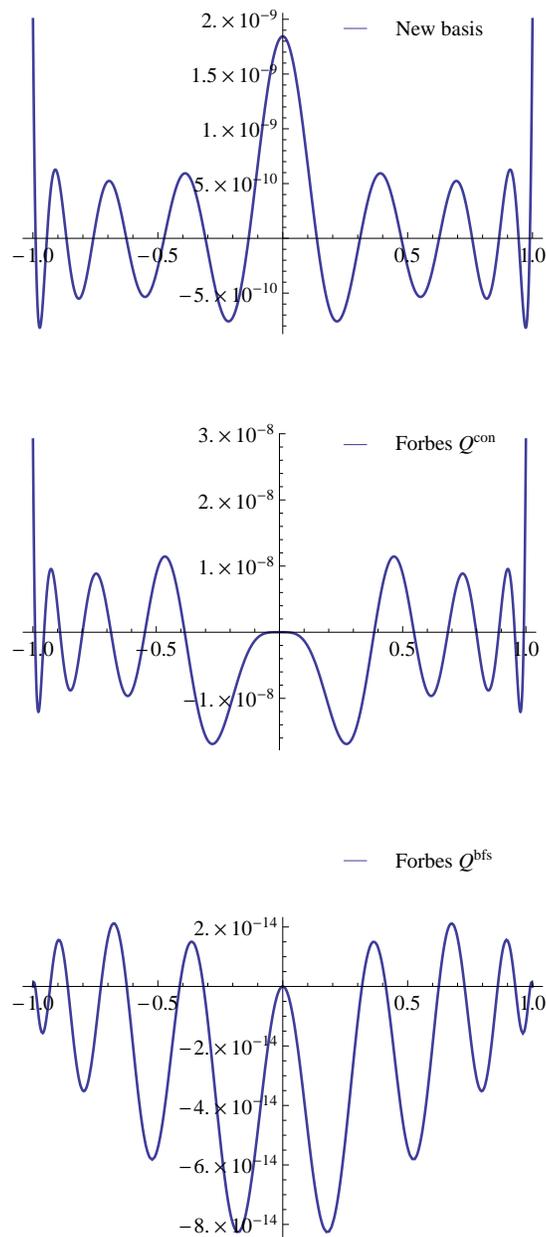
Fig. 4. Plots of the fit error obtained using our basis functions and Forbes polynomials, Q_m^{con} and Q_m^{bfs}, for the same number of terms, eight, for the three cases.

with $c \simeq 0.261082$, $r_{max} = 1$, and the new quasi-orthonomal basis $\{q_n^m(r,\theta)\}_{n,m(n)=0,1,\dots}$ given in Eq. (34), but choosing $q_0^0(r)$ the portion of a hyperboloid with semiaxes a and b like Eq. (14) with $s = -1$ and $L = 1$, $q_0^0(r) = b\sqrt{1 + r^2/a^2}$. The least square fitting provides the values
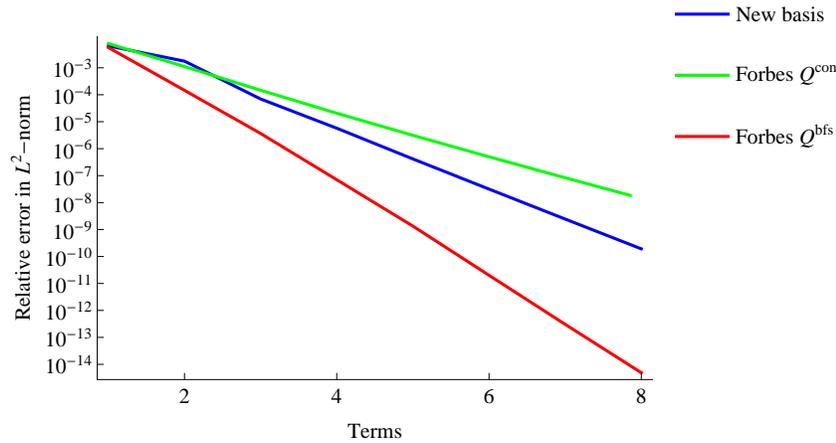
Fig. 5. Relative errors in the approximation of the function Eq. (35) using our approximation (see Eq. (19)) and Forbes' approximations (see Eqs. (36) and (37)) in the L^2-norm for the same number of terms.
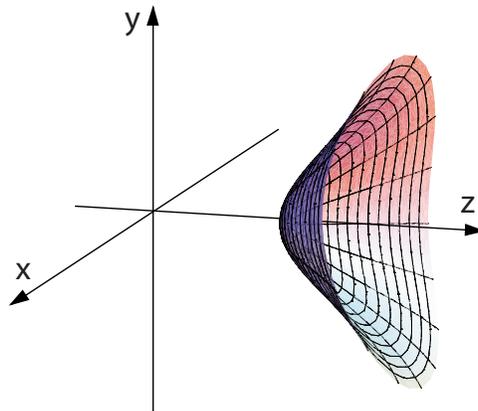

Fig. 6. Non-symmetric elliptical gaussian surface given in Eq. (41).

$a \simeq 0.653754$ and $b \simeq 1.02225$. Moreover, using the idea given by Forbes, we also approximate the function using the new quasi-orthonomal basis in Eq. (34) where now $q_0^0(r)$ is the best-fit sphere, and then $s = 1$, and $a = b = (0.261082)^{-1}$. We also compare with ZPs Eq. (40). The comparison is given using the number of terms (calculated coefficients) where we take the best RMSE given by Forbes approximation (42) for different values of N and $M \neq 0$ up to $2N + M = 10$. We also compare with the ZPs approximation. Our basis functions provide lower errors in comparison with Forbes approximation, and the difference decreases with the number of terms considered. In comparison with ZPs approximation, our basis funcions provide lower errors up to 10 terms, and ZPs provides lower errors from 15 terms, and also the difference decreases with the number of terms considered.
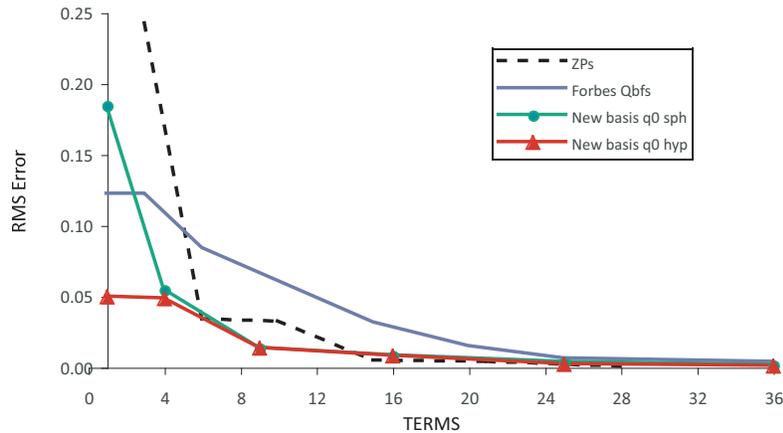
Fig. 7. RMSE in the approximation of the function Eq. (41) provided by the different approximations using the number of terms.

5. Discussion and conclusions

In summary, we developed a general, rigorous and powerful framework to obtain orthogonal systems to represent freeform optical surfaces. The method consists of first selecting the desired first basis function q_0, a sphere or conicoid in our case, and then choosing an appropriate orthogonal system on the desired domain or support. For the case of rotational symmetry we considered the interval $[0,1]$ and the unit disk in general. For these $1D$ and $2D$ domains we selected Legendre polynomials and spherical harmonics respectively, as the initial systems. Then the method consists of finding a change of variables φ which transforms these systems $\{p_n\}_{n=0,1,2,...}$ into another orthogonal system $\{q_n\}_{n=0,1,2,...}$ in which the first function q_0 is the conicoid that we chose. Note that this theoretical framework is powerful enough to obtain orthogonal systems for different expressions of the conicoids: canonic expression of Eq. (14), or the sag of Eq. (20) commonly used in optical design and testing. We also included the semidiameter L of the surface explicitly, which permits to avoid the need of normalizing the radial coordinate r, that is necessary when using ZPs or similar systems. For the implementation and examples we used the canonical expression for the sake of simplicity, since it allows more compact expressions as compared to the sag equation.

Compared to other systems used in optics, this work is more general in different aspects. Probably, ZPs form the most widely used orthogonal system in optics. It is the standard basis for representing wave aberration (or optical path difference), and it is also used for representing aspherical and freeform surfaces. This system includes the paraboloid of zero mean (defocus term $Z_2^0 = 2r^2 - 1$, $r \in [0,1]$), but it requires to apply a higher order expansion to approximate spheres or other conicoids. In fact, it is common to use ZPs, Z_n^m to represent only the aspherical terms in the form $z = C + A$ where C is the conicoid and $A = \sum_{n,m} a_n^m Z_n^m$; a_n^m are the Zernike expansion coefficients. The approach developed by Forbes [10] (actually he proposed two different approaches for mild and strong aspheres respectively) is similar in the sense that it splits the surface sag into the same conic and aspherical parts $z = C + A$, but applying a smarter *ad hoc* expansion $A = r^4 \sum_n f_n Q_n(r^2)$ where Q_n are orthogonal. Among other advantages, extracting the common factor r^4 permits to reduce the order (and number) of Forbes polynomials Q_n

needed to approximate A. Nevertheless, the main difference of our approach is that the conicoid is the first element of the system and hence it is orthogonal with the rest. In this sense we propose a fully orthogonal system as opposed to these partially orthogonal systems applied only to the aspherical part A.

For the numerical implementation and testing we have chosen a Gaussian as it is expected to require higher order expansions (high values of n) to obtain reasonable accuracies. Our results with $1D$ (Legendre polynomials) and $2D$ (spherical harmonics) implementations provide results which compare favorably with Forbes polynomials and ZPs in the general freeform two-dimensional case, whereas Forbes Q^{bfs} polynomials show a better performance in the one-dimensional rotationally-symmetric case. We want to remark that the cases and implementations presented here are particular examples of a much more general theoretical framework. In fact, it may be possible to consider a wide variety and types of initial surface q_0 as well as to use different types of polynomials, etc. The main restriction to obtain systems with analytical expressions is that the integral equations, Eqs. (4) and (5), that we have to solve to obtain the change of variables, must have an analytical solution. Nevertheless, we hope that the examples of conicoid-based orthogonal systems presented here are useful and general enough for designing, manufacturing and testing freeform optical surfaces.

Acknowledgments