Effective Characterization of the Nominal Shape of Aspheric Optics

Greg Forbes
QED Technologies Inc., 1040 University Avenue, Rochester, NY 14607, USA

Abstract: Advances in versatile fabrication and metrology systems enable the wider penetration of optical aspheres. Starting from the hands of designers, a more effective specification of shape can support these developments in surprising ways.

Keywords: aspheric optics, optical design, design for manufacture

1. Introduction

A few clearly desirable characteristics of any scheme for specifying aspheric surface shape are that it

i) gives the required accuracy with a minimal number of decimal digits,

ii) facilitates simple estimates of difficulty for fabrication and testing, and

iii) supports direct options for specifying tolerance requirements.

The conventional specification expresses the sag as a combination of a conic section and a sum of monomial powers, namely[1]

$$z(\rho) = \frac{c\rho^2}{1+\sqrt{1-(1+\kappa)c^2\rho^2}} + \sum_{m=0}^{M} A_m \rho^{2m+4}, \qquad (1)$$

where c is the axial curvature and κ the conic constant. As is becoming well known through firsthand struggles across the industry, this approach scores poorly on each of the three measures proposed above. Some of its weaknesses are demonstrated in this talk, and a more promising option is explained.

Escalating requirements for optical performance have been matched by innovations in tools for design and fabrication. One of the natural outcomes, however, is that it is increasingly common to see a significant number of terms in the sum in Eq.(1). Five or six terms are not unusual now (i.e. $M \approx 5$), and recent patents contain surfaces with up to ten terms[2]. In fact, we have seen designs that involve as many as eighteen. When only two or three terms are used, Eq.(1) is workable and it has served the industry well. For current applications, however, it is increasingly problematic.

2. Degeneracy

The parameters in Eq.(1) are said to be degenerate if they offer multiple options to characterise what is effectively a single shape. Eq.(1) is found to approach degeneracy in several ways: If $c = 0$, κ can obviously be given any value without changing the resulting shape at all. It can also be seen that, for small c, changes to κ can effectively be undone by adjusting A_0. [When, as is sometimes done, the sum starts at ρ^2 instead of ρ^4, c then becomes similarly degenerate with that second-order term.] On a moment's reflection, it can be appreciated that such degeneracies can cause wasted digits, obscure the part's shape, and mean that tolerancing the parameters directly is not an option. That is, degeneracy leads to trouble with each of the desirable criteria listed in the Introduction.

These are minor issues, however, compared to the potentially catastrophic degeneracy associated with the linear parameters, namely A_m. Degeneracy for these means that a combination of a set of terms with non-zero coefficients can lead to an end result that is much smaller than the individual terms themselves. To clarify this point, consider a case where the full aperture is 2.0 units in diameter, so $0 < \rho < 1$. In this case, when plotted across the aperture, each of the terms in the sum in Eq.(1) runs between zero and A_m in value. The key question therefore is "Can the result of this sum be much smaller than the individual terms?". For the case of ten terms, a compelling answer is given by the plot of the function presented in Figure 1.

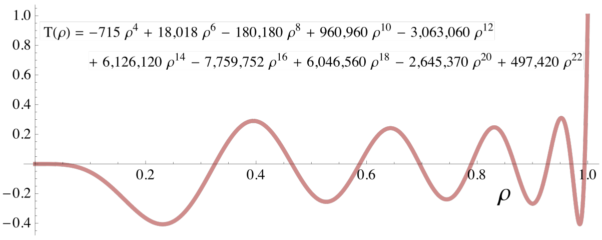

Figure 1. Approaching catastrophic cancellation.

The magnitudes of the individual terms in the superposition plotted above approach ten million, but their sum never exceeds unity. That is, six or seven digits are lost through cancellation. Consider using such a scheme to describe functions that have rms values of about unity while requiring the result to be accurate to ±0.001, say. It is now evident that even though the function is being specified with a relative precision of about one part in a thousand, many of the coefficients may need to be specified with ten digits (seven above and three below the decimal point). This means that we may need as many as one hundred digits when using ten terms at this level of accuracy.

Importantly, these observations reveal that the magnitude of any one coefficient carries almost no information on its own. To appreciate this, notice that if we divide $T(\rho)$ of Fig.1 by 10^3, the result remains below the nominated fit tolerance of ±0.001. That is, for our purposes, the result is effectively zero. The magnitude of the individual (scaled) coefficients still approaches 10^4, however. For any particular asphere, it follows that these coefficients can be changed by this amount with only negligible change to its shape. This signals one of the typical problems encountered in such work: When, for example, we are specifying a shape to within about one part in a thousand of its PV range (think some number of nm's in some μm's of aspheric departure), coordinated changes can be made to these coefficients at a level that is thousands of times larger than the function itself while having effectively zero impact. Obviously these coefficient values should not be toleranced individually.

When the number of terms approaches twenty, as we have started to see in real designs, such problems are even further exacerbated. For example, the relative waste is then doubled which makes the cancellation catastrophic: all fifteen or so digits in a double-precision floating-point number can be lost through cancellation. A new option is evidently vital.

Degeneracy is the root of these difficulties. The natural solution therefore is to adopt a superposition of elements that are orthogonal. That is, the elements must be defined so that, in some average or mean-square sense, they cannot cancel one another. One such average that is relevant in our context is the mean square value of just the sum in Eq.(1). If, as in Fourier series, the product of two different elements averages to zero, the mean square value of a sum of them is simply related to a sum of the squares of the coefficients, recall Parseval's theorem[3]. This property ensures that, in an rms sense, *the elements in such a sum can never cancel one another:* making any one coefficient larger in magnitude necessarily increases the function's rms value. These coefficients then form a familiar *spectral decomposition where each term has a direct meaning on its own*. In this way, the functions considered above that needed about a hundred digits can be characterized with no more than four or so digits per term, hence under half the number of digits. And this is just one of the significant gains to be had.

Just such a linear combination of a particular set of orthogonal polynomials has been proposed as a minor modification[4] of the sum in Eq.(1). [In fact, the curve plotted in Fig. 1 is precisely the tenth member of that new set.] This talk concentrates on a second, more interesting option that was also introduced in Ref.4. It is more directly tailored to considerations related to fabrication and design for manufacture.

3. An orthogonal basis

Whether an asphere's shape is suitable for various cost-effective fabrication and metrology processes can be coupled to the rate of change of its departure from a best-fit sphere. With this in mind, another alternative to Eq.(1) can be written as[4]

$$z(\rho) = \frac{c_{\text{bfs}}\,\rho^2}{1+\sqrt{1-c_{\text{bfs}}^2\rho^2}} + \frac{s\,(1-s)}{\sqrt{1-c_{\text{bfs}}^2\rho_{\max}^2\,s}} \sum_{m=0}^{M} a_m\, Q_m^{\text{bfs}}(s), \quad (2)$$

where c_{bfs} is the curvature of the best-fit sphere, $\rho_{\max}$ is one half of the CA, $s \equiv \rho^2/\rho_{\max}^2$, and $Q_m^{\text{bfs}}(s)$ is a polynomial of order m in s. Notice that κ is no longer present. All other degeneracy has been removed by choosing the elements of $Q_m^{\text{bfs}}(s)$ to be orthogonal in that the mean square slope of the normal aspheric departure is a constant times the sum over m of a_m^2.

Designing under a constraint on this simple sum of squares can ensure full-aperture interferometric testability[5]. A similar process can be used when the aspheres can be tested with stitched interferometry[6]. In the latter case, the mean square difference between the local principal curvatures across the part's aperture is constrained, and a similar option is just what is needed to facilitate manufacturing steps such as pad polishing.

To point out advantages, a variety of sample lenses will be shown in the talk in both the new and old representations. These include cases where Eq.(1) used both even and odd powers. Others are shown in Ref.7. To give some idea, two modest examples with $\kappa = 0$ are presented in Table 1. The difference between the new and old sags is at the few nm level. Notice that, while the A_m values give no easy indication of shape, the stronger asphere (with over 100μm of departure in sag) is readily identified by reference to the a_0 values. The new representations have just one third of the number of digits and it becomes self-evident when terms can be dropped, as was so for both cases here.

Table 1. Old and new representations of two sample lenses.

US Pat#:	6,646,718		6,646,718	
Lens ID:	#L624 (~6μm)		#L625 (~100μm)	
m	A_m (ord 4…16)	a_m (nm)	A_m (ord 4…16)	a_m (nm)
0	3.02835805E-10	**-45,901**	-3.99248993E-10	**-349,932**
1	-2.40484062E-14	**30,708**	5.79276562E-14	**-6,442**
2	-3.22339189E-19	**-16,955**	3.53241478E-18	**46,321**
3	1.64516979E-22	**9,035**	-4.57872308E-23	**-6,940**
4	-8.51268614E-27	**-1,634**	-6.29695208E-27	**-591**
5	2.09276792E-31	**115**	1.57844931E-31	**74**
6	-4.74605669E-36	dropped	-2.19266130E-36	dropped
digit count:	**77**	**26**	**77**	**24**
rad of curv:	-251.154571510		-193.582989843	
CA:	269.706		268.202	

4. Conclusions

As described in the talk, orthogonal bases offer many benefits: significantly fewer digits are needed, deleting terms becomes straightforward in design (all the other terms then remain unchanged), simple constraints exist to support design for manufacture, the coefficients have direct interpretations, and it is possible to tolerance them individually. It's a change whose time has come.

5. References

1) D.S. Goodman: *Handbook of Optics,* (Optical Society of America, McGraw Hill, 1995) Sec. 1.10
2) See, for example, US Patent #7,315,423
3) See, for example, http://mathworld.wolfram.com/ParsevalsTheorem.html
4) G.W. Forbes: Opt. Exp. **15**, (2007) 5218
5) G.W. Forbes and C.P. Brophy: SPIE Proceedings **7100**, (2008) 710002
6) G.W. Forbes and C.P. Brophy: SPIE Optifab (2009) TD06-25 (1)
7) G.W. Forbes: OSA FiO (2009) TD06-25 (1)

forbes@qedmrf.com www.qedmrf.com