

Optimal local shape description for rotationally non-symmetric optical surface design and analysis

Ozan Cakmakci[◇], Brendan Moore[‡], Hassan Foroosh[‡], Jannick P. Rolland^{◇,‡}

[◇] College of Optics, Center for Research and Education in Optics and Lasers (CREOL),

[‡]School of Electrical Engineering and Computer Science

University of Central Florida,

Orlando, FL, 32816

ozan.cakmakci@gmail.com

Abstract: A local optical surface representation as a sum of basis functions is proposed and implemented. Specifically, we investigate the use of linear combination of Gaussians. The proposed approach is a local descriptor of shape and we show how such surfaces are optimized to represent rotationally non-symmetric surfaces as well as rotationally symmetric surfaces. As an optical design example, a single surface off-axis mirror with multiple fields is optimized, analyzed, and compared to existing shape descriptors. For the specific case of the single surface off-axis magnifier with a 3 mm pupil, >15mm eye relief, 24 degree diagonal full field of view, we found the linear combination of Gaussians surface to yield an 18.5% gain in the average MTF across 17 field points compared to a Zernike polynomial up to and including 10th order. The sum of local basis representation is not limited to circular apertures.

© 2008 Optical Society of America

OCIS codes: (080.2740) Geometric optical design; (220.4830) Optical systems design; (230.4040) Mirrors; (220.1250) Aspherics.

References and links

1. G.D. Wasserman and E. Wolf, "On the theory of aplanatic aspheric systems," *Proc. Phys. Soc. B* **62**, 2–8 (1949).
2. J. M. Rodgers, "Nonstandard representations of aspheric surfaces in optical design," Ph.D. Thesis, University of Arizona (1984).
3. D. Knapp, "Conformal Optical Design," Ph.D. Thesis, University of Arizona (2002).
4. J.E. Stacy, "Design of Unobscured Reflective Optical Systems with General Surfaces," Ph.D. Thesis, University of Arizona (1983).
5. T. Davenport, "Creation of a uniform circular illuminance distribution using faceted reflective NURBS", Ph.D. Thesis, University of Arizona (2002).
6. H. Chase, "Optical design with rotationally symmetric NURBS," *Proc. SPIE* **4832**, 10-24 (2002).
7. S. Lerner, "Optical Design Using Novel Aspheric Surfaces," Ph.D. Thesis, University of Arizona (2003).
8. M.D. Buhmann, *Radial Basis Functions: Theory and Implementations* (Cambridge University Press, 2003).
9. G.W. Forbes, "Shape specification for axially symmetric optical surfaces," *Opt. Express* (**15**), 5218-5226 (2007).
10. O. Cakmakci and J. Rolland, "Head-Worn Displays: A Review," *J. Disp. Tech.* (**2**), 199-216 (2006).
11. O. Cakmakci and J. Rolland, "Dual-Element Off-Axis Near-Eye Optical Magnifier," *Opt. Lett.* (**32**), 1363-1365 (2007).
12. J. M. Rodgers and K. P. Thompson, "Benefits of Freeform Mirror Surfaces in Optical Design," *Proceedings of the American Society of Precision Engineering 2004 Winter Topical Meeting on Freeform Optics*, **31**, 73-78 (2004).

1. Introduction

Several optical designs require aspheric or freeform surfaces to achieve compact and lightweight solutions. The traditional description of an aspheric surface in the optical design community is accepted to be

$$z(r) = \frac{cr^2}{1 + \sqrt{1 - (1+k)c^2r^2}} + \sum_{i=0}^n a_i r^{2i+4}, \quad (1)$$

where c represents the curvature, r is the $\sqrt{x^2 + y^2}$, and k is the conic constant.

Michael Rodgers considered alternative aspheric representations to the standard method of adding a power series to a base conic[2]. Rodgers' thesis discussed nonpolynomial basis functions added to a conic in order to yield perfect axial imagery. Examples of nonstandard aspheric functions are given in Table 2.1 of the Rodgers' thesis to be hyperbolic cosine, logarithm, secant, inverse sine, tangent and a Gaussian. Convergence properties of such alternative representations were studied. Rodgers' dissertation considered rotationally symmetric systems only. Rodgers' thesis studied global and compact representations for optical surfaces. Rodgers' thesis includes noniterative as well as iterative approaches to generating aspheric profiles. The noniterative surface generation technique was based on Wasserman-Wolf, which is briefly summarized in the following paragraph. Imaging and energy redistribution examples of 2, 3, and 4 mirror reflective designs were given. These examples illustrated that basis functions like the hyperbolic cosine, logarithm, and the secant can represent a surface shape significantly better than the same number of power series terms.

Wasserman-Wolf is a non-iterative technique generating a pair of differential equations for stigmatic imaging of a point object[1]. Coma correction is obtained by satisfying the sine condition. Essentially, Wasserman-Wolf generates a slope field representing the tangents on each point of each aspheric surface, the slope field is integrated using Runge-Kutta or Adam's method to get the surface profile. The original paper deals with the case of rotationally symmetric aspheres. David Knapp's thesis work generalized the Wasserman-Wolf technique removing the axial symmetry assumption in the original work[3].

Stacy explored splines as general surfaces in the design of unobscured reflective optical systems[4]. The basic aim in Stacy's work was to compare spline based designs to the classically designed Galileo narrow angle camera. Stacy found that the spline designs allowed wider field angles and higher transmission than the Galileo with lower but still acceptable image quality. Chase discussed the application of parametric curves such as Non-uniform Rational B-Splines (NURBS) to the description of rotationally symmetric asphere design. Chase compared the spherical aberration correction of a NURBS surface against an even and odd polynomial in a Cassegrain system[6]. Davenport investigated NURBS as a tool for creating an incoherent uniform circular illuminance distribution on a target plane[5]. Benefits of freeform surfaces have been summarized by Rodgers and Thompson as "greater control of the location in the field of nodes in the aberration field, and potentially the larger number of nodes in the field"[12].

Scott Lerner introduced a novel explicit superconic surface[7]. Lerner also explored parametrically defined optical surfaces and implicitly defined optical surfaces. A truncated parametric Taylor surface and an xyz-polynomial were shown to be general surface descriptions. Representations were compared for ray tracing speed, optimization complexity, the ability to correct highly aspheric wavefronts, and the ability to represent steeply sloped surfaces. Specifically, Table 5.2 in [7] lists a superconic, explicit superconic, truncated parametric Taylor, and an xyz polynomial as the surfaces that were compared.

Greg Forbes recently proposed a sum of Jacobi polynomials to represent axisymmetrical aspheres[9]. Forbes emphasizes the use of Jacobi polynomials which are global and orthogo-

nal. Forbes representation has the key property that the mean square slope of the normal departure from a best-fit sphere is related to the sum of squares of individual coefficients of Jacobi polynomials. This property facilitates the enforcement of fabrication constraints.

Our emphasis in this paper is on a local representation of shape that is suitable for the optimization of rotationally non-symmetric freeform systems (example 2 in this paper) as well as rotationally symmetric (example 1 in this paper) systems. At this point in our development, we have not explicitly constrained the slopes for manufacturability. However, we provide an interferogram of the surface with the base sphere subtracted to show that the surface is well behaved even with unconstrained optimization with respect to the slopes on the surface.

2. Optical Surface Representation using Local Basis Functions

An optical surface can be represented as a sum of basis functions

$$z(x, y) = \sum \phi_i(x, y) w_i, \quad (2)$$

where ϕ_i are the basis functions and w_i are the coefficients. In eqn.(2) we have shown the surface to be a function of two variables for clarity. Alternatively, we will use the vector \mathbf{x} to represent these two dimensions. In this paper, we propose and explore the use of 2D Gaussians as basis functions to describe optical surfaces. A 2D Gaussian is described by

$$\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathbf{e}^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad (3)$$

where $\boldsymbol{\mu}$ represents the mean vector and $\boldsymbol{\Sigma}$ represents the covariance matrix. Gaussians have several desirable properties in the context of optical design. First, Gaussians can be considered to be local functions since the value of a Gaussian outside of 3 sigma is small (<0.011 for a zero mean, unit variance Gaussian). Second, Gaussians are smooth (C^∞) having derivatives of all orders. Third, the Fourier transform of a Gaussian is a Gaussian which gives us an analytical description of the Power Spectral Density (PSD) function that is represented by a linear combination of Gaussians. Figure 1 illustrates the concept of a linear combination of Gaussian basis functions summing to approximate a sphere along with 1-dimensional slices through the original and the fit.

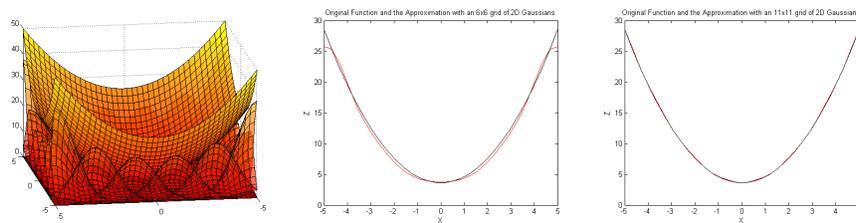


Fig. 1. Illustration of a linear combination of Gaussian basis functions approximating a sphere. (Left) Original sphere and the weighted basis functions are shown. The 2D Gaussians are spaced uniformly with means centered on a 6x6 grid. Each 2D Gaussian has a variance of unity along the x and y dimensions. The weighted Gaussian basis functions shown underneath the sphere are found through least squares. (Center) 1D slice through the original function (black) and the approximation with a 6x6 grid (red). (Right) 1D slice through the original function (black) and the approximation with a 11x11 grid (red). We observe that the fit error reduces when the number of basis functions is increased.

Equation (2) can be written in vector form as

$$\Phi \mathbf{w} = \mathbf{Z}, \quad (4)$$

where the $m \times n$ Φ matrix contains the basis functions, w is a vector of weights, and Z is the resulting surface. Further theoretical background into this framework is given by Buhmann[8].

3. Optimization Procedure

In this section we will describe the optimization procedure. The first step was to set the grid size. Optimal setting of the grid size for a particular problem requires further study and will not be considered in this paper. For the magnifier example given below, we have found a 17x17 grid to perform well when compared against relatively general surface representations such as x-y polynomials or Zernike polynomials. The second step was to initialize the starting point. Our raytrace software requires a base sphere in order to perform paraxial image calculations. Therefore, we modified the description given in equation 3 by adding a base conic to the sum of basis representation as follows

$$z(x,y) = \frac{c(x^2 + y^2)}{1 + \sqrt{1 - (1+k)c^2(x^2 + y^2)}} + \sum_i \phi_i(x,y)w_i, \quad (5)$$

where c is the curvature in the x and y directions and k is the conic constant. The third step was the initialization of the starting point. During optimization we set the weights of the basis functions to zero, starting with the base conic. The fourth step was the construction of the Φ matrix. Each column of the Φ matrix contains a vectorized form of equation 3 and is written as

$$\Phi = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \dots & \phi_n(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \dots & \phi_n(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_m) & \phi_1(\mathbf{x}_m) & \dots & \phi_n(\mathbf{x}_m) \end{pmatrix}.$$

Given a number of basis functions, we divided the aperture into x_{num} pieces in the x -dimension and y_{num} pieces in the y -dimension. The number of columns in the matrix was set by the product of x_{num} and y_{num} . The number of rows in the Φ matrix controlled the spatial resolution of the Gaussians and was set by the user. We make a rectangular aperture assumption in this case. However, the sum of basis representation accommodates any aperture shape since the Gaussians can be moved spatially using their means. We divided the aperture diameter into x_{num} pieces in the x -dimension and placed each x -mean $1/x_{num}$ apart from each other. Similarly, we divided the aperture into y_{num} pieces in the y -dimension and placed each y -mean $1/y_{num}$ apart from each other. The variances in the covariance matrix were set to 1. The weights of 400 Gaussians in the illustration given in Fig. 1 were found through least squares by

$$w = (\Phi^T \Phi)^{-1} \Phi^T Z. \quad (6)$$

Unlike the function fitting example given in Fig. 1 where we have a surface Z (sphere) to fit, in the context of the optical design problem, the surface Z is unknown a priori, and the goal of an iterative optimizer is to adjust the weights w in equation 4 with the goal of reaching a minimum of the merit function given a starting point. The fifth step is an optional step to represent a starting point only with the sum of basis, without the base conic. The intention with step 5 is to allow the exploration of alternative optimization techniques (for example using the MATLAB optimization toolbox) such as the trust region dogleg, Gauss-Newton or simplex. We remind that the addition of the base conic was only required for the paraxial calculations in the raytrace code. Alternative optimization environments could utilize step 5 as the preferred way to experiment with this surface representation. The sixth step is the construction or choice of the error function. We used the transverse error in the image plane, which is the sum of squares of the deviations of the rays from their respective reference wavelength chief rays, as our merit

function. The seventh step was to choose an optimization technique and to optimize the error function. The results reported in this paper are based on the damped least squares algorithm.

4. Optical System Design Examples

A user defined surface type 1 has been implemented in C for Code V as a dynamically linked library to test the surface representation. A full description of user defined type 1 surfaces is provided in the Code V documentation. Code V interacts with the surface one point at a time meaning that Code V will ask the sag of the surface for a specific x , y and z point. Therefore, the Φ matrix reduces to a row vector and the sag calculation becomes a dot product operation with the weights. The first elementary test case was to optimize the system with a single on-axis field (object at infinity) and check whether we get a parabola represented by the sum of basis. The user defined surface type 1 was setup for this test case with a 6x6 grid (112 user defined coefficients). The first coefficient represented the x curvature of the base sphere, the second coefficient represented the conic constant in x , and the third coefficient represented the conic constant in y . The fourth coefficient represented the aperture size. The next 36 coefficients were the weights of the Gaussians (w_i). The last 72 coefficients contained the x and y means of the Gaussians. During the optimization runs presented in this paper, the means were frozen (not variable) during optimization runs. Means were included in the user coefficient array solely for diagnostics purposes to make sure that the gridding functions were working properly. The variables in this elementary test case included the x and y curvatures, image plane defocus, and the 36 weights of the 6x6 Gaussians across the aperture. The conic constant was set to zero and not varied during optimization. The optimization converged in a few cycles (< 10) to a parabola with a Strehl ratio of 1.

The second test case we will discuss is an off-axis magnifier. Applications such as head-worn displays demand compact and lightweight optical solutions [9]. Recently, we designed single and dual-element optical magnifiers having free-form surfaces described with x - y polynomials for head-worn display applications, and a dual-element design was fabricated [10]. An ideal solution for a head-worn display would be a single surface mirror design. A single surface mirror does not have dispersion, therefore, color correction is not required. A single surface mirror can be made see-through by machining the appropriate surface shape on the opposite side to form a zero power shell. In the second test case, we address the question of “what is the optimal shape for a single surface mirror constrained in an off-axis magnifier configuration?”.

We designed four systems to address the question of optimal shape for the off-axis magnifier problem. A 10th order anamorphic sphere, an x - y polynomial, a 10th order Zernike polynomial and a linear combination of Gaussians were compared. Each system under comparison had a > 15 mm eye clearance, 3mm pupil, 24 degree diagonal full field of view (9.56° x semi-field and 7.2° y semi-field), and a 14.7° mirror tilt angle. Each system was optimized with the minimum set of constraints such as the effective focal length (14.25mm), real ray based distortion constraints, and the field weights. Each system had 17 field points defined. The variables in each system included the surface coefficients, image plane defocus and tilt. Distance from the pupil to the mirror vertex was 16.9mm. Distance from the vertex of the mirror to the image plane was kept around 13mm since it is undesirable to have a microdisplay physically close to a human eye. The image plane has a rectangular aperture with a size of 4.8mm by 3.6mm in the x and y dimensions, respectively. The image plane needs to stay clear out of the ray path between the pupil and the mirror. 140 rays across the pupil were traced in each system during optimization. Figure 2 (left) shows the layout of the optimal off-axis magnifier. Modulation Transfer Function (MTF), evaluated at $\lambda = 550nm$ (no dispersion), for the optimized linear combination of Gaussians surface is shown in Fig. 2 (center). Interferogram of the surface with the base sphere subtracted is shown in Fig. 2 (right). Distortion characteristics exhibit similar behavior with

each design having a maximum of about 3 to 4%. Table 1 shows a comparison of the surface representation proposed and implemented in this paper against an anamorphic asphere, Zernike polynomial up to and including 10th order, an x-y polynomial up to and including 10th order (good balancing achieved with order 5) with the maximum distortion and the average MTF across 17 field points as the comparison metrics. The sum of local basis representation proposed in this paper achieves the highest MTF performance averaged across 17 field points by 18.5% in the field with an acceptable level of maximum distortion, among the functions that were compared. Zernike optimization has been confirmed independently in Zemax and Code-V and the results agree to within 1.5% of the reported MTF value in Table 1.

Table 1. Comparison of the transverse error function value and 17 average tangential and sagittal MTF values between an anamorphic asphere, x-y polynomial, Zernike polynomial, and a linear combination of Gaussians surface type.

Surface Type	Average MTF (17 fields at 23cyc/mm)	Max. Distortion
Anamorphic Asphere	26.5%	3.8%
x-y polynomial	43.6%	2.65%
Zernike Polynomial	42%	3.74%
Linear combination of Gaussians	60.5%	3.6%

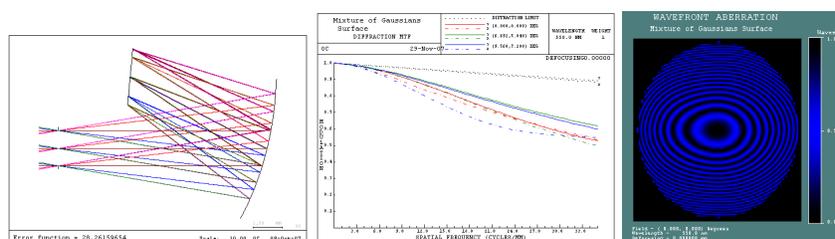


Fig. 2. (Left) Optical layout of the off-axis magnifier. (Center) MTF evaluated on-axis, 0.7 in the field, and at the maximum field. (Right) Interferogram of the surface represented with a linear combination of Gaussians

5. Conclusion and Future Work

A sum of local basis representation, specifically the linear combination of Gaussians, was proposed and implemented to represent and optimize freeform optical surface shapes. A rotationally symmetric and a rotationally non-symmetric example were given. A single surface off-axis mirror with multiple fields was optimized with a linear combination of Gaussian surface, analyzed, and compared to existing shape descriptors such as an anamorphic asphere, x-y polynomial, and a Zernike polynomial. Even though Zernike polynomials are an orthogonal and a complete set of basis over the unit circle and they can be orthogonalized for rectangular or hexagonal pupils using the Gram-Schmidt process, taking into account practical considerations, such as optimization time and the maximum number of variables, for the specific case of the single off-axis magnifier with a 3 mm pupil, >15mm eye relief, 24 degree diagonal full field of view, we found the linear combination of Gaussians surface to yield an 18.5% gain in the average MTF across 17 field points compared to a Zernike polynomial up to and including 10th order. The largest grid size we have been able to explore has been 17x17 because currently our raytrace code has a maximum number of 300 variables available for optimization at the time of this writing.

Acknowledgements

We thank the Link Foundation and this work has been supported in part by National Science Foundation (NSF) IIS/HCI 03-07189. We thank Optical Research Associates for the educational license of Code V and excellent technical support.