Official journal of the CIOMP 2047-7538
www.nature.com/lsa

ARTICLE Open Access

Freeform imaging systems: Fermat's principle unlocks "first time right" design

Fabian Duerr[1] and Hugo Thienpont[1]

Abstract

For more than 150 years, scientists have advanced aberration theory to describe, analyze and eliminate imperfections that disturb the imaging quality of optical components and systems. Simultaneously, they have developed optical design methods for and manufacturing techniques of imaging systems with ever-increasing complexity and performance up to the point where they are now including optical elements that are unrestricted in their surface shape. These so-called optical freeform elements offer degrees of freedom that can greatly extend the functionalities and further boost the specifications of state-of-the-art imaging systems. However, the drastically increased number of surface coefficients of these freeform surfaces poses severe challenges for the optical design process, such that the deployment of freeform optics remained limited until today. In this paper, we present a deterministic direct optical design method for freeform imaging systems based on differential equations derived from Fermat's principle and solved using power series. The method allows calculating the optical surface coefficients that ensure minimal image blurring for each individual order of aberrations. We demonstrate the systematic, deterministic, scalable, and holistic character of our method with catoptric and catadioptric design examples. As such we introduce a disruptive methodology to design optical imaging systems from scratch, we largely reduce the "trial-and-error" approach in present-day optical design, and we pave the way to a fast-track uptake of freeform elements to create the next-generation high-end optics. We include a user application that allows users to experience this unique design method hands-on.

Introduction

Optical imaging systems have been playing an essential role in scientific discovery and societal progress for several centuries[1,2]. For more than 150 years scientists and engineers have used aberration theory to describe and quantify the deviation of light rays from ideal focusing in an imaging system, and to develop methods to design diffraction-limited imaging systems[3]. Until recently most of these imaging systems included spherical and aspherical refractive lenses (dioptric systems) or reflective mirrors (catoptric systems) or a combination of both (catadioptric systems). The last decennia, with the introduction of new ultra-precision manufacturing methods, such as single-point diamond turning and multi-axis polishing[4], two-photon polymerization[5,6] or other additive manufacturing technologies[7], it has become possible to fabricate lenses and mirrors that have at least one optical surface that lacks the common translational or rotational symmetry about a plane or an axis. Such optical components are called freeform optical elements[8] and they can be used to greatly extend the functionalities, improve performance, and reduce volume and weight of optical imaging systems that are principal parts of spectrometers[9], telescopes[10,11], medical imaging systems[12,13], augmented and virtual reality systems[14], or lithography platforms[15,16]. As such imaging systems including freeform optical elements will be key to tremendously advance science and engineering in a wide range of application domains such as astronomy, material research, chip fabrication, visualization, metrology and inspection, energy production, safety and security, biotechnology and medical imaging and diagnosis[17].

Today the design of optical systems largely relies on efficient ray-tracing and optimization algorithms for which

Correspondence: Fabian Duerr (fduerr@b-phot.org)
[1]Brussels Photonics, Department of Applied Physics and Photonics (B-PHOTONA) Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

a variety of commercial software[18] and optimization algorithms[19] are available. During an optical design cycle, different parameters of the optical system such as the optical material parameters, radii, coefficients, and positions of the optical surfaces are varied to optimize a defined merit function that indicates the image quality for a given field of view[20]. These merit functions are typically "wild" with many local minima, and there is no guarantee that local or global optimization algorithms will lead to a satisfactory design solution[21]. A successful and widely used optimization-based optical design strategy therefore consists of choosing a well-known optical system as a starting point (e.g. from literature) and steadily achieving incremental improvements. Such a "step-and-repeat" approach to optical design, however, requires considerable experience, intuition and guesswork, which is why it is sometimes referred to as "art and science"[22]. This applies especially to freeform optical systems. With the potential of freeform optics widely recognized and its advantages essential to the creation of next-generation ultra-performing optical systems, it has become a top priority to overcome the laborious and hardly reproducible trial-and-error approaches currently used for their design.

Freeform optical design strategies

So far four main design strategies for freeform optical systems have been proposed and developed. All of them target to guide researchers and optical designers to a sufficiently good freeform design that, if needed, can then be used as starting point for subsequent design optimization.

A first strategy involves the so-called direct design methods, such as the Simultaneous Multiple Surface (SMS) design method[23]. They rely on solving geometrical or differential equations describing the freeform optical system under study to achieve a well-performing initial design[23–31]. Although these methods show clear merits, so far, they lack a straightforward path to increase the number of optical surfaces that can be calculated[25,29,31]. Because of this limitation in scalability, their applicability remained limited until today.

A second strategy focuses on the automated design and advanced optimization methods to generate high-performance freeform systems with reduced efforts by the optical designer[32–37]. Although this strategy aims to provide practical design tools to achieve systems with better optical performance, they do not offer valuable insights in the optical design process. Therefore, the cause for not reaching a satisfactory solution often remains unclear, with the only option to restart the process repeatedly.

A third strategy consists in calculating an initial design that is free of certain aberrations and then rely on optimization techniques to balance all aberrations and yield the best overall imaging performance. One can for example start with an initial rotationally symmetric on-axis design, that has been corrected for several aberrations[38–41], then introduce freeform surfaces and iterate from there, while unobscuring the light path in the system by introducing tilts for the optical surfaces. Alternatively, first-order unobstructed, plane-symmetric systems of three or four spherical surfaces can be calculated as starting points[42,43]. So far, with this design strategy, only a limited number of distinct low-order aberration terms have been canceled or controlled.

A fourth strategy is based on nodal aberration theory[44–46] that has been extended to freeform surfaces[47–49]. This approach allows us to predict and visualize the contribution of Zernike terms of an optical surface to the aberration fields and provides information to the designer which Zernike terms are required to correct an aberration of the system[50,51]. So far, nodal aberration theory is one of the more systematic freeform design strategies. It can guide the experienced designer towards a successful design provided a substantial and in-depth understanding of the underlying theory.

In this paper, we present a novel hybrid direct design method (patent no. WO/2019/129872[52]) for optical imaging systems that also allows us to systematically expand the range of aberration terms that need to be controlled, suppressed, or canceled. The method is holistic because it can be used to design imaging systems that include catoptric and/or dioptric spherical, aspherical, and/or freeform surfaces and components. It allows us to match user-defined conditions such as minimal blurring for each individual order of aberrations by calculating the corresponding coefficients of the optical surfaces, only neglecting order-crossing aberration correlations. As a result, our proposed method allows us to calculate initial imaging systems "first time right" for a given geometry. In that sense, it provides a local and deterministic solution rather than a global solution. Moreover, the method is scalable, as it allows us to add additional optical surfaces in an unrestricted manner. The method also allows the calculation of aberrations of sufficiently high orders to immediately and accurately estimate the overall imaging performance of the system under design. We highlight the substantial added value for the optical design community by demonstrating the systematic, scalable, deterministic, flexible, and holistic character of the design approach with distinct high-end examples, both for catoptric and catadioptric systems. Furthermore, we provide the opportunity for hands-on experience with a design application for imaging systems based on three freeform mirrors.

Materials and methods

Freeform optical systems, aberrations, and their mathematical description

There are various possibilities to classify aberrations in freeform optical systems[53,54]. For the design method

presented in this paper we distinguish three different categories, according to the overall system symmetry present: (1) systems with two orthogonal planes of symmetry, (2) systems with one plane of symmetry, and (3) systems without any symmetry. Here, we will only consider systems with one plane of symmetry as it is the most common category. The two other symmetry categories can be treated similarly.

To describe the transverse ray aberrations (see Fig. 1) in the case of an optical system with the x-z plane of symmetry we follow the systematic description of aberrations of non-rotationally symmetric systems by Barakat and Houston[53]. Following an ansatz similar to that used in most aberration theories, we assume an arbitrary ray that is emitted from an object at infinity with a field direction vector $\mathbf{H} = (\sin(H_x), \sin(H_y), (1\text{-}\sin(H_x)^2\text{-}\sin(H_y)^2)^{1/2})$ for field angles H_x and H_y that passes through the pupil of the system (first surface in Fig. 1) at $(x_p, y_p, f1(\mathrm{x_p},\mathrm{y_p}))$, and that finally intersects the image plane at $\mathbf{h} = (h_x, h_y, D)$. This intersection can be written in vector form as a series expansion $\mathbf{h} = \mathbf{h_0} + \boldsymbol{\epsilon}^{(1)} + \boldsymbol{\epsilon}^{(2)} + \boldsymbol{\epsilon}^{(3)} + \cdots$ where $\mathbf{h_0} = (g(H_x), g(H_y), D)$ denotes the ideal image location as a function of H_x and H_y. Note that H_x and H_y can describe an infinitely distant as well as a finite object where rays are emitted from an object with coordinates $\mathbf{H} = (H_x, H_y, z_0)$. The deviation from the ideal image location is then described by the ray aberration polynomials $\boldsymbol{\epsilon}^{(i)} = (\epsilon_x, \epsilon_y, 0)$ as the sum over the transverse ray aberration coefficients $\epsilon_{x,j,k,l,m}$ and $\epsilon_{y,j,k,l,m}$

$$\epsilon_x(x_p, y_p, H_x, H_y) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\epsilon_{x,j,k,l,m}x_p^j y_p^k H_x^l H_y^m \tag{1}$$

$$\epsilon_y(x_p, y_p, H_x, H_y) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\epsilon_{y,j,k,l,m}x_p^j y_p^k H_x^l H_y^m \tag{2}$$

in the x- and y-directions, respectively, where the sum of the indices $(j + k + l + m)$ denotes the aberration order. It is important to remark that the number of independent and non-vanishing aberration coefficients per order depends on the considered symmetry of the freeform optical system. The transverse aberration series expansions are very similar to "standard" aberrations and thus related to Nodal or vectorial aberration theory so that a conversion of aberration coefficients is possible. For example, all $\epsilon_{j,k,l,m}$ coefficients with $l = m = 0$ can be identified as spherical aberrations, coefficients with $l = 1$, $m = 0$ or $l = 0$, $m = 1$ as comatic aberrations, and so on.

We now consider a sequence of N refractive and/or reflective optical surfaces f_i $(i = 1\ldots N)$ (in Fig.1 these are three mirrors) aligned along a principal ray path. To describe their surface functions, we use a power series representation with coefficients $f_{i,st}$:

$$f_i(x, y) = \sum_{s=0}^{S_i}\sum_{t=0}^{T_i}\frac{1}{s!t!}f_{i,st}x^s y^t \tag{3}$$

In the case of one plane of symmetry (here the x–z-plane), all coefficients in y with odd t vanish. Next, we introduce the ray mapping functions $(u_i(x_p, y_p, H_x, H_y), v_i(x_p, y_p, H_x, H_y))$ in the x- and y-directions that describe where an arbitrary ray, described by variables (x_p, y_p, H_x, H_y), will intersect each optical surface f_i. The ray mapping functions (u_i, v_i) for the ith surface are power series in (x_p, y_p, H_x, H_y) that we write in a similar way as the ray aberration expansions in Eqs. (1) and (2):

$$u_i(x_p, y_p, H_x, H_y) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}u_{i,jklm}x_p^j y_p^k H_x^l H_y^m \tag{4}$$

$$v_i\left(x_p, y_p, H_x, H_y\right) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}v_{i,jklm}x_p^j y_p^k H_x^l H_y^m \tag{5}$$

The chosen principal ray path then determines all vertices of the optical surfaces in a 3D geometry through the series coefficients $(u_{i,0000}, v_{i,0000}, f_{i,00})$. As the surfaces may be

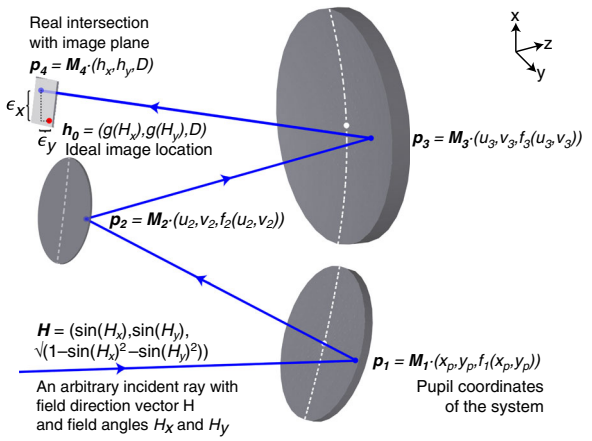

Fig. 1 Schematic layout of a typical plane-symmetric freeform optical system, here, consisting of three mirrors. The plane of symmetry is indicated by the white dotted lines on each surface. The ray path of an arbitrary ray is shown including the introduced mathematical description of the initial ray direction and intersections with each mirror and the local transverse ray aberrations ϵ_x and ϵ_y in the image plane

tilted and thus not share one common optical axis, the principal ray path direction can change from surface to surface. We can define the respective orientation (tilts) of all elements, namely the object, the individual surfaces of the optical elements, and the image plane, through appropriate rotation matrices $\mathbf{M_H}$, $\mathbf{M_i}$ and $\mathbf{M_h}$, applied to $\mathbf{H}$, $_{\mathbf{pi}} = (u_i, v_i, f_i)$ and $\mathbf{h}$, respectively. Consequently, the linear surface coefficients $f_{i,10}$ and $f_{i,01}$ are zero for all surfaces. The ray path from the object to the image plane then consists of $(N+1)$ segments $d_1 \ldots d_{N+1}$ that describe the intermediate optical path length distances for an arbitrary ray. Vector geometry and the Pythagorean theorem enable us to express these distances weighted by the refractive indices $n_{i,i+1}$ of the surrounding materials. For example, the distance d_i is calculated as $n_{i-1,i} \left| \; \right| \; \mathbf{M_i \cdot pi} \; - \; \mathbf{M_i \cdot pi\text{-}1} \left| \; \right|$ where $n_{i-1,i}$ denotes the refractive index of the material between two consecutive surfaces and $||\ldots||$ is the Euclidean norm.

Let us consider a fixed arbitrary point on the object (finite or infinite object distance), and a fixed but arbitrary point on the second surface $\mathbf{p_2}=\mathbf{M_2} \cdot (u_2, v_2, f_2\,(u_2, v_2))$. A ray emerging from the object and passing through $\mathbf{p_1}$ towards $\mathbf{p_2}$ must be such that the total optical path length $d_1 + d_2$ is an extremum according to the modern formulation of Fermat's principle. With points at the boundaries kept fixed, the only remaining variables that can be changed to reach an extremum are u_1 and v_1 at the point $\mathbf{p_1}=\mathbf{M_1} \cdot (u_1, v_1, f_1\,(u_1, v_1))$ on the first (in-between) surface $\mathbf{M_1}$. Fermat's principle thus implies that $D_{1x} = \partial_{u1}(d_1 + d_2) = 0$ and $D_{1y} = \partial_{v1}(d_1 + d_2) = 0$. Following similar arguments, we can derive two sets of differential equations for all defined distances d_i pairwise from object to image space:

$$D_{i,x} = \partial_{ui}(d_i + d_{i+1}) = 0 (i = 1 \ldots N) \tag{6}$$

$$D_{i,y} = \partial_{vi}(d_i + d_{i+1}) = 0 (i = 1 \ldots N) \tag{7}$$

An optical system consisting of N surfaces is thus fully described by N differential equations $D_{i,x}$ and N differential equations $D_{i,y}$ for $i = 1 \ldots N$, for a given but arbitrary pupil plane.

Suppose that all functions h_x, h_y, u_i, v_i, and f_i are analytic and smooth solutions of the differential equations $D_{i,x}$ and $D_{i,y}$, then Taylor's theorem implies that the functions must be infinitely differentiable and have power series representations as defined in Eqs. (1)–(5). We can employ a power series method[55] to find solutions to the derived differential equations $D_{i,x}$ and $D_{i,y}$. This method substitutes the power series of Eqs. (1)–(5) into the differential equations in Eqs. (6) and (7) in order to determine the values of the series coefficients. To calculate certain coefficients, we differentiate Eqs. (6) and (7) with respect to x_p, y_p, H_x, H_y evaluated at $x_p = y_p = H_x = H_y = 0$.

$$\lim_{x_p \to 0} \lim_{y_p \to 0} \lim_{H_x \to 0} \lim_{H_y \to 0} \frac{\partial^j}{\partial x_p^j} \frac{\partial^k}{\partial y_p^k} \frac{\partial^l}{\partial H_x^l} \frac{\partial^m}{\partial H_y^m} D_{i,x} = 0 (i = 1 \ldots N) \tag{8}$$

$$\lim_{x_p \to 0} \lim_{y_p \to 0} \lim_{H_x \to 0} \lim_{H_y \to 0} \frac{\partial^j}{\partial x_p^j} \frac{\partial^k}{\partial y_p^k} \frac{\partial^l}{\partial H_x^l} \frac{\partial^m}{\partial H_y^m} D_{i,y} = 0 (i = 1 \ldots N) \tag{9}$$

It is important to remark that Eqs. (8) and (9) correspond to the x-components and the y-components of the mapping functions and ray aberrations in the local image plane.

Deterministic freeform optical design and system evaluation

A preparatory step in the design of a freeform optical system is to specify the layout, number and types of surfaces to be designed and the location of the stop. If we define for example the ith surface as the stop, then $(u_i, v_i, f_i(u_i, v_i))$ will be replaced by $(x_p, y_p, f_i(x_p, y_p))$ in all Eqs. (3)–(9). The layout of the optical system is defined by the path of the chief ray for the central field. The tilts of optical surfaces are entered using rotation matrices. We can now derive the differential Eqs. (6) and (7) and evaluate Eqs. (8) and (9) for the zero-order case for $j = k = l = m = 0$. This condition is mathematically equivalent to the defined chief ray path and is as such automatically fulfilled. At this stage we can include optical specifications such as the entrance pupil diameter (ENPD), focal length (FL), field of view (FOV) and design wavelength as required by the targeted application.

With the differential equations established and the overall system specifications introduced, two design steps need to be taken: (1) solve the nonlinear first-order case using a standard nonlinear solver or by making use of equivalent first-order optics tools that will provide structurally similar nonlinear equations[42,43]; (2) solve the linear systems of equations in ascending order by setting unwanted aberrations to zero or by minimizing a combination thereof as required by the targeted specifications of the imaging freeform system. These two steps are identical for all freeform optical designs and are implemented as follows:

1. We evaluate Eqs. (8) and (9) for all (i, j, k, l, m) for the first-order case with $j + k + l + m = 1$. This results in a nonlinear system of equations for the second order surface coefficients $f_{i,st}$ with $s+t = 2$, the mapping coefficients $u_{i,1,0,0,0}$, $u_{i,0,0,1,0}$, $v_{i,0,1,0,0}$, $v_{i,0,0,0,1}$ and the aberration coefficients $\epsilon_{x,1,0,0,0}$, $\epsilon_{x,0,0,1,0}$, $\epsilon_{y,0,1,0,0}$, $\epsilon_{y,0,0,0,1}$. The latter are four first-order ray aberration coefficients that can be set to

zero or that can be minimized to calculate the unknown surface and the mapping coefficients. Extra conditions for the second order surface coefficients can be imposed as additional equations if desired. In cases where the number of equations will exceed the number of unknowns, the first-order optical powers of some surfaces can serve as additional initial degrees of freedom. The equations can now be solved using a standard nonlinear solver.

2. For each of the higher orders with $\Omega = j + k + l + m = 2, 3\ldots$, we can determine the exact relationship between the surface, mapping and aberration coefficients by evaluating Eqs. (8) and (9) for all corresponding (i, j, k, l, m). This becomes a linear process once the previous order has been solved. Each linear system of order o then relates the surface coefficients $f_{i,st}$ with $s+t = \Omega+1$ to the mapping coefficients $u_{i,jklm}$, $v_{i,jklm}$ and the aberration coefficients $\epsilon_{x,jklm,}$, $\epsilon_{y,jklm}$ with $j + k + l + m = \Omega$. For each order, we apply the elimination method for solving linear systems to eliminate the unknown mapping coefficients and to obtain a reduced linear system that expresses the aberration coefficients as linear equations of the unknown surface coefficients of that order. The resulting linear system can be either squared, overdetermined or underdetermined. In all cases, we use the Moore-Penrose pseudoinverse (MATLAB's *pinv* function) to solve the linear system for the respective surface coefficients to get the minimum-norm least-squares solution. If we would calculate the least-squares solution for an overdetermined linear system, the aberrations of the considered order would be weighted equally. This, however, would not take the system specifications into account. We therefore define a basic set of weighting factors $\mathrm{WF}_{j,k,l,m} = (\mathrm{ENPD}/2)^{j+k}\left(\mathrm{FOV}/\sqrt{2}\right)^{l+m}$ that is multiplied with the reduced linear equations of same index (j, k, l, m). It is important to stress that an exact but weighted linear least-squares solution[56] is thus very different from commonly used weighting factors in optimization. The weighted least-squares solution for the reduced linear system then takes both the maximum entrance pupil diameter and largest diagonal field diameter into account to simultaneously minimize all properly weighted aberrations for each order, and to calculate the corresponding surface and aberration coefficients. The calculated coefficients are now substituted into the original linear system to calculate the remaining unknown mapping coefficients of that order.

The set of surface coefficients $f_{i,st}$ that have as such been obtained in a deterministic way, now fully describe the N reflective and/or refractive optical surfaces. As such they form a satisfactory "first time right" solution of the free-form system-under-design, while the aberrations and mapping coefficients can be used to evaluate the imaging quality and the geometry of the optical system. Alternative solutions can be further explored, for example by introducing different weighting factors to rebalance the corrected aberrations, e.g. in favor of a wider field of view.

In addition, the proposed power series method enables a function-based analytic ray-tracing evaluation by accurately calculating higher order mapping and aberration coefficients in ascending order and for all combinations of $j + k + l + m = 1, 2, 3\ldots$ up to a desired order that is typically higher than that of the preceding surface coefficient calculations. If not stated otherwise, we always calculate the surface coefficients up to the sixth order, and the mapping and aberration coefficients up to the eighth order, and this throughout the paper. Thus, the full physical behavior of the optical system can be immediately interpreted by the optical designer. The obtained "first time right" solution thus provides an initial imaging system that can be further fine-tuned, for example by using a classical ray-tracing software and applying the embedded advanced optimization algorithms. Figure 2 summarizes the overall design process in a flowchart, which highlights the difference between input parameters for the optical designer (upper row in blue) and the calculated series coefficients (lower row in orange).

Results

Two-dimensional design example for illustration purpose

First, we apply the outlined general process of the previous section to design a Cassegrain-type aplanatic on-axis (all rotation angles are zero) two-mirror system (see Fig. 3a). The two distances between mirror 1 and 2, and mirror 2 and the image plane are 250 and 300 mm, respectively. The focal length (f_L) is 2000 mm and the ENPD is 150 mm. Due to the rotational symmetry of the problem, we can calculate the solution in the x–z-plane. As such the given Eqs. (1)–(5) simplify with partially vanishing sums and indices k=m=t=0. The chosen principal ray path determines the constant terms of the polynomial series in Eqs. (3) and (4) to be $f_{1,00} = u_{1,0000} = u_{2,0000} = 0$ and $f_{2,00} = -250$. The optical path length segments are $d_1 = \boldsymbol{H} \cdot (u_1, f(u_1))$ with $\mathbf{H} = (\sin(H_x), \cos(H_x))$, $d_2 = ||(u_2, f(u_2)) - (u_1, f(u_1))||$ and $d_3 = ||(h_x, D) - (u_2, f(u_2))||$ with $h_x = f_L \tan(H_x) + \epsilon_x(x_p, H_x)$ and $D = 50$. Next, we can derive two differential equations $D_{1,x}$ and $D_{2,x}$ according to Eq. (6) for the defined path length segments and afterwards replace u_1 by x_p to position the stop at the first mirror. The two subsequent calculation steps are executed as follows:

1. We evaluate $\partial x_p^j \partial H_x^l D_{i,x}$ (cf. Eq. (8)) for i=1,2 and for $j = 1, l = 0$ and $j = 0, l = 1$ at $x_p = \mathrm{H_x} = 0$, respectively, which results in four nonlinear

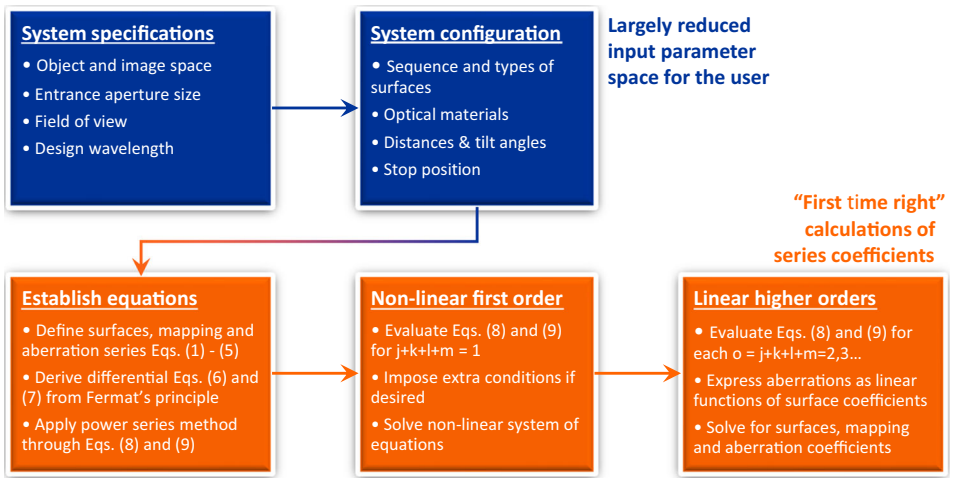

Fig. 2 Flowchart of the "first time right" design process. It highlights the required input by the user (upper row in blue) and the directly calculated coefficients by the design method (lower row in orange).

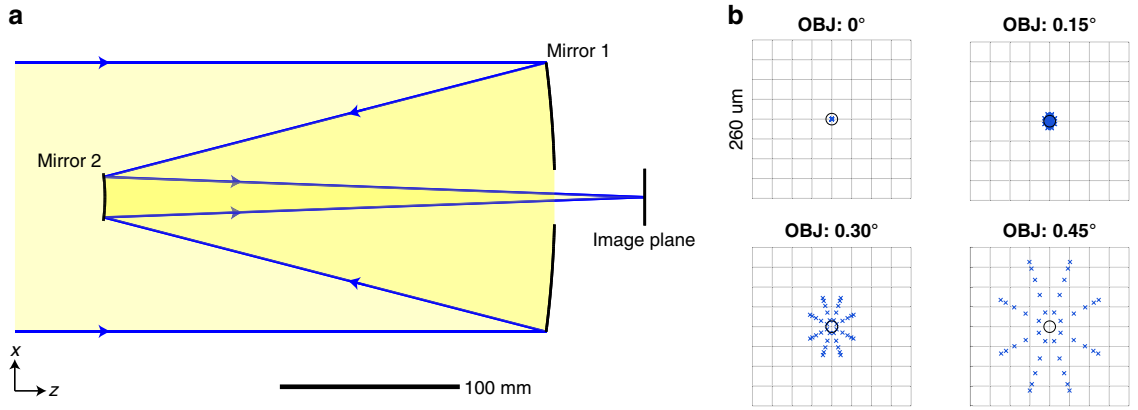

Fig. 3 "First time right" design result for the on-axis aplanatic two-mirror imager. a cross-section of the directly calculated initial system. **b** Corresponding spot diagrams for four selected fields based on ray tracing

algebraic equations. By setting the first-order aberrations $\epsilon_{x,1,0,0,0}$ and $\epsilon_{x,0,0,1,0}$ equal to zero, there are an equal number of four unknown coefficients $f_{1,20}$, $f_{2,20}$, $u_{2,1000}$, and $u_{2,0010}$ to solve for by using MATLAB's *fsolve* function.

2. For each higher order $\Omega = j + l = 2, 3, 4, 5$ we continue to evaluate $\partial x_p^j \partial H_x^l D_{i,x}$ for i = 1,2 and for both $j = \Omega, l = 0$ (spherical aberrations) and $j = \Omega - 1, l = 1$ (comatic aberrations) at $x_p = H_x = 0$, respectively. Thus, we obtain a constant number of four linear equations for every order. By setting the aberration coefficients $\epsilon_{x,\Omega,0,0,0}$ and $\epsilon_{x,(\Omega-1),0,1,0}$ equal to zero for each order, there is an equal number (squared system) of four unknown coefficients of the surfaces $f_{1,(\Omega+1)0}$, $f_{2,(\Omega+1)0}$ and the mapping function $u_{2,\Omega000}$, $u_{2,(\Omega-1)010}$ to immediately solve for. As expected from the imposed on-axis symmetry, all polynomial series coefficients for $\Omega = 2, 4$ vanish.

The aplanatic solution that we calculated is shown in Fig. 3a and the respective spot diagrams after creating a rotationally symmetric system out of the calculated mirror profiles are shown in Fig. 3b.

As expected from our imposed design conditions, both third-order spherical aberration and third-order coma vanish (calculated from Seidel coefficients using Zemax, the file is made available).

Advanced catoptric and catadioptric freeform design examples

To further illustrate the potential of our deterministic direct optical design method, we apply the "first time right" method to two highly advanced state-of-the-art optical freeform systems that were recently developed by

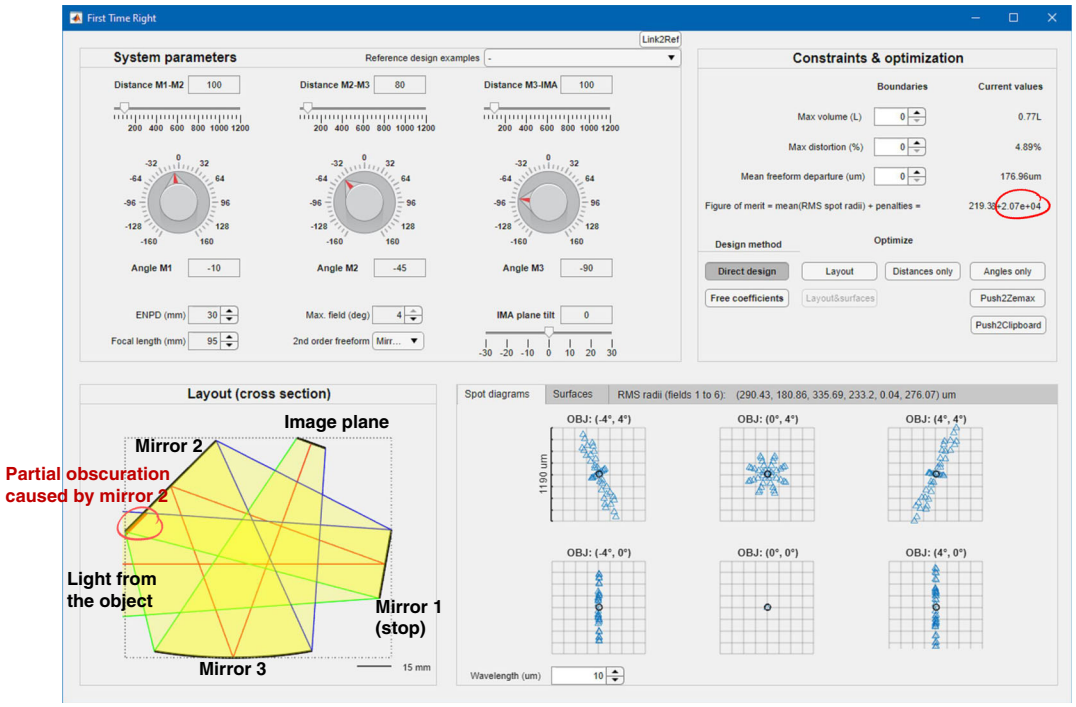

Fig. 4 Graphical user interface of our developed MATLAB Web Application. The overlaid information in the Layout (cross-section) window illustrates the mirror sequence. The partial obscuration caused by mirror 2 (blocking parts of the light from the object) is highlighted and adds a penalty term to the figure of merit. All input and control functions for the user are further explained in the Supplementary Information

expert groups through elaborate design procedures. Further design examples from the literature are provided in the Supplementary Information. To allow readers to follow and evaluate the design method hands-on and in real-time we have programmed and compiled a graphical user interface (GUI) enhanced user application in C++ for deployment to MATLAB Web Application Server. In the shown screenshot in Fig. 4, partial obscuration is caused by mirror 2, which is directly reflected in a penalty term that is added to the figure of merit function. The Supplementary Information provides all necessary additional information on this user application.

The first reference design example is a catoptric imaging system by Bauer et al.[50] for the visible spectrum with x-z-plane symmetry and the object at infinity. It consists of three freeform mirrors with the stop at the first mirror. The targeted system volume is 60 L. The layout of the optical components can be described by three distances (between mirror 1 and 2, mirror 2 and 3, and mirror 3 and the image) and four rotation angles (one for each of the three mirrors, and one for the image), which provides seven degrees of freedom for the optical designer. These distances and angles can now be adjusted to create an unobscured starting geometry that meets certain geometrical constraints such as the given target system volume. This can be done manually and in real time with the user application. Details on geometry selection procedures are given in the Supplementary Information. With the entrance pupil diameter, focal length, field of view and design wavelength specified, we can initiate the approach as described in the previous section:

1. Evaluating Eqs. (8) and (9) for all $(j+k+l+m) = 1$ and $i = 1\ldots3$ results in a nonlinear system of 12 non-vanishing equations with 18 unknowns. Setting the four first-order ray aberration coefficients $\epsilon_{x,1,0,0,0}$, $\epsilon_{x,0,0,1,0}$, $\epsilon_{y,0,1,0,0}$, $\epsilon_{y,0,0,0,1}$ to zero leaves eight mapping coefficients $u_{i,1,0,0,0}$, $u_{i,0,0,1,0}$, $v_{i,0,1,0,0}$, $v_{i,0,0,0,1}$ (i=2,3) and six second order surface coefficients $f_{i,st}$ $(s+t = 2)$ as unknowns. We can define two of the three mirrors to have a base curvature, for example $f_{1,20} = f_{1,02} = c_1$ and $f_{3,20} = f_{3,02} = c_3$, reducing the number of unknown surface coefficients from six to four. The nonlinear system can now be solved.
2. We calculate the surface coefficients up to the sixth order and the mapping and aberration coefficients up to the fifth order. These calculations result in 40 surface coefficients, 250 mapping, and 125 aberration coefficients. We continue the linear solution scheme by further increasing the order of the mapping and aberration coefficients to the eighth

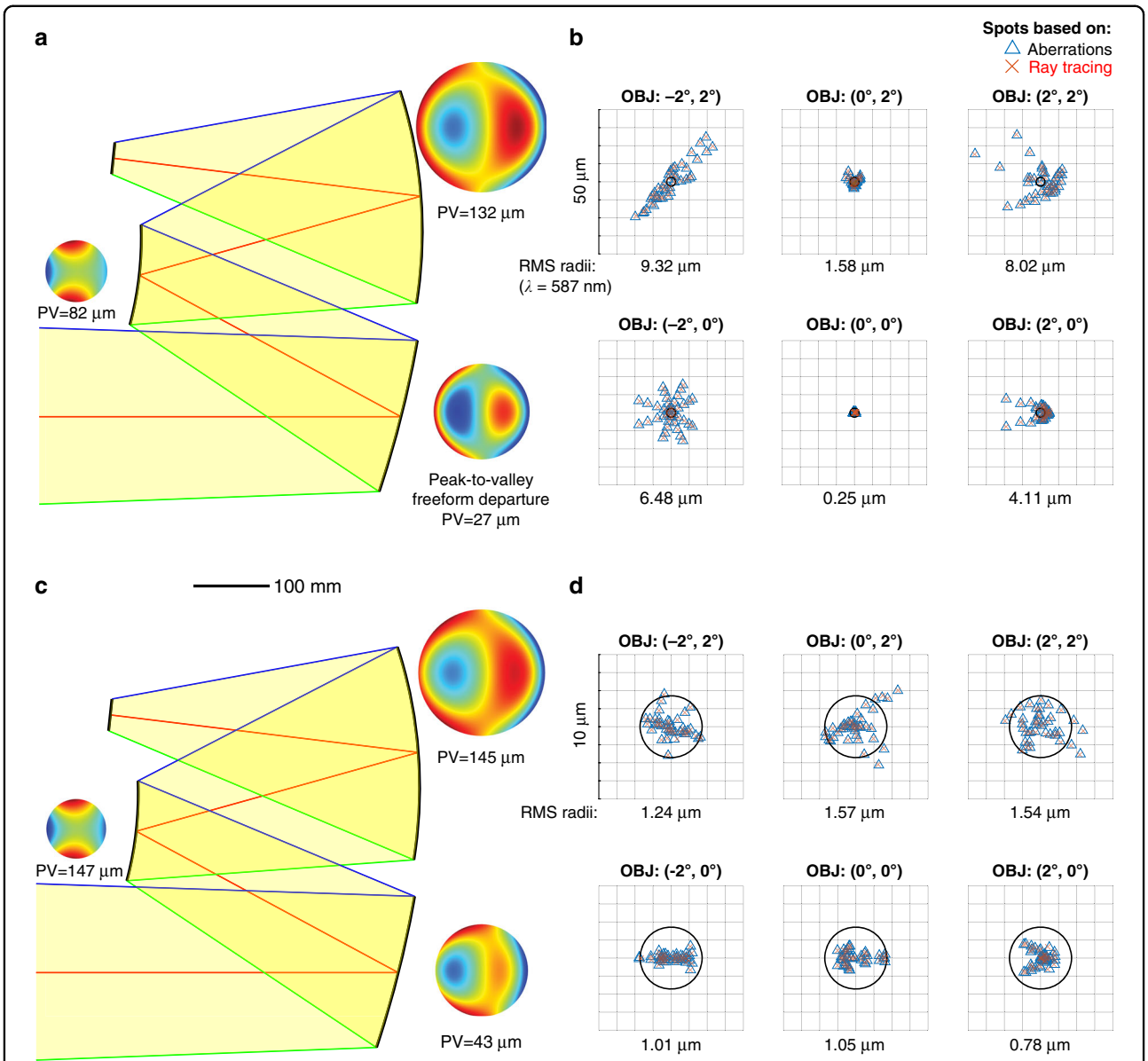

Fig. 5 "First time right" design results for the three-mirror imager. a Cross-section of the directly calculated initial system combined with peak-to-valley freeform departures (PV) from the base sphere for the primary, secondary and tertiary mirror. **b** Corresponding spot diagrams for six selected fields based on aberration calculations (blue triangles) and ray tracing (red crosses) in comparison. **c** Cross-section of the subsequently optimized system combined with peak-to-valley freeform departures (PV) from the base sphere for each mirror. **d** Corresponding spot diagrams for the same six fields based on aberration calculations (blue triangles) and ray tracing (red crosses) in comparison

order (or higher if needed) to ensure a sufficiently accurate prediction of the performance of the system.

The calculated design that we obtained by optimizing the seven degrees of freedom of the geometry (three distances and four angles) is shown in Fig. 5a, where the system layout cross-section is combined with the full 3D peak-to-valley freeform departures (PV) from the best-fit base sphere for each mirror, respectively.

Figure 5b shows the spot diagrams for six selected fields based on our aberration calculations. With an average RMS spot radius of about 5 µm, our directly calculated system provides an already well-corrected "first time right" solution that can be readily further optimized. Next, all forty previously calculated surface coefficients and the initial seven degrees of freedom are used as variables for further optimization, e.g. using MATLAB's *lsqnonlin* solver. After ten iterations, which only take a few minutes

of calculation, the system already reaches diffraction-limited performance close to the reported performance[50] for almost the full field of view. The optimized design shows slightly increased and moderate freeform departures distributed among the three mirrors. The results are shown in Figs. 5c, d accordingly. The aberration-based performance estimation of our method was found to be in excellent agreement with spot diagram data from classical ray tracing (calculated using Zemax, overlaid red cross symbols in Figs. 5b, d).

It is important to emphasize that these results mainly highlight the very effective nature of this new initial freeform design and evaluation tool. Any further final optimization can be performed by using a suitable multivariable optimizer or a classical ray-tracing software with embedded advanced optimization algorithms. Thus, the directly calculated system solution provides an excellent starting point that can be used for subsequent optimization. By using a non-uniform field weighting, and/or a wave front-error-based optimization merit function for the "first time right" design, a diffraction-limited performance can be readily reached.

As stated in the introduction, increasing the number of calculated surfaces is a major issue for most current direct design methods. As a second example from literature[57], we therefore have selected a monolithic freeform objective for a very compact infrared camera with four optical surfaces, two of which are refractive and two reflective. We chose this example to illustrate the scalability of our design method, well beyond the capabilities of most present-day direct freeform design approaches. We furthermore highlight its holistic character, since in this case we are dealing with a catadioptric system with freeform and aspherical surfaces. Here we design an objective consisting of three aspheres and one freeform at the second surface, with the stop placed at the first surface. It was originally designed by the company Jenoptik AG Jena and discussed by Kiontke et al [57,58]. An all-freeform redesign of the monolithic objective has been proposed and discussed by Reshidko et al [59]. We follow these references as closely as possible with the following system requirements: an $F/1.4$ design covering a 37×25-degree field-of-view, an 8.4 mm entrance aperture, made of optical-grade germanium and operating in the long-wave infrared region (LWIR) from 8–12 μm. With the object at infinity, the layout of the system can be described by four distances and five angles, defining the principal ray path from object to image and the respective positions and orientation of all elements and corresponding rotation matrices. It is important to emphasize that the following calculation steps are almost identical to the previous three-mirror example, whereas only the nonlinear first design step is slightly altered:

1. We can evaluate Eqs. (8) and (9) for all $(j+k+l+m)=1$ and for $i=1\ldots4$, which results in a nonlinear system of 16 non-vanishing equations. We set the four first-order ray aberration coefficients $\epsilon_{x,1,0,0,0}$, $\epsilon_{x,0,0,1,0}$, $\epsilon_{y,0,1,0,0}$, $\epsilon_{y,0,0,0,1}$ to zero, leaving 12 mapping coefficients $u_{i,1,0,0,0}$, $u_{i,0,0,1,0}$, $v_{i,0,1,0,0}$, $v_{i,0,0,0,1}$ ($i = 2, 3, 4$) and eight second order surfaces coefficients$f_{i,st}$ ($s+t=2$) as unknowns. Here, three surfaces ($i = 1, 3, 4$) have a base curvature, that is $f_{1,20}=f_{1,02}=c_1$, $f_{3,20}=f_{3,02}=c_3$ and $f_{4,20}=f_{4,02}=c_4$, reducing the number of unknown surface coefficients from eight to five. For the second surface with $i = 2$, a freeform surface, we define the mean base curvature as $c_3=(f_{3,20}+f_{3,02})/2$, which allows us to apply Petzval's curvature condition[54] for the four curvatures c_1 to c_4 as a 17th equation. We now solve the nonlinear system.
2. We calculate the surface coefficients up to sixth order for the surfaces and up to eighth order for the mapping and aberration coefficients.

In order to use a volume close to the reference systems, we constrained the on-axis chief ray path length within the material to 58 mm, a value we estimated through manual measurements. Optimizing the initial degrees of freedom of the geometry yields an already well-corrected and unobscured system. Figure 6a shows the configuration of the system of comparable size to the cited reference designs. The peak-to-valley freeform departure (PV) from the best-fit base sphere for the second freeform surface is added next to it. Figure 6b shows the corresponding spot diagrams for six selected fields (we used 15 fields for the optimization) based on our aberration calculations. With an average RMS spot radius of about 51 μm, we achieved an excellent starting point for further optimization. All surface coefficients are now set as variables for further optimization using e.g. MATLAB's *lsqnonlin* solver. The results are shown in Figs. 6c, d.

After ten iterations and few minutes, the system reaches a close to diffraction-limited performance in the LWIR band for the full field of view. Chromatic aberrations are very well controlled due to the relatively low dispersion of Germanium in the LWIR band (the refractive index ranges from 4.0067 at 8 μm to 4.0029 at 12 μm) compared to common glasses in the visible band. In addition, the relatively small angles within the monolith due to the high refractive index lead to a better aberration control in general. This spot diagram-based performance is in excellent agreement with spot diagram data from classical ray tracing and seems to exceed the reported performances of the reference systems[57,59]. However, as the authors did not reveal all relevant information such as the nominal focal length, the acceptable distortion (here, less than 5%) or the target volume of their systems, a truly quantitative comparison was not possible.

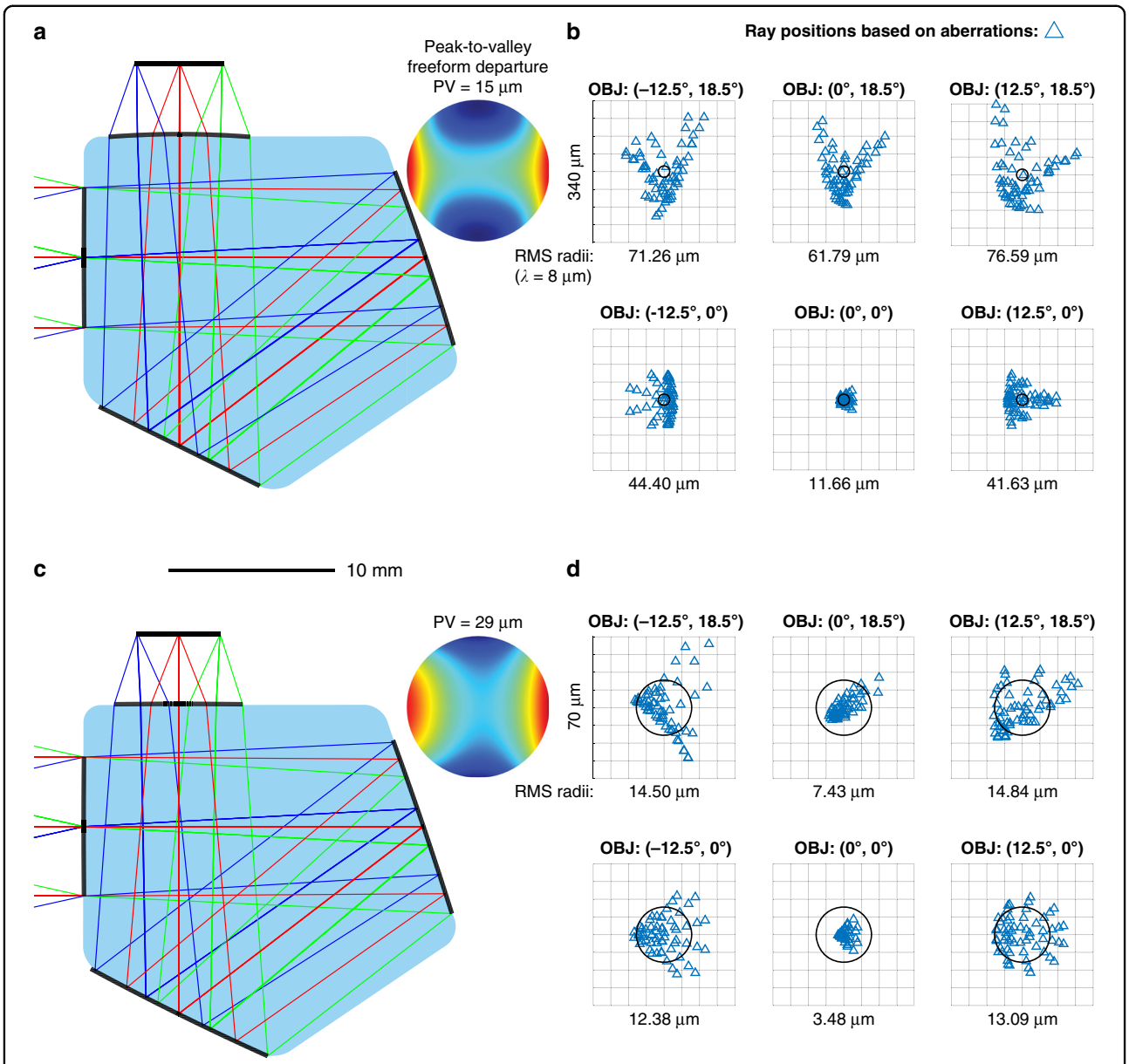

Fig. 6 "First time right" design results for the four-surfaces monolithic infrared camera objective. a Cross-section of the system from direct calculations with peak-to-valley freeform departure (PV) from the base sphere for the second and only freeform surface. **b** Corresponding spot diagrams for six selected fields based on aberration calculations. **c** Cross-section of the subsequently optimized system combined with peak-to-valley freeform departure (PV) from the base sphere for the second surface. **d** Corresponding spot diagrams for the same six fields based on aberration calculations

A comparison with the catoptric three surfaces design example makes clear that the only differences with the catadioptric system designed here are the altered set of underlying Eqs. (6)–(9) and the adapted solution of the nonlinear first-order equations. Everything else remains the same no matter what types or how many surfaces are used. Important to remark here is that all orders higher than the first-order result in linear systems of equations that can be fast and very efficiently transformed and solved.

This example clearly demonstrates again the highly effective nature of our proposed design and evaluation method. In addition, it highlights two further important features: (1) our method allows us to simply combine refractive and/or reflective surfaces of spherical, aspheric or freeform shapes; (2) the method also straightforwardly enables to increase the number of calculated surfaces of an optical system without considerably increasing the computational complexity of the problem. This direct

path to scaling the number of calculated surfaces is a result of the fact that the intermediate optical path length distance expressions do not depend on the number of considered surfaces. Thus, all intermediate optical path length distances can be implemented in a modular fashion and reused even when the system is changed (e.g., number of surfaces).

Discussion

Equipping lens- and/or mirror-based optical systems with freeform optical surfaces makes it possible to deliver highly original imaging functionalities with superior performance compared to their more traditional (a)spherical counterparts, such as enhanced field-of-view, increased light-collection efficiencies, larger spectral band, and higher compactness. Until now mathematical models and design strategies for freeform optics remained limited and failed to provide deterministic solutions. In particular, the identification of a suitable initial design has often proven to be a painstaking and time-consuming trial-and-error process.

In this paper, we reported on the first deterministic direct design method for freeform optical systems that is not restricted by the aberration terms that can be controlled. The method relies on Fermat's principle and allows a highly systematic generation and evaluation of directly calculated freeform design solutions that can be readily used as starting point for further and final optimization. As such, this new method allows the straightforward generation of "first time right" initial designs that enable a rigorous, extensive and real-time evaluation in solution space when combined with available local or global optimization algorithms.

We can summarize the main features of our new method as follows:

1. *Holistic*: The method can be used to design imaging systems that include catoptric and/or dioptric spherical, aspherical, and/or freeform surfaces and components. It works for all present aberrations in freeform optical systems up to a desired order.
2. *Deterministic*: As many current design strategies strongly rely on multi-step optimization routines, final design results often differ considerably for slightly different starting points. The here presented method allows a fast and deterministic solution as well as a very systematic evaluation of the solution space while providing more detailed insights into the fundamental physics of the optical system than most conventional design approaches.
3. *Scalable:* So far, several direct design methods have been reported for freeform optical systems. However, these methods have in common, that they are developed for a specific system layout, e.g. two or three freeform mirrors. Once a designer would like to add one or more mirrors to the design, the methods do not typically scale with the number of calculated optical surfaces. In contrast, the here presented method follows a clear set of rules that allows us to add additional refractive and/or reflective surfaces to a system.
4. *All-round:* the method can be beneficial for "junior" and "senior" optical designers alike. The in-depth insights into aberrations with this approach are very valuable for skilled senior optical designers. Inexperienced optical designers with a less advanced understanding of aberration theory on the other hand can equally benefit from it, as this in-depth insight is not a must-have to make excellent use of the method. Finally, the drastic reduction in parameter space, i.e. the large ratio of calculated coefficients over input coefficients, make this method very attractive for designers with a strong background in optimization.

The deterministic, holistic, and scalable nature of this all-round method therefore has the full potential to create a true paradigm shift in how freeform imaging systems will be designed and developed from now on to suit a wide range of applications.

Author contributions

F.D. conceived the idea and developed the 'first time right' method and algorithms. F.D. and H.T. wrote the paper. Both authors contributed to the discussion and preparation of the manuscript.

Funding

In part by VUB-OZR and FWO fellowship 12C9817N [Duerr]. In part by BELSPO-IAP, IOF, Hercules/FWO and the Methusalem foundations [Thienpont].

Conflict of interest

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41377-021-00538-1.

Received: 26 October 2020 Revised: 7 April 2021 Accepted: 18 April 2021
Published online: 06 May 2021

References

1. Maddison, F. Early astronomical and mathematical instruments: a brief survey of sources and modern studies. *Hist. Sci.* **2**, 17–50 (1963).
2. Turner, G. L. The history of optical instruments: a brief survey of sources and modern studies. *Hist. Sci.* **8**, 53–93 (1969).
3. Johnson, R. B. A historical perspective on understanding optical aberrations. *Proc. SPIE* **10263**, (1992).
4. Fang, F. Z. et al. Manufacturing and measurement of freeform optics. *CIRP Ann.* **62**, 823–846 (2013).
5. Gissibl, T. et al. Two-photon direct laser writing of ultracompact multi-lens objectives. *Nat. Photonics* **10**, 554–560 (2016).
6. Gissibl, T. et al. Sub-micrometre accurate free-form optics by three-dimensional printing on single-mode fibres. *Nat. Commun.* **7**, 11763 (2016).
7. Hong, Z. H. & Liang, R. G. IR-laser assisted additive freeform optics manufacturing. *Sci. Rep.* **7**, 7145 (2017).
8. Thompson, K. P. & Rolland, J. P. Freeform optical surfaces: a revolution in imaging optical design. *Opt. Photonics N.* **23**, 30–35 (2012).
9. Reimers, J. et al. Freeform spectrometer enabling increased compactness. *Light.: Sci. Appl.* **6**, e17026 (2017).
10. Jahn, W., Ferrari, M. & Hugot, E. Innovative focal plane design for large space telescope using freeform mirrors. *Optica* **4**, 1188–1195 (2017).

11. West, G. J. & Howard, J. M. Application for freeform optics at NASA. https://ntrs.nasa.gov/search.jsp?R=20170010419 (2017).
12. Zou, Y. C. et al. Miniature adjustable-focus endoscope with a solid electrically tunable lens. *Opt. Express* **23**, 20582–20592 (2015).
13. Li, J. W. et al. Two-photon polymerisation 3D printed freeform micro-optics for optical coherence tomography fibre probes. *Sci. Rep.* **8**, 14789 (2018).
14. Kress, B. C. *Optical Architectures for Augmented-, Virtual-, and Mixed-Reality Headsets* (SPIE Press, 2020).
15. Wagner, C. & Harned, N. EUV lithography: lithography gets extreme. *Nat. Photonics* **4**, 24–26 (2010).
16. Williamson, D. M. Freeforms in EUV lithography projection optics. In *Proceedings of Freeform Optics 2015* FM3B.4 (Optical Society of America, 2015) https://www.osapublishing.org/abstract.cfm?uri=freeform-2015-FM3B.4.
17. Wills, S. Freeform optics: notes from the revolution. *Opt. Photonics N.* **28**, 34–41 (2017).
18. The Scott Partnership. Optical simulation software. *Nat. Photonics* **4**, 256–257 (2010).
19. Sahin, F. E. Open-source optimization algorithms for optical design. *Optik* **178**, 1016–1022 (2019).
20. Feder, D. P. Automatic optical design. *Appl. Opt.* **2**, 1209–1226 (1963).
21. van Turnhout, M. & Bociort, F. Instabilities and fractal basins of attraction in optical system optimization. *Opt. Express* **17**, 314–328 (2009).
22. Shannon, R. R. *The Art and Science of Optical Design* (Cambridge University Press, 1997).
23. Gimenez-Benitez, P. et al. Simultaneous multiple surface optical design method in three dimensions. *Optical Eng.* **43**, 1489–1502 (2004).
24. Liu, J. Y., Benítez, P. & Miñano, J. C. Single freeform surface imaging design with unconstrained object to image mapping. *Opt. Express* **22**, 30538–30546 (2014).
25. Duerr, F. et al. Analytic free-form lens design in 3D: coupling three ray sets using two lens surfaces. *Opt. Express* **20**, 10839–10846 (2012).
26. Duerr, F. et al. Analytic design method for optimal imaging: coupling three ray sets using two free-form lens profiles. *Opt. Express* **20**, 5576–5585 (2012).
27. Hicks, R. A. Controlling a ray bundle with a free-form reflector. *Opt. Lett.* **33**, 1672–1674 (2008).
28. Yang, T. et al. Design method of freeform off-axis reflective imaging systems with a direct construction process. *Opt. Express* **22**, 9193–9205 (2014).
29. Volatier, J. B. & Druart, G. Differential method for freeform optics applied to two-mirror off-axis telescope design. *Opt. Lett.* **44**, 1174–1177 (2019).
30. Volatier, J. B., Duveau, L. & Druart, G. An exploration of the freeform two-mirror off-axis solution space. *J. Phys.: Photonics* **2**, 014004 (2020).
31. Croke, C. B. & Hicks, R. A. Solution to the bundle-to-bundle mapping problem of geometric optics using four freeform reflectors. *J. Optical Soc. Am. A* **31**, 2097–2104 (2014).
32. Yang, T., Jin, G. F. & Zhu, J. Automated design of freeform imaging systems. *Light.: Sci. Appl.* **6**, e17081 (2017).
33. Xu, C. et al. Automatic optical path configuration variation in off-axis mirror system design. *Opt. Express* **27**, 15251–15261 (2019).
34. Carneiro de Albuquerque, B. F., Luis de Sousa, F. & Montes, A. S. Multi-objective approach for the automatic design of optical systems. *Opt. Express* **24**, 6619–6643 (2016).
35. Wu, R. M., Sasián, J. & Liang, R. G. Algorithm for designing free-form imaging optics with nonrational B-spline surfaces. *Appl. Opt.* **56**, 2517–2522 (2017).
36. Chrisp, M. P., Primeau, B. & Echter, M. A. Imaging freeform optical systems designed with NURBS surfaces. *Optical Eng.* **55**, 071208 (2016).
37. Houllier, T. & Lépine, T. Comparing optimization algorithms for conventional and freeform optical design. *Opt. Express* **27**, 18940–18957 (2019).
38. Korsch, D. *Reflective Optics* 207–259 (Academic Press, 1991).
39. Rakich, A. & Rumsey, N. Method for deriving the complete solution set for three-mirror anastigmatic telescopes with two spherical mirrors. *J. Optical Soc. Am. A* **19**, 1398–1405 (2002).
40. Nie, Y. F. et al. Freeform optical design for a nonscanning corneal imaging system with a convexly curved image. *Appl. Opt.* **56**, 5630–5638 (2017).
41. González-Acuña, R. G., Chaparro-Romo, H. A. & Gutiérrez-Vega, J. C. Analytic solution of the Eikonal for a stigmatic singlet lens. *Phys. Scr.* **95**, 085201 (2020).
42. Howard, J. M. & Stone, B. D. Imaging with three spherical mirrors. *Appl. Opt.* **39**, 3216–3231 (2000).
43. Howard, J. M. & Stone, B. D. Imaging with four spherical mirrors. *Appl. Opt.* **39**, 3232–3242 (2000).
44. Thompson, K. Description of the third-order optical aberrations of near-circular pupil optical systems without symmetry. *J. Optical Soc. Am. A* **22**, 1389–1401 (2005).
45. Thompson, K. P. Multinodal fifth-order optical aberrations of optical systems without rotational symmetry: spherical aberration. *J. Optical Soc. Am. A* **26**, 1090–1100 (2009).
46. Thompson, K. P. Multinodal fifth-order optical aberrations of optical systems without rotational symmetry: the comatic aberrations. *J. Optical Soc. Am. A* **27**, 1490–1504 (2010).
47. Fuerschbach, K., Rolland, J. P. & Thompson, K. P. Extending Nodal Aberration Theory to include mount-induced aberrations with application to freeform surfaces. *Opt. Express* **20**, 20139–20155 (2012).
48. Yang, T., Zhu, J. & Jin, G. F. Nodal aberration properties of coaxial imaging systems using Zernike polynomial surfaces. *J. Optical Soc. Am. A* **32**, 822–836 (2015).
49. Shi, H. D. et al. Analysis of nodal aberration properties in off-axis freeform system design. *Appl. Opt.* **55**, 6782–6790 (2016).
50. Bauer, A., Schiesser, E. M. & Rolland, J. P. Starting geometry creation and design method for freeform optics. *Nat. Commun.* **9**, 1756 (2018).
51. Zhong, Y. & Gross, H. Initial system design method for non-rotationally symmetric systems based on Gaussian brackets and Nodal aberration theory. *Opt. Express* **25**, 10016–10030 (2017).
52. Duerr, F. & Thienpont, H. Optical design methods for imaging systems and optical systems designed therewith. Patent No. WO/2019/129872 (2019).
53. Barakat, R. & Houston, A. The aberrations of non-rotationally symmetric systems and their diffraction effects. *Opt. Acta.: Int. J. Opt.* **13**, 1–30 (1966).
54. Gross, H. et al. *Handbook of Optical Systems: Volume 3: Aberration Theory and Correction of Optical Systems* (Wiley-VCH, 2007).
55. Said-Houari, B. in *Differential Equations: Methods and Applications* (ed Said-Houari, B.) 125–140 (Springer, 2015).
56. Lawson, C. L. & Hanson, R. J. *Solving Least Squares Problems* (Society for Industrial and Applied Mathematics, 1995), https://doi.org/10.1137/1.9781611971217.
57. Kiontke, S. R. Monolithic freeform element. *Proc. SPIE* **9575**, 95750G (2015).
58. Fuchs, U. & Kiontke, S. R. *Proc. SPIE* **9948**, 99480L (2016).
59. Reshidko, D. & Sasián, J. Method for the design of nonaxially symmetric optical systems using free-form surfaces. *Opt. Eng.* **57**, 101704 (2018).