Research Article | Vol. 27, No. 5 | 4 Mar 2019 | OPTICS EXPRESS 6129
Optics EXPRESS

On-the-fly surface manufacturability constraints for freeform optical design enabled by orthogonal polynomials

NICK TAKAKI,[1,*] AARON BAUER,[1] AND JANNICK P. ROLLAND[1]

[1]University of Rochester, 275 Hutchison Rd, Rochester, NY 14620, USA
[*]ntakaki@ur.rochester.edu

Abstract: When leveraging orthogonal polynomials for describing freeform optics, designers typically focus on the computational efficiency of convergence and the optical performance of the resulting designs. However, to physically realize these designs, the freeform surfaces need to be fabricated and tested. An optimization constraint is described that allows on-the-fly calculation and constraint of manufacturability estimates for freeform surfaces, namely peak-to-valley sag departure and maximum gradient normal departure. This constraint's construction is demonstrated in general for orthogonal polynomials, and in particular for both Zernike polynomials and Forbes 2D-Q polynomials. Lastly, this optimization constraint's impact during design is shown via two design studies: a redesign of a published unobscured three-mirror telescope in the ball geometry for use in LWIR imaging and a freeform prism combiner for use in AR/VR applications. It is shown that using the optimization penalty with a fixed number of coefficients enables an improvement in manufacturability in exchange for a tradeoff in optical performance. It is further shown that, when the number of coefficients is increased in conjunction with the optimization penalty, manufacturability estimates can be improved without sacrificing optical performance.

1. Introduction

When seeking to correct the aberrations that arise in non-axisymmetric optical systems, designers often reach for rotationally-variant "freeform" optical surfaces. One of the first choices they face is the selection of mathematical freeform surface description.

One accepted method is to characterize a freeform optic as a base surface (e.g. a sphere, conic, or biconic) plus sag departure described with a linear combination of polynomial terms [1]. If the base surface is a sphere with curvature c and sag departure is described by arbitrary polynomials $\{P_n\}$, then the freeform surface sag z is given as

$$z = f(\rho,\theta) = \frac{c\rho^2}{1+\sqrt{1-c^2\rho^2}} + \sum_{n=0}^{N} s_n P_n(u,\theta). \tag{1.1}$$

Without loss of generality, the polynomials P_n are expressed in polar coordinates, s_n is the weight for P_n, and $u = \rho/\rho_{\max}$ is the normalized radial coordinate. Further, while non-orthogonal (or even non-polynomial) surface descriptions exist, the methods of this paper leverage orthogonality. Thus, it is required in this work that $\{P_n\}$ is orthogonal.

Even with the requirement of orthogonality, there is an abundance of options when choosing a polynomial set $\{P_n\}$. Zernike polynomials, which are widely known in many applications in optics [2,3], are also used to characterize freeform surfaces [4–7]. Cartesian products of Chebyshev and Legendre polynomials, each a set of orthogonal polynomials with a long history in mathematics, can also be used to describe freeform surfaces [8,9]. Other orthogonal polynomials have been custom-introduced specifically for freeform surface

#352984 https://doi.org/10.1364/OE.27.006129
Journal © 2019 Received 28 Nov 2018; revised 29 Jan 2019; accepted 31 Jan 2019; published 20 Feb 2019

Optics EXPRESS

description. Forbes' 2D-Q polynomials, hereafter referred to as 2D-Qs, were introduced by extending the 1-dimensional Q_{bfs} and are orthogonal in gradient normal departure [10–12]. Zernike difference polynomials, defined as the difference between specific Zernikes, are also orthogonal in slope but can leverage existing computational knowledge about Zernikes [13]. Re-orthogonalizing existing polynomial sets over different apertures also yields new descriptions. Zernikes have been re-orthogonalized over rectangular, hexagonal, annular, and arbitrarily shaped apertures [14–17]. The Q-Legendre and A-polynomials also seek to preserve some properties of the 2D-Qs while being orthogonal over a square aperture [18,19]. This list is by no means exhaustive, and the questions of how surface descriptions compare and whether there is an optimal description for a given design problem are still open and beyond the scope of this work.

Once a surface description is chosen, the description must be leveraged to achieve efficient convergence to a well-corrected minimum. The first step is often to understand the aberrations present in an optical system and how these aberrations can be corrected using orthogonal polynomial terms. Fuerschbach et al. presented a theory of understanding the aberrations of φ- type polynomial surfaces based on their location in the optical system, and uses this method to guide the design of a three-mirror telescope in the ball geometry [7,20]. Yang et al. extended these derivations to include all Zernike terms up to 16th order [21], while Zhong et al. examined the vectorial aberration contributions from biconic surfaces [22]. Bauer et al. developed this framework into a method of design based on vectorial aberration theory, including filters for choosing an optical starting point and strategies for correcting the aberrations identified in a system [23]. By linking design methods with the aberrations present in the system, each orthogonal polynomial term is introduced with the correction of specific aberrations in mind.

It is not uncommon to introduce optimization constraints that enforce the manufacturability of the surfaces in a design. These constraints are often constructed with specific fabrication and testing regimes in mind and can involve sampling the surfaces over a large grid of points. Consequently, including these constraints from the beginning of optimization can be prohibitively time-consuming. Similarly, it is not uncommon to inform the design process with tolerancing information [24–27]. These approaches may lead to improved as-built performance but do not address the freeform departures that drive testability and fabrication time.

By linking the surface manufacturability estimates to the orthogonality of the polynomials used in the surface description, the connection between the sum of the squares of orthogonal polynomial coefficients and the physical quantity associated with the orthogonality of the polynomials can be used to facilitate rapid calculation of the manufacturability estimates and their inclusion as optimization constraints. Both peak-to-valley (PV) sag departure and the maximum of the gradient of the normal departure are used here as manufacturability estimates; in this paper, departures reported are from the base sphere. These manufacturability estimates are linked to the surface testability in null or quasi-null interferometric testing regimes, which are common in optics. Further, improvements in PV sag departure and maximum gradient normal departure may also lead to surfaces that can be fabricated more rapidly or tested using a broader variety of metrology techniques [28–31].

In Section 2 of this paper, we show the mathematical construction of the optimization constraint from the square-sum of the orthogonal polynomial coefficients. The constraint, hereafter referred to as the square-sum penalty, is constructed in general for any orthogonal polynomial basis and in particular for both Zernikes and 2D-Qs. In Section 3 of this paper, we report two designs studies, a three-mirror telescope for use in LWIR imaging and a freeform prism for use as a combiner in AR/VR, to demonstrate that the square-sum penalty enables the fine-tuning of a design to improve manufacturability estimates.

Optics EXPRESS

2. Construction of the square-sum penalty

2.1 Square-sum penalty for orthogonal polynomials

The building blocks of the square-sum penalty are mathematical properties of orthogonal polynomials, which are described by Forbes [32].

A polynomial set $\{P_n\}$ is orthogonal with respect to an angle bracket if

$$\langle P_n, P_{n'} \rangle = 0 \text{ when } n \neq n'. \tag{2.1}$$

Here, the angle bracket $\langle f, g \rangle$ represents an inner product of f and g, which can be thought of as a measure of how much f and g overlap [33]. For orthogonal polynomials, $\langle f, g \rangle$ is typically a weighted integral of some product of f and g over a domain of interest [32]. Equations (2.6) and (2.10) show Zernike and 2D-Q angle brackets, respectively.

Because orthogonal polynomials do not overlap in the sense of Eq. (2.1), terms within an orthogonal basis can be added or removed without concern for cross-term interaction. Consider a sag departure profile defined as a linear combination of polynomial terms, say $S = \sum_n s_n P_n$. An algebraic rearrangement of terms from [32] shows that, for a given polynomial P_n,

$$\frac{\langle S, P_n \rangle}{\langle P_n, P_n \rangle} = \sum_{n'} \frac{\langle s_{n'} P_{n'}, P_n \rangle}{\langle P_n, P_n \rangle} = s_n. \tag{2.2}$$

In other words, each coefficient s_n in an orthogonal expansion can be calculated using only the surface S and the specific corresponding P_n. This equation may be more familiar if $\langle P_n, P_n \rangle$ is normalized to one and suppressed.

Given the angle bracket, the orthogonal polynomial coefficients can be directly linked to physical properties of the freeform surface. Specifically, the angle bracket of a surface with itself often corresponds to some physical property of the surface. For example, the Zernike angle bracket of a Zernike freeform surface with itself is equal to the surface's mean-square sag departure. For a general surface, this angle bracket is calculated as

$$\langle S, S \rangle = \sum_n \langle S, s_n P_n \rangle = \sum_n \langle P_n, P_n \rangle s_n^2. \tag{2.3}$$

This last equation, which is shared by all orthogonal polynomials, enables the manufacturability estimates associated with the angle bracket to be calculated by summing the squares of the orthogonal polynomial coefficients (possibly with weights). For most orthogonal polynomials used to specify freeform optical surfaces, the $\langle P_n, P_n \rangle$ weight factors often have either value one or simple, known formulas. Thus, Eq. (2.3) enables on-the-fly surface manufacturability estimates, in the sense that the manufacturability estimates can be included as optimization penalties without interrupting or prohibitively slowing down the optimization process. To return to the example with Zernikes, Eq. (2.3) enables the mean-square sag departure of a Zernike freeform surface to be calculated with no sampling of the surface and only a handful of arithmetic operations per orthogonal polynomial coefficient.

Using Eq. (2.3), it is possible to calculate – and therefore constrain – the manufacturing or testing estimates associated with the orthogonal polynomials. Sometimes, this constraint can be informed by the application. Ma et al., when designing rotationally-invariant systems with Q_{bfs} for applications including lithography and mobile camera devices, showed that fringe density constraints were effective at reducing design sensitivity [34,35].

Optics EXPRESS

Here, we explore improvement in manufacturability estimates without compromising optical performance. To accomplish this goal, the square sum from Eq. (2.3) is incorporated as an error function penalty. This constraint yields an error function of the form

$$\mathrm{Err}_{\mathrm{SS}} = \mathrm{Err}_{\mathrm{Aber}} + \zeta \cdot \sum_{n} \langle P_n, P_n \rangle s_n^2, \tag{2.4}$$

where $\mathrm{Err}_{\mathrm{Aber}}$ is the error function contribution from system optical performance and ζ is the square-sum penalty weight. This square-sum penalty encourages the optimizer to minimize freeform departure during design. Further, calculating the square-sum penalty involves only a handful of arithmetic operations per orthogonal polynomial term and does not involve sampling the freeform surface over a grid of points.

It is important to note that the manufacturability of individual surfaces within a system is only one of many factors that determine the overall manufacturability of an optical system [36]. Nevertheless, when comparing optical designs differing primarily in freeform surface shape, improvements in the PV sag departures and maximum gradient normal departures can be used to estimate improvements in the manufacturability of the surfaces.

2.2 Square-sum penalty for Zernikes and 2D-Qs

The square-sum penalty is now constructed for Zernikes and 2D-Qs, which are used in the designs in Section 3. Both Zernike and 2D-Q freeform surfaces can be thought of as a base surface plus sag departure described by orthogonal polynomials, with sag equations like Eq. (1.1). Both Zernikes and 2D-Qs have orthogonality properties with respect to angle brackets that are averages over a circular domain. In Eqs. (2.5) to (2.11), angle brackets with only one argument represent averages over a circular domain.

A Zernike freeform surface with a base sphere and sag departure $D(u,\theta)$ has sag equation

$$\begin{aligned} z = f(\rho,\theta) &= \frac{c\rho^2}{1+\sqrt{1-c^2\rho^2}} + D(u,\theta), \\ D(u,\theta) &:= \sum_{n=0}^{N}\sum_{m=-n}^{n} C_n^m Z_n^m(u,\theta). \end{aligned} \tag{2.5}$$

Here, the two-index Zernike notation used is $Z_n^m = u^{|m|} P_{(n-|m|)/2}^{(0,|m|)}(2u^2-1)\begin{Bmatrix}\cos(m\theta) & \text{if } m \ge 0\\ \sin(-m\theta) & \text{if } m < 0\end{Bmatrix}$ and $P_{(n-|m|)/2}^{(0,|m|)}(2u^2-1)$ is a Jacobi polynomial, so long as $(n-|m|)$ is even [37]. This two-index notation is mathematically convenient, although, during design, we use the single-index FRINGE Zernikes that follow the order of the traditional Seidel aberrations [38]. The square-sum penalty does not depend on which Zernike ordering scheme is chosen.

For the Zernikes, the angle bracket used here is

$$\langle f\, g \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi}\int_{0}^{1} f(u,\theta) g(u,\theta)\, u\, du\, d\theta. \tag{2.6}$$

In other words, Zernikes are orthogonal in sag departure, which enables a link between the sag departure and the Zernike coefficients. By replacing S with $D(u,\theta)$ in Eq. (2.3), it can be seen that the mean-square sag departure of a Zernike freeform surface is equal to the (weighted) square sum of the Zernike coefficients given as

Optics EXPRESS

$$\left\langle \left| D(u,\theta) \right|^2 \right\rangle = \left\langle \left(\sum_{n,m} C_n^m Z_n^m \right) \left(\sum_{n',m'} C_{n'}^{m'} Z_{n'}^{m'} \right) \right\rangle = \sum_{n,m} (C_n^m)^2 \frac{\varepsilon_m}{2n+2}, \tag{2.7}$$

where $\varepsilon_m = \begin{Bmatrix} 2 & m=0 \\ 1 & m \neq 0 \end{Bmatrix}$ emerges from evaluating $\left\langle Z_n^m Z_n^m \right\rangle$ in Eq. (2.3). If Eq. (2.7) is introduced into Eq. (2.4), the Zernike square-sum-penalized error function is now given as

$$\mathrm{Err_{SS}} = \mathrm{Err_{Aber}} + \zeta \cdot \sum_{n,m} (C_n^m)^2 \frac{\varepsilon_m}{2n+2}. \tag{2.8}$$

A similar analysis can be done for the 2D-Q polynomials. A 2D-Q freeform surface with a base sphere and normal departure $\delta(u,\theta)$ is described as

$$\begin{aligned} z = f(\rho,\theta) &= \frac{c\rho^2}{1+\sqrt{1-c^2\rho^2}} + \frac{\delta(u,\theta)}{\sqrt{1-c^2\rho^2}}, \\ \delta(u,\theta) &:= u^2(1-u^2)\sum_{n=0}^{N} a_n^0 Q_n^0(u^2) + \sum_{m=1}^{M} u^m \sum_{n=0}^{N} [a_n^m \cos(m\theta) + b_n^m \sin(m\theta)] Q_n^m(u^2). \end{aligned} \tag{2.9}$$

The cosine factor $\sqrt{1-c^2\rho^2}$ scales sag departure to normal departure, while c is the curvature of the sphere that is the unique best fit of the 2D-Q freeform surface [10]. For the 2D-Qs, the angle bracket is

$$\left\langle \nabla f \cdot \nabla g \right\rangle = \int_{-\pi}^{\pi} \int_{0}^{1} \nabla f(u,\theta) \cdot \nabla g(u,\theta) \frac{u\, du\, d\theta}{\sqrt{u^2(1-u^2)}} \Bigg/ \int_{-\pi}^{\pi} \int_{0}^{1} \frac{u\, du\, d\theta}{\sqrt{u^2(1-u^2)}}. \tag{2.10}$$

In other words, the 2D-Qs are orthogonal in gradient normal departure. Correspondingly, the mean-square gradient normal departure of a 2D-Q freeform surface is equal to the square sum of the 2D-Q coefficients given as

$$\left\langle \left| \nabla \delta(u,\theta) \right|^2 \right\rangle = \left\langle \left(\frac{\partial \delta}{\partial u} \right)^2 + \frac{1}{u^2} \left(\frac{\partial \delta}{\partial \theta} \right)^2 \right\rangle = \sum_{m,n} [(a_n^m)^2 + (b_n^m)^2]. \tag{2.11}$$

Equation (2.11) is a well-known result central to the definition of the 2D-Qs. If Eq. (2.11) is introduced into Eq. (2.4), the 2D-Q square-sum-penalized error function is given as

$$\mathrm{Err_{SS}} = \mathrm{Err_{Aber}} + \zeta \cdot \frac{1}{\rho_{\max}^2} \sum_{m,n} [(a_n^m)^2 + (b_n^m)^2]. \tag{2.12}$$

Here, $\frac{1}{\rho_{\max}^2}$ is a renormalization factor, necessary due to converting from u to ρ.

When specifying surfaces with Zernikes or 2D-Qs in design, the region of a freeform surface illuminated by the totality of the footprints of the fields, referred to as the effective aperture of the surface, is not always the same as the circular domain over which the Zernikes or 2D-Qs are orthogonal. Since the square-sum penalty calculates manufacturability estimates over this circular domain of orthogonality, the square-sum penalty may not exactly represent freeform surfaces with highly non-circular effective apertures.

In practice, the effective apertures of freeform surfaces do not need to be perfectly circular. In our Reflective Prism design example in Section 3.4, the effective apertures are non-circular and have an aspect ratio of about 5:4. This aspect ratio is caused by the placement of the freeform surfaces away from the stop in conjunction with the aspect ratio of

Optics EXPRESS

the field of view. Even so, as shown by the results in Section 3.4, the square-sum penalty can improve manufacturability estimates for these non-circular effective apertures.

That said, if the effective apertures in a design are highly non-circular (e.g., if their aspect ratios approach or exceed 2:1), then designers may wish to use other orthogonal polynomials, such as those described by Nikolic et al. or Broemel et al. [18,19]. For this reason, Section 2.1 was written to facilitate the construction of the square-sum penalty for orthogonal polynomials beyond just Zernikes and 2D-Qs.

3. Analysis of square-sum penalty in design

3.1 Design study overview

Two design case studies are presented to analyze the impact of the square sum penalty. For each design, a starting point was chosen that had the same layout and satisfied the same geometry constraints as the completed design but included no freeform terms and consequently had virtually no aberration correction. Orthogonal polynomial coefficients and system parameters were then introduced as optimization variables based on their ability to correct system aberrations, as detailed by Bauer et al. [23]. Use of up to the 16th FRINGE Zernike term or corresponding 2D-Q (see Eqs. (5) and (7) of Takaki et al. [39] or Figs. 16 and 17 of Menke and Forbes [11] for the Zernike-Q correspondence) was allowed on each freeform surface, but only terms necessary for aberration correction were used. The resulting designs, which were created without the square-sum penalty, are used as benchmarks.

The aberration correction process was then repeated with a range of values of ζ, the square-sum penalty weight. To facilitate comparison with the benchmark, no coefficients were used beyond those in the benchmark design, and the square-sum penalty was introduced at the beginning of the optimization process and applied equally to each surface. An initial value for ζ was chosen so that the optical performance and manufacturability estimates of the design were within 5% of the benchmark. Square-sum penalized designs were then created with increasing ζ values, and optical performance and manufacturability estimates (PV sag departure and maximum gradient normal departure) versus ζ were plotted. These plots, in Figs. 2 and 7 below, show that manufacturability estimates improve when the square-sum penalty weight is increased, in exchange for a tradeoff in optical performance.

In practice, the optical performance of a design often must achieve some minimum specification and, therefore, is not always available for tradeoff. At the same time, designers are typically free to choose which orthogonal polynomial coefficients to use, provided that the surfaces in their designs can be manufactured and tested. Consequently, we explored the impact of the square-sum penalty when use of up to the 36th FRINGE Zernike or corresponding 2D-Q was allowed but optical performance equal that of the 16-term unpenalized benchmark was required. To conduct this exploration, we started with 16-term unpenalized benchmark designs and introduced additional higher-order polynomial terms while also introducing and adjusting the square-sum penalty so that optical performance remained equivalent to that of the benchmark. By leveraging the square-sum penalty in conjunction with additional polynomial terms, we improved manufacturability estimates without sacrificing optical performance, as shown in Sections 3.3 and 3.4, below.

3.2 Notes on design procedures

For all Zernike designs, the base sphere was constrained to be the best-fit sphere that eliminates the local sag departure at the origin and average sag departure around the edge, which is the same best-fit sphere used by the 2D-Qs, although different definitions of best-fit sphere can also be used for the Zernikes. To eliminate redundancy between the surfaces' physical tilt parameters and the polynomial tilt terms, which are defined as tilt-degeneracy in [39], the mean tilt terms (i.e., terms with coefficients $C2/C3$ for FRINGE Zernikes and a_0^1/b_0^1 for 2D-Qs) were removed and the freeform surface coordinate systems were required

Optics EXPRESS

to be centered on the optical axis ray. For all designs, the normalization radius of each freeform surface was chosen to enclose the effective aperture tightly while also allowing for increases in the size of the effective aperture during optimization. For all design studies, the optical performance reported is the nominal optical performance.

All numerical optimization was conducted in CODE V. Both the step optimization algorithm and the finite differences method for computing derivative increments were each used at points during the design, based on designer discretion. Global search methods were not used. For all designs, root-mean-square spot size was used by CODE V as the merit function during optimization.

As a general note on the use of the square-sum penalty, we consider it best practice to include the square-sum penalty as early in the design process as allowed by the design study. As another general note, we found that enabling the finite differences method during the ending stages of optimization improved the numerical stability of convergence.

Surface coefficients will be made available upon request via communication with the primary author.

3.3 First design study: three-mirror telescope

The first design study is a plane-symmetric, ball-geometry three-mirror telescope for use in LWIR imaging, based on a design by Fuerschbach et al. [7]. Fuerschbach et al.'s design achieved a root-mean-square wavefront error (RMS WFE) of less than λ/100 at 10 μm while maintaining a ball geometry and prioritizing low volume. We used the same field, wavelength, aperture size, and F/# as Fuerschbach et al. and aimed for our unpenalized benchmark designs to achieve similar optical performance. The overall volume of each design is no greater than 820 ml. In all designs, the projections of the image plane and all three mirrors onto the system's plane of symmetry was constrained to lie within a circle of radius 70 mm, and in this way maintained the ball geometry. Full specifications for the three-mirror telescope design study can be seen in Table 1.

For both Zernikes and 2D-Qs, a benchmark design was created without use of the square sum penalty. All terms up to the 16th FRINGE Zernike or corresponding 2D-Q term were necessary for aberration correction except for coma and astigmatism on the second surface. Cross-sections and full-field displays (FFDs) of the RMS WFE of these benchmark designs are shown in Fig. 1.

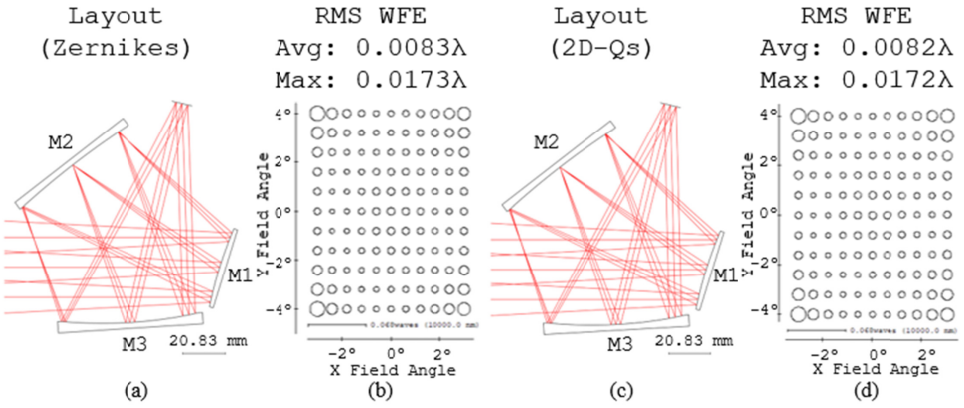

Fig. 1. Layout and optical performance (λ = 10 μm) for unpenalized designs used as benchmarks with (a,b) Zernikes and (c,d) 2D-Qs.

The square-sum penalty was then introduced, and the design process was repeated with ζ values increasing by orders-of-magnitude. Figure 2 reports the optical performance (in terms of average RMS WFE over the field) and manufacturability estimates for square-sum

Optics EXPRESS

penalized designs. In this design study, the manufacturability estimates of the third mirror were always much greater than the manufacturability estimates of the other two mirrors. Consequently, we report both the manufacturability estimates for the third mirror individually as well as the sum of each manufacturability estimate across all three mirrors together.

As we can see in Fig. 2, manufacturability estimates improve as ζ increases. For the Zernikes, PV sag departure and maximum gradient normal departure of the third mirror improved from 362 μm and 1.04° in the benchmark to 124 μm and 0.47°, respectively, while summed manufacturability estimates improved from 395 μm and 1.36° in the benchmark to 153 μm and 0.68°. For the 2D-Qs, the manufacturability estimates of the third mirror improved from 321 μm and 0.97° in the benchmark to 121 μm and 0.41° while summed manufacturability estimates improved from 377 μm and 1.24° in the benchmark to 148 μm and 0.56°. Optical performance was traded off from 0.0083 λ to 0.0726 λ for the Zernikes and from 0.0082 λ to 0.0612 λ for the 2D-Qs (λ = 10 μm), so that the upper-bound square-sum penalized designs are both approximately diffraction limited.

We also note from Fig. 2 that the square-sum penalty is maximally effective during the first three orders of magnitude of increase of the square-sum penalty weight. For the Zernikes, manufacturability estimates for the third mirror improved to 138 μm and 0.49° when the weight was 10^{-2}, a factor of two improvement. The tradeoff in optical performance was comparatively less, from 0.0083 λ to 0.0142 λ. For the 2D-Qs, manufacturability estimates for the third mirror improved to 128 μm and 0.43° when the weight was 10^{0}, and again the tradeoff in optical performance was comparatively less, from 0.0082 λ to 0.0143 λ.

Table 1. Three-Mirror Telescope Specifications

Parameters	Specifications
Full Field of View	10° diagonal
Entrance Pupil Diameter	30 mm
Detector Size	6 mm x 8 mm
Wavelength	10 μm
Focal Length	57 mm
Volume	< 820 ml
Distortion	< 3%
Average RMS WFE of Benchmark	0.0085 waves (λ = 10 μm)
Ball Geometry Radius	70 mm

This region of maximal square-sum penalty effectiveness coincides with the region in which the layout shifts. Figure 3 shows the layouts of the 2D-Q benchmark design alongside the 2D- Q square-sum penalized designs with weights 10^{-2}, 10^{0} and 10^{2}. By comparing the benchmark design to the square-sum penalized design with weight 10^{0}, we see that, although both designs maintain the ball geometry, the layout has evolved. At the same time, Fig. 2 shows that manufacturability estimates have improved substantially. On the other hand, clearance and ball geometry constraints prevent the layout from continuing to change, so the design with weight 10^{2} sees virtually no improvement in manufacturability estimates and a significant optical performance tradeoff. The Zernike designs see a similar trend (not shown).

Next, we leveraged the square-sum penalty in conjunction with an increased number of coefficients to improve manufacturability estimates without decreasing optical performance. Use of up to the 36th FRINGE Zernike or corresponding 2D-Q term was allowed, and the weight of the square-sum penalty was adjusted so that the optical performance remained equivalent to that of the unpenalized benchmark design. We also created 36-term designs without use of the square-sum penalty. Both 36-term designs are compared to the 16-term unpenalized benchmark designs, in Table 2 for the Zernikes and Table 3 for the 2D-Qs.

With the introduction of the square-sum penalty, manufacturability estimates improved relative to the benchmark without any decrease in optical performance. For both the Zernikes and 2D-Qs, PV sag departure of the third mirror improved by about 40%, from 362 μm to 225 μm and 322 μm to 189 μm, respectively. Maximum gradient normal departure of the

third mirror also improved by about 40%, from 1.04° to 0.68° for the Zernikes and from 0.97° to 0.60° for the 2D-Qs. Summed PV sag departure and maximum gradient normal departure show similar improvements. At the same time, optical performance remains equivalent to that of the benchmark, as intended. While a full comparison of polynomial sets is beyond the scope of this paper, we note that Zernike and 2D-Q designs achieve similar optical performance and manufacturability estimates.

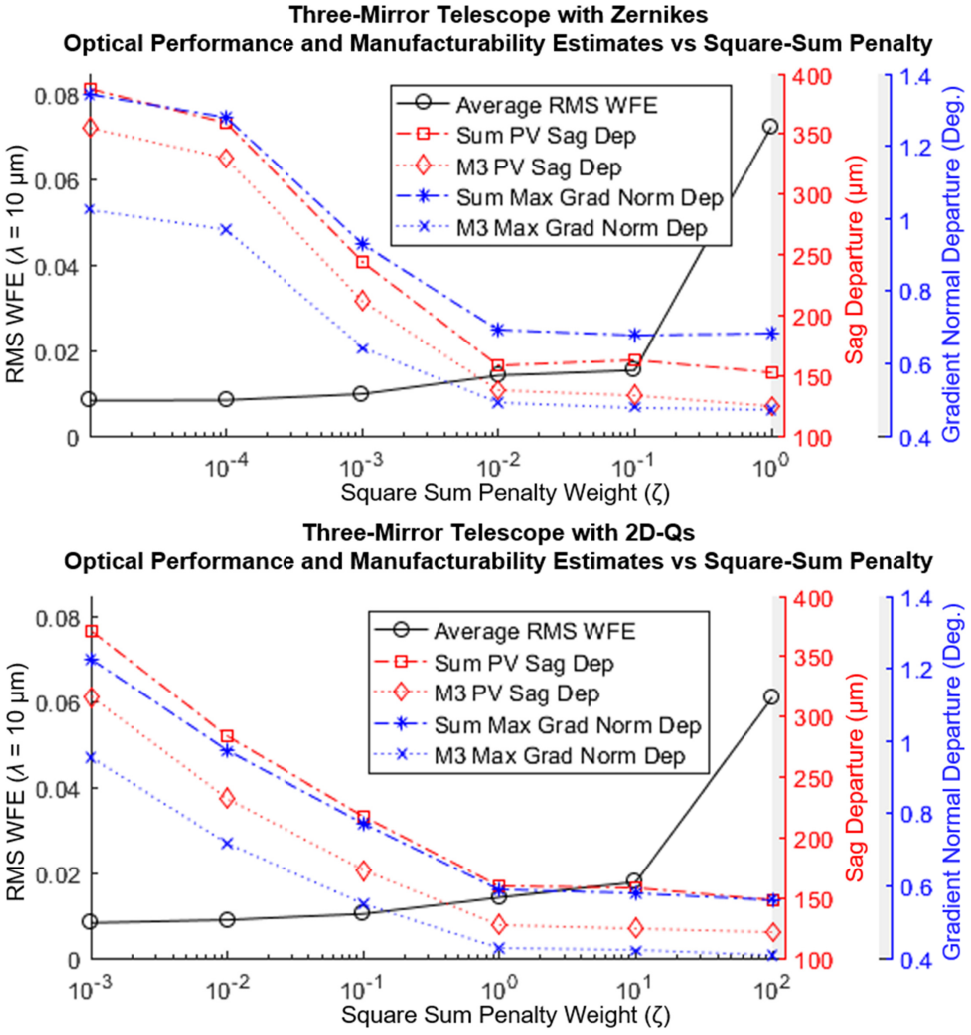

Fig. 2. For the three-mirror telescope, optical performance and manufacturability estimates versus square-sum penalty weight with (top) Zernikes and (bottom) 2D-Qs. All values of the lower-bound designs are within 5% of the benchmark.

To demonstrate that the optical performance is equivalent over the field and that the ball geometry has been maintained, we compare the layout and FFDs of the RMS WFE for the 36-term square-sum penalized designs, reported in Fig. 4, with the benchmark, reported in Fig. 1. The designs' layouts see a slight change: the 36-term square-sum penalized designs are both similar in layout to that of the design with weight 10^{-2} shown in Fig. 3. This pattern is consistent with the observation that the square-sum penalty is most effective when able to adjust the geometry. The optical performance is also comparable: the average RMS WFE is approximately 0.0085 λ and performance decreases towards the edge of the field.

Optics EXPRESS

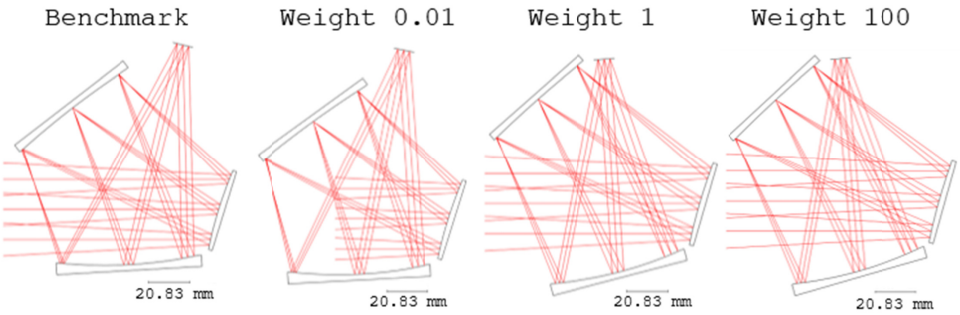

Fig. 3. Layout for (left) the benchmark design and square-sum penalized designs with 2D-Qs for weights (middle-left) 10^{-2}, (middle-right) 10^{0}, and (right) 10^{2}. The layout changes from the benchmark to the weight 10^{-2} and weight 10^{0} designs, but not from the weight 10^{0} design to the weight 10^{2} design. The Zernike designs (not shown) have a similar trend.

Table 2. For the three-mirror telescope with Zernikes, optical performance and manufacturability estimates of (left) the 16-term unpenalized benchmark design, (middle) the 36-term unpenalized design, and (right) the 36-term square-sum penalized design.

Zernike Three-Mirror Telescope Designs						
	16 Zernikes, Unpenalized		36 Zernikes, Unpenalized		36 Zernikes, Square-Sum Penalized	
	Mean RMS WFE		Mean RMS WFE		Mean RMS WFE	
	0.0083 λ (λ = 10 μm)		0.0058 λ (λ = 10 μm)		0.0084 λ (λ = 10 μm)	
	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure
M1	28 μm	0.25°	28 μm	0.28°	26 μm	0.25°
M2	5 μm	0.07°	11 μm	0.09°	9 μm	0.08°
M3	362 μm	1.04°	377 μm	1.09°	225 μm	0.68°
SUM	395 μm	1.36°	416 μm	1.46°	260 μm	1.01°

Table 3. For the three-mirror telescope with 2D-Qs, optical performance and manufacturability estimates of (left) the 16-term unpenalized benchmark design, (middle) the 36-term unpenalized design, and (right) the 36-term square-sum penalized design.

2D-Q Three-Mirror Telescope Designs						
	16 Qs, Unpenalized		36 Qs, Unpenalized		36 Qs, Square-Sum Penalized	
	Mean RMS WFE		Mean RMS WFE		Mean RMS WFE	
	0.0082 λ (λ = 10 μm)		0.0057 λ (λ = 10 μm)		0.0085 λ (λ = 10 μm)	
	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure
M1	50 μm	0.21°	50 μm	0.23°	41 μm	0.20°
M2	5 μm	0.06°	13 μm	0.08°	9 μm	0.08°
M3	322 μm	0.97°	320 μm	0.99°	189 μm	0.60°
SUM	377 μm	1.24°	383 μm	1.30°	239 μm	0.88°

Lastly, we confirm that the improvement is not achieved simply by increasing the number of coefficients. For both Zernikes and 2D-Qs, the average RMS WFE improved by about 30% from the 16-term unpenalized benchmark to the 36-term unpenalized designs, from 0.0083 λ to 0.0058 λ and 0.0082 λ to 0.0057 λ, respectively. However, manufacturability estimates are almost entirely unchanged, meaning that the improvement in manufacturability estimates observed with the 36-term square-sum penalized designs is driven by the square-sum penalty working in conjunction with the additional orthogonal polynomial coefficients.

Optics EXPRESS

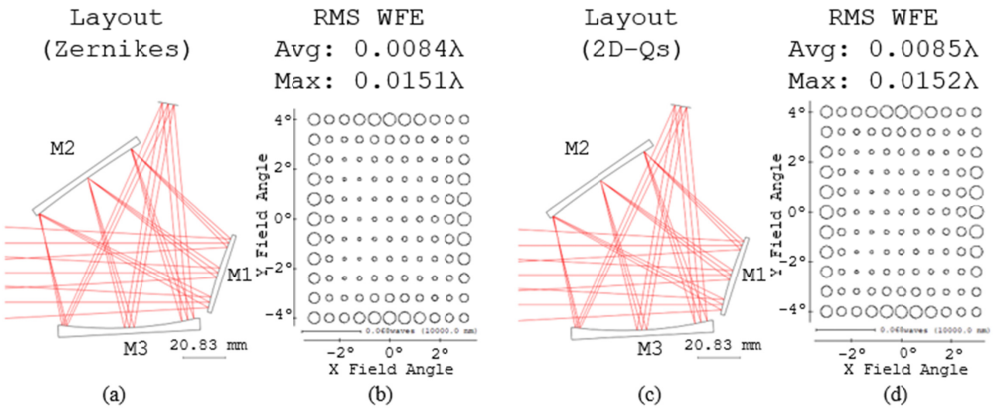

Fig. 4. Layout and optical performance (λ = 10 μm) for 36-term square-sum penalized designs with (a,b) Zernikes and (c,d) 2D-Qs.

3.4 Second design study: freeform prism combiner

The second design study is a freeform prism for use as a combiner in AR/VR applications. This design study is based on a reflective freeform prism design proposed by Chen and Herkommer, who document that the reflective freeform prism geometry offers advantage in optical performance versus other combiner geometries [40]. Our designs differ from that of Chen and Herkommer primarily in terms of pupil size and metric for optical performance. We designed the system with an 8 mm diameter eyebox and required the performance to achieve >10% MTF at the Nyquist frequency of the micro-display for all fields, evaluated over 3 mm sub-pupils which sample the eyebox. The micro-display was chosen so that the pixels span 1.5 arcminutes when projected to the human eye. Chen and Herkommer prioritized having a flat first surface to facilitate additive manufacturing; to achieve a reduced volume with the large eyebox, the designs in the current paper also remove the flat surface requirement.

Full specifications of the freeform prism are presented in Table 4. The centered and outermost sub-pupils used for optical performance evaluation are shown in Fig. 5; sub-pupils located halfway between the outermost and centered sub-pupils were also used.

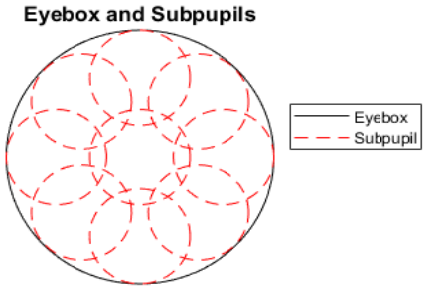

Fig. 5. Location of the centered and outermost 3 mm sub-pupils used for evaluating optical performance within the 8 mm eyebox. Performance was also evaluated over sub-pupils located halfway between the outermost and center sub-pupils (not shown).

As with the first design study, an unpenalized benchmark design was created. Unlike the first design study, in which nearly every orthogonal polynomial coefficient up to the 16th FRINGE Zernike or corresponding 2D-Q term was used on each freeform surface, suitable optical performance could be achieved primarily using terms on the second freeform surface; only two orthogonal polynomial terms (astigmatism and coma) were varied on the first freeform surface. Figure 6 shows XZ cross-sections and optical performance of the Zernike and 2D-Q benchmark designs. Optical performance is reported in terms of MTF at 50

Optics EXPRESS

cyc/mm over 3 mm sub-pupils, and full-field displays of both sagittal and tangential MTF are shown for the best-case and worst-case sub-pupils. Manufacturability estimates are evaluated over the effective aperture of each surface.

Table 4. Specifications of Reflective Prism

Parameters	Specifications
Full Field of View	26° by 41°
Eyebox Diameter	8 mm
Wavelength	587 nm
Focal Length	21 mm
Volume	6.5 ml
Distortion	< 12.5%
Image Quality (Benchmark)	MTF >10% @ 50 cyc/mm Evaluated over 3 mm sub-pupils (See Fig. 5)
Eye Clearance	>18.25 mm
Micro-display	
Diagonal Length of Active Display	21.8 mm
Resolution in Pixels	1080 x 1920
Pixel Pitch	10 μm
Pixel Angular Subtense	1.5 arcminutes

The square-sum penalty was then introduced, and the designs recreated with increasing ζ value. Figure 7 reports that the optical performance and manufacturability estimates for the 16- term square-sum penalized designs. Optical performance is reported as the minimum MTF at 50 cyc/mm across all fields, evaluated over fifteen 3 mm sub-pupils which sample the 8 mm eyebox, while manufacturability estimates are reported in the same terms as the first design study. As with the first design study, the lower bound of the range of ζ values was chosen so that the manufacturability and optical performance are within 5% of benchmark values. The square-sum penalty weight was increased by half-orders of magnitude until optical performance drops below 10% MTF. For both Zernikes and 2D-Qs, increasing the square-sum penalty by another half-order of magnitude beyond this upper bound decreased MTF to nearly zero percent MTF.

As ζ increases, the manufacturability estimates improve, and optical performance is traded off. For both Zernikes and 2D-Qs, the second freeform surface experienced most of the manufacturability estimate improvement from benchmark to endpoint: from 201 μm to 111 μm and 2.01° to 1.08° for the Zernikes, and from 254 μm to 154 μm and 1.77° to 1.10° for the 2DQs. The shape of the first mirror is strongly dominated by astigmatism in all designs. In the Zernike design, Mirror 1 has 129 μm of Zernike astigmatism in the benchmark versus 120 μm of Zernike astigmatism in the endpoint, while, in the 2D-Q design, Mirror 1 has 186 μm of Q astigmatism (i.e., terms with coefficients a_0^2, b_0^2 for the 2D-Qs) in the benchmark versus 173 μm of Q astigmatism in the endpoint. Significant adjustment of astigmatism on the first mirror requires a more significant tradeoff of optical performance than is allowed in Fig. 7 or, alternatively, an increase in the square-sum penalty weight on Mirror 1 relative to that of Mirror 2, which was not allowed in this design study.

Optics EXPRESS

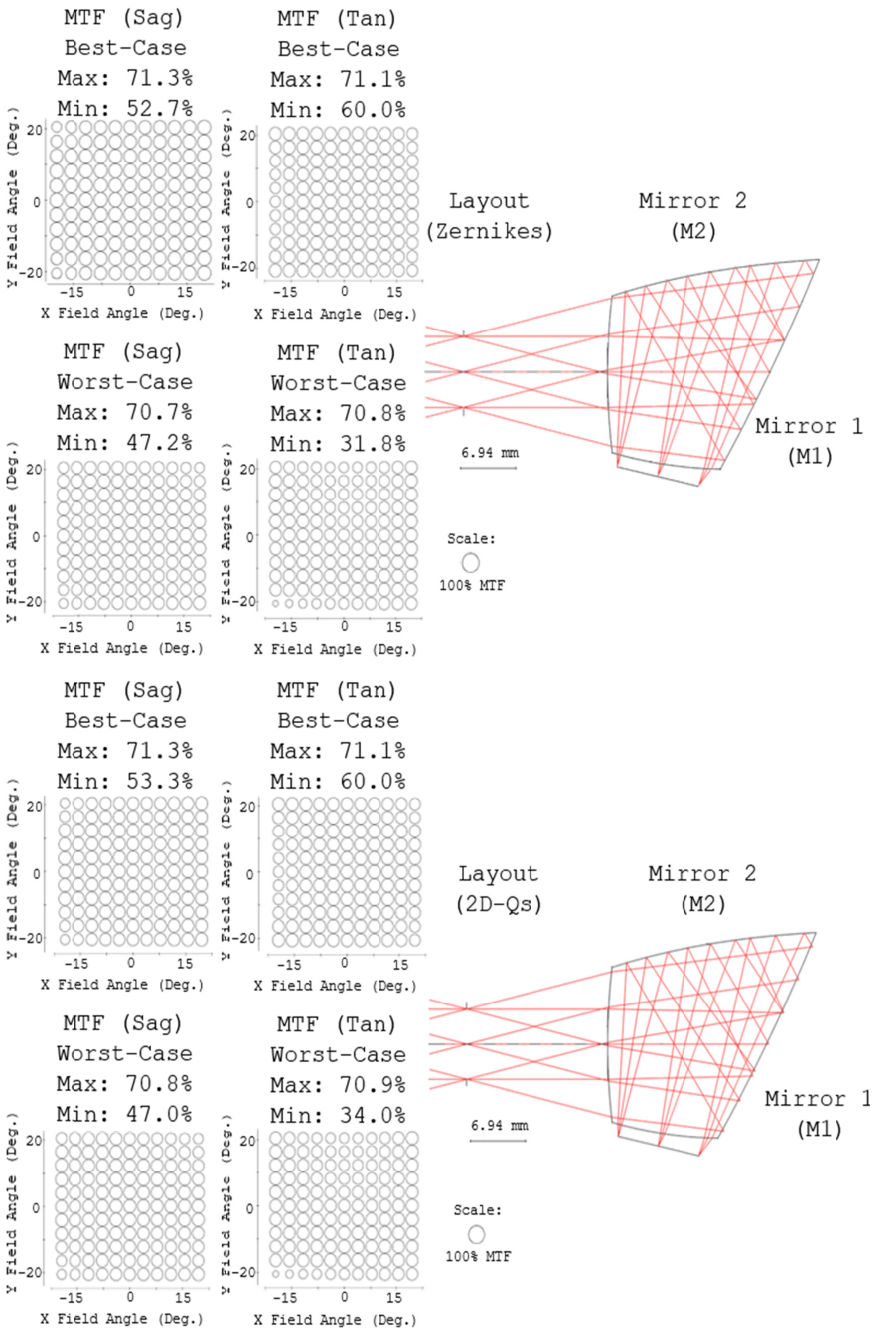

Fig. 6. Layout and optical performance for unpenalized benchmark reflective prism designs with (top) Zernikes and (bottom) 2D-Qs. FFDs of sagittal and tangential MTF at 50 cyc/mm are shown for best-case and worst-case sub-pupils.

As with the first design study, this region of effectiveness for the square-sum penalty is coupled with a change in layout of the freeform prism designs. Figure 8 shows the change in layout from the Zernike benchmark design to the highest-weight ($\zeta = 2*10^{-3/2}$) square-sum penalized Zernike design. The 2D-Qs see a similar trend (not shown).

Optics EXPRESS

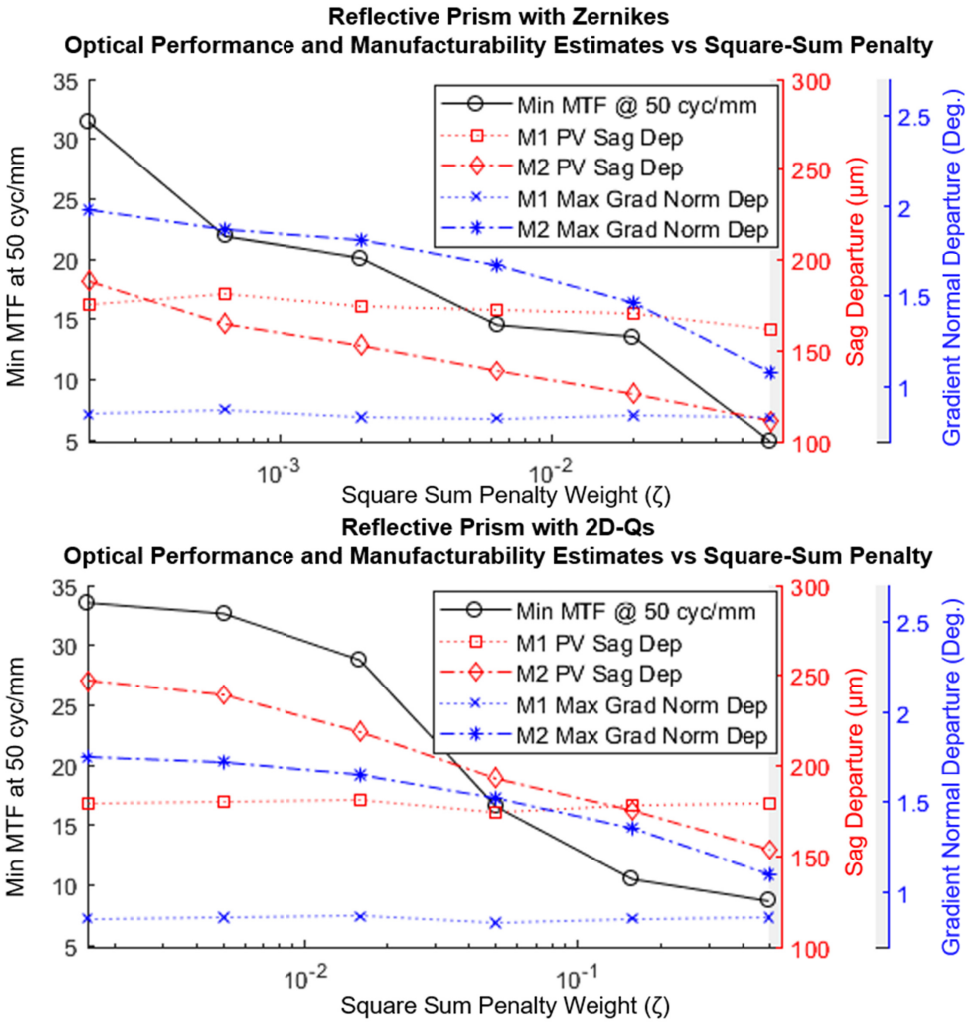

Fig. 7. For the reflective prism, optical performance versus manufacturability estimates as square-sum penalty changes with (top) Zernikes and (bottom) 2D-Qs. All values of the lower-bound design are within 5% of the benchmark.

As with the first design study, we also leveraged the square-sum penalty in conjunction with additional coefficients to improve manufacturability estimates without decreasing optical performance. Tables 5 and 6 compare the 16-term unpenalized benchmark designs with the 36-term unpenalized and 36-term square-sum penalized designs for both Zernikes and 2D-Qs, respectively. As in the first design study, square-sum penalized designs achieve improved manufacturability estimates without a decrease in optical performance relative to the benchmark. The PV sag departure of the second mirror improved by about 40%, from 201 μm to 119 μm for Zernikes and from 254 μm to 153 μm for 2D-Qs. Maximum gradient normal departure of the second mirror improved similarly, from 2.01° to 1.17° for the Zernikes and from 1.77° to 0.98° for the 2D-Qs. For both Zernikes and 2D-Qs, manufacturability estimates of the first mirror remained effectively unchanged, so that the summed manufacturability estimates improved only by about 25-30%. As in the first design study, Zernike and 2D-Q designs are very similar in both optical performance and manufacturability estimates.

Optics EXPRESS

We lastly note simply increasing the number of coefficients improved optical performance but not manufacturability estimates. The improvement in manufacturability estimates is a consequence of the square-sum penalty in conjunction with the additional coefficients.

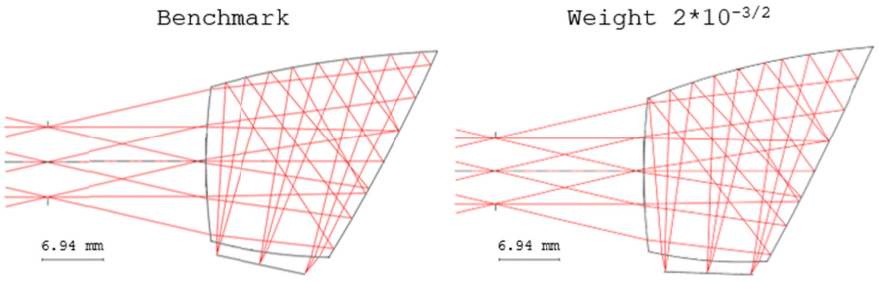

Fig. 8. Layout of Zernike prism designs for (left) the unpenalized benchmark and (right) the highest-weight square-sum penalized designs. Notice the change in layout. The 2D-Q designs (not shown) show a similar trend.

Table 5. For the freeform prism with Zernikes, optical performance and manufacturability estimates of (left) the 16-term unpenalized benchmark design, (middle) the 36-term unpenalized design, and (right) the 36-term square-sum penalized design.

Zernike Freeform Prism Designs						
	16 Zernikes, Unpenalized		36 Zernikes, Unpenalized		36 Zernikes, Square-Sum Penalized	
	Min MTF @ 50 cyc/mm Evaluated over sub-pupils		Min MTF @ 50 cyc/mm Evaluated over sub-pupils		Min MTF @ 50 cyc/mm Evaluated over sub-pupils	
	31.8%		45.2%		31.8%	
	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure
M1	168 µm	0.82°	147 µm	0.81°	162 µm	0.77°
M2	201 µm	2.01°	235 µm	2.37°	119 µm	1.17°
SUM	369 µm	2.83°	382 µm	3.17°	281 µm	1.94°

Table 6. For the freeform prism with 2D-Qs, optical performance and manufacturability estimates of (left) the 16-term unpenalized benchmark design, (middle) the 36-term unpenalized design, and (right) the 36-term square-sum penalized design.

2D-Q Freeform Prism Designs						
	16 2D-Qs, Unpenalized		36 2D-Qs, Unpenalized		36 2D-Qs, Square-Sum Penalized	
	Min MTF @ 50 cyc/mm Evaluated over sub-pupils		Min MTF @ 50 cyc/mm Evaluated over sub-pupils		Min MTF @ 50 cyc/mm Evaluated over sub-pupils	
	34.0%		47.9%		30.0%	
	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure	PV Sag Departure	Max Gradient Normal Departure
M1	177 µm	0.85°	171 µm	0.73°	178 µm	0.78°
M2	254 µm	1.77°	326 µm	2.06°	153 µm	0.98°
SUM	431 µm	2.62°	497 µm	2.79°	331 µm	1.76°

Optics EXPRESS

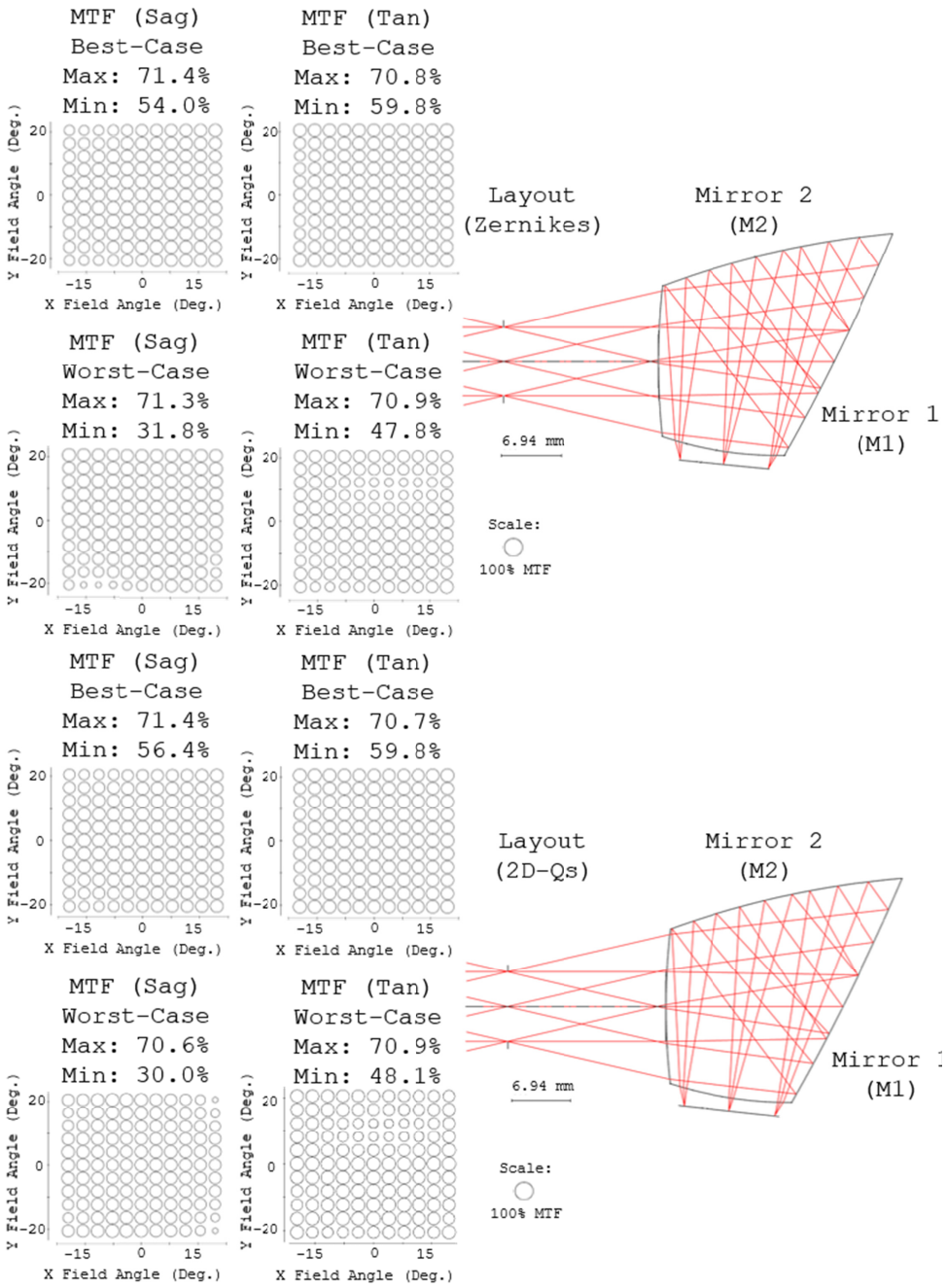

Fig. 9. Layout and optical performance for 36-term square-sum penalized prism designs with (top) Zernikes and (bottom) 2D-Qs. FFDs of sagittal and tangential MTF at 50 cyc/mm are shown for best-case and worst-case sub-pupils.

The layout and optical performance of the 36-term square-sum designs reported in Fig. 9 are compared with those of the 16-term unpenalized benchmark designs reported in Fig. 6. Optical performance for both 36-term square-sum penalized designs exceeds 30% MTF at 50 cyc/mm, which is equivalent to that of the benchmark. All designs also maintain the same freeform prism geometry with a volume of 6.5 ml.

Optics EXPRESS

4. Conclusion

In this paper, an optimization constraint for use in freeform optical design, which calculates manufacturability estimates of a freeform surface, was constructed from the square sum of orthogonal polynomial coefficients. This square-sum penalty can be constructed for general orthogonal polynomials, with examples provided for both Zernikes and 2D-Qs. The square-sum penalty was then introduced in two design examples: a three-mirror ball geometry telescope and a freeform prism combiner. It is shown that using the square-sum penalty with a fixed number of coefficients leads to an improvement in manufacturability estimates in exchange for a tradeoff of optical performance. Further, if the number of coefficients is increased in conjunction with the introduction of the square-sum penalty, manufacturability estimates can be improved with no sacrifice of optical performance.

Funding

National Science Foundation I/UCRC Center for Freeform Optics (IIP-1338877, IIP-1338898, IIP-1822049, and IIP-1822026).

Acknowledgements

We thank Jonathan Papa, Jacob Reimers, and Eric Schiesser for insight about designing with Zernike freeform surfaces. We thank Greg Forbes for insight learned from discussions of freeform surface descriptions with both Zernikes and 2D-Qs. We thank Synopsys for the student license of CODE V.

References

1. "ISO 10110-19 Optics and Photonics - Preparation of drawings for optical elements and systems - Part 19," (ISO, Geneva, Switzerland, 2015).
2. F. Zernike, "Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode," Physica **1**(7-12), 689–704 (1934).
3. D. S. Grey, "Orthogonal Polynomials As Lens-Aberration Coefficients," Proc. SPIE **0237**, 85–90 (1980).
4. L. Cook, G. Perron, B. Zellers, and B. Cohn, "Display system having coma-control plate in relay lens," U.S. Patent 4,826,287 (May 2 1989).
5. T. Nakano and Y. Tamagawa, "Configuration of an off-axis three-mirror system focused on compactness and brightness," Appl. Opt. **44**(5), 776–783 (2005).
6. Z. Zheng, X. Sun, X. Liu, and P. Gu, "Design of reflective projection lens with Zernike polynomials surfaces," Displays **29**(4), 412–417 (2008).
7. K. Fuerschbach, J. P. Rolland, and K. P. Thompson, "A new family of optical systems employing φ-polynomial surfaces," Opt. Express **19**(22), 21919–21928 (2011).
8. E. W. Weisstein, "Chebyshev Polynomial of the First Kind" (from MathWorld - A Wolfram Web Resource), retrieved September 1, 2018, *http://mathworld.wolfram.com/ChebyshevPolynomialoftheFirstKind.html*.
9. E. W. Weisstein, "Legendre Polynomial" (from MathWorld - A Wolfram Web Resource), retrieved September 1, 2018, *http://mathworld.wolfram.com/LegendrePolynomial.html*.
10. G. W. Forbes, "Characterizing the shape of freeform optics," Opt. Express **20**(3), 2483–2499 (2012).
11. C. Menke and G. W. Forbes, "Optical design with orthogonal representations of rotationally symmetric and freeform aspheres," Advanced Optical Testing **2**(1), 97–109 (2013).
12. Synopsys, "CODE V Q-Freeform UD1 Release Notes," (Synopsys, Inc, 2017).
13. D. Ochse, K. Uhlendorf, and L. Reichmann, "Describing freeform surfaces with orthogonal functions," Proc. SPIE **9626**, 962612 (2015).
14. V. Mahajan, "Zernike annular polynomials for imaging systems with annular pupils," J. Opt. Soc. Am. **71**(1), 75–85 (1981).
15. W. Swantner and W. W. Chow, "Gram-Schmidt orthonormalization of Zernike polynomials for general aperture shapes," Appl. Opt. **33**(10), 1832–1837 (1994).
16. M. Bray, "Orthogonal Polynomials: A set for square areas," Proc. SPIE **5252**, 314–321 (2004).
17. C. Ferreira, J. L. López, R. Navarro, and E. P. Sinusía, "Zernike-like systems in polygons and polygonal facets," Appl. Opt. **54**(21), 6575–6583 (2015).
18. M. Nikolic, P. Benítez, B. Narasimhan, D. Grabovickic, J. Liu, and J. Minano, "Optical design through optimization for rectangular apertures using freeform orthogonal polynomials: a case study," Optical Engineering **55**, 071204(071201–071206) (2016).
19. A. Broemel, U. Lippman, and H. Gross, "Freeform surface descriptions - part I: Mathematical representations," Adv. Opt. Technol. **6**, 327–336 (2017).

Optics EXPRESS

20. K. Fuerschbach, J. P. Rolland, and K. P. Thompson, "Theory of aberration fields for general optical systems with freeform surfaces," Opt. Express **22**(22), 26585–26606 (2014).
21. T. Yang, J. Zhu, and G. Jin, "Nodal aberration properties of coaxial imaging systems using Zernike polynomial surfaces," J. Opt. Soc. Am. A **32**(5), 822–836 (2015).
22. Y. Zhong and H. Gross, "Vectorial aberrations of biconic surfaces," J. Opt. Soc. Am. A **35**(8), 1385–1392 (2018).
23. A. Bauer, E. M. Schiesser, and J. P. Rolland, "Starting geometry creation and design method for freeform optics," Nat. Commun. **9**(1), 1756 (2018).
24. D. S. Grey, "Tolerance sensitivity and optimization," Appl. Opt. **9**(3), 523–526 (1970).
25. J. R. Rogers, "Using Global Synthesis to find tolerance-insensitive design," Proc. SPIE **6342**, 63420M (2006).
26. B. G. Crowther and J. R. Rogers, "Desensitization in aspheric and freeform optical designs," Proc. SPIE **10590**, 1059010 (2017).
27. B. J. Bauman and M. D. Schneider, "Design of optical systems that maximize as-built performance using tolerance/compensator-informed optimization," Opt. Express **26**(11), 13819–13840 (2018).
28. G. W. Forbes, "Manufacturability estimates for optical aspheres," Opt. Express **19**(10), 9923–9941 (2011).
29. F. Z. Fang, X. D. Zhang, A. Weckenmann, G. X. Zhang, and C. Evans, "Manufacturing and measurement of freeform optics," CIRP Ann. **62**(2), 823–846 (2013).
30. T. Blalock, B. Myer, I. Ferralli, M. Brunelle, and T. Lynch, "Metrology for the manufacturing of freeform optics," Proc. SPIE **10448**, 46 (2017).
31. M. Davies, Mechanical Engineering and Engineering Science, The University of North Carolina at Charlotte, (personal communication, October 2, 2018).
32. G. W. Forbes, "Robust and fast computation for the polynomials of optics," Opt. Express **18**(13), 13851–13862 (2010).
33. J. Renze, C. Stover, and E. W. Weisstein, "Inner Product" (from MathWorld - A Wolfram Web Resource), retrieved January 25, 2019, *http://mathworld.wolfram.com/InnerProduct.html*.
34. B. Ma, L. Li, K. P. Thompson, and J. P. Rolland, "Applying slope constrained Q-type aspheres to develop higher performance lenses," Opt. Express **19**(22), 21174–21179 (2011).
35. B. Ma, K. Sharma, K. P. Thompson, and J. P. Rolland, "Mobile device camera design with Q-type polynomials to achieve higher production yield," Opt. Express **21**(15), 17454–17463 (2013).
36. U. Fuchs and S. R. Kiontke, "Discussing design for manufacturability for two freeform imaging systems," Proc. SPIE **9948**, 99480L (2016).
37. E. W. Weisstein, "Jacobi Polynomials" (from MathWorld - A Wolfram Web Resource), retrieved October 30, 2018, *http://mathworld.wolfram.com/JacobiPolynomial.html*.
38. "Appendix A Zernike Polynomials," in *CODEV Lens System Setup Reference Manual* (2015).
39. N. Takaki, A. Bauer, and J. P. Rolland, "Degeneracy in freeform surfaces described with orthogonal polynomials," Appl. Opt. **57**(35), 10348–10354 (2018).
40. B. Chen and A. M. Herkommer, "Alternate optical designs for head-mounted displays with a wide field of view," Appl. Opt. **56**(4), 901–906 (2017).