UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN**

TESIS DOCTORAL

**SMS FOR IMAGING SYSTEMS
USING FREE-FORMS**

Milena Nikolić
Ingeniera en Electrónica

2017

UNIVERSIDAD POLITÉCNICA DE MADRID

Instituto de Energía Solar

Departamento de Electrónica Física

Escuela Técnica Superior de Ingenieros de Telecomunicación

TESIS DOCTORAL

**SMS FOR IMAGING SYSTEMS
USING FREE-FORMS**

AUTOR: Milena Nikolić
DIRECTOR: Pablo Benítez Giménez
TITULACIÓN: Ingeniera en Electrónica

2017

**Tribunal nombrado por el Magfco. Y Excmo. Sr. Rector de la
Universidad Politécnica de Madrid.**

PRESIDENTE:

VOCALES:

SECRETARIO:

SUPLENTE:

Realizado el acto de defensa y lectura de la Tesis en Madrid,
el día___ de _____ de 2017.

Calificación:

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer al profesor Juan Carlos Miñano y al profesor Pablo Benítez por haberme dado la oportunidad de formar parte del grupo de óptica de Cedint-UPM. En esta tesis se ha resumido una parte del amplio trabajo que ha sido desarrollado durante varios años en el grupo. Esto no hubiera sido posible sin Pablo Benítez. A él le agradezco toda la confianza ofrecida en mi formación. Durante todo este tiempo ha tenido mucha paciencia ante mis continuas dudas y con gran humildad ha dirigido y guiado los trabajos de esta tesis. Ha sido un honor y un placer aprender de él.

Me gustaría agradecer a todos mis compañeros del grupo de óptica. A toda la gente que ha pasado por el grupo durante estos años, Hamed, Mao, Yingfei, Jayao Liu, Wang Lin, Guille, José, Profesor y muchos más que en algún momento han formado parte del grupo. A mis compañeros de despacho, Bharat, Joao y Jesús por crear un ambiente en el que daba gusto trabajar. A Edu, que no siendo del grupo de óptica sí que lo era. A toda la gente de Limbak: Pablo, Dejan, Marina; LPI: Rúben, Maikel, Blen, Juan, Simone, Julio, Fernanda y en general a toda la gente de Cedint. A Sasha por introducirme a este grupo e indirectamente influir en numerosas decisiones precedentes que me han traído aquí.

Gracias a todos los que forman parte de proyecto Adopsys, a Rafael, Lijo, Ton, Nenad, Olga, Priya, Yunfeng, Marco, Adam, Simone y especialmente a Zhe, a quien le agradezco toda la ayuda en el trabajo realizado en el capítulo 3 de esta tesis. A toda la gente del grupo de óptica de TU Delft, por hacer mi estancia allí muy agradable. Al profesor Florian Bociort y al profesor Paul Urbach por varias discusiones sobre SMS y sobre el diseño óptico en general. A todo el equipo de Carl Zeiss CRT Oberkochen, al dr. Bernd Kleeman y Marco Mout por toda su ayuda durante los tres meses pasados en Aalen y Oberkochen. Me gustaría agradecer además al dr. Wolfgang Singer, por su ayuda con la especificación de tolerancias.

Una mención especial para José Infante, por ayudarme en mis primeros pasos con el CODE V® y en general en óptica de formación de imagen. Y además por darme la oportunidad de hacer una parte de las medidas de resolución en los laboratorios de Escribano. Debo agradecimientos a todo el equipo de Escribano, y especialmente a Javier Muñoz, por responder a todas mis dudas (que no han sido pocas) sobre las medidas de resolución.

Y muy especialmente a Jorge, a mis padres y mi hermana, a toda mi familia (tanto la serbia como la española) y amigos por todo su amor y apoyo.

RESUMEN

El diseño de sistemas ópticos de formación de imagen goza de una historia larga e ilustre. Desde los primeros magnificadores ópticos del antiguo Egipto, pasando por los telescopios de Galileo y Newton al comienzo de la ciencia moderna, hasta la ubicuidad de la óptica actual, el proceso de diseño óptico se ha considerado tanto un arte como una ciencia. El diseño óptico ha evolucionado desde la teoría de óptica paraxial de la época de Newton hasta los modernos diseños realizados mediante la utilización de diferentes técnicas de optimización multiparamétrica. Aunque los métodos mencionados pueden proporcionar soluciones elegantes para muchos de los problemas tradicionales, la creciente demanda de novedosos sistemas ópticos con superficies no-esféricas y geometrías fuera de eje hace que los problemas de diseño ya no puedan resolverse fácilmente utilizando esos métodos tradicionales.

Con el desarrollo acelerado de las nuevas técnicas de fabricación se ha llegado a la omnipresencia de las superficies asféricas en los diseños ópticos actuales. Sin embargo, la creciente complejidad de los problemas ópticos relacionados con la formación de imagen hace que hagan falta incluso más grados de libertad que los que propone la óptica esférica. Esos grados de libertad se pueden lograr, bien incrementando el número de superficies del sistema, lo cual a menudo lleva a una solución voluminosa no apropiada para muchas aplicaciones, o abandonando la simetría de revolución del sistema, lo cual permite la introducción de nuevas formas de superficies sin simetría de revolución, conocidas como superficies anamórficas (*“free-form”*).

Las superficies anamórficas aparecieron por primera vez en los diseños de óptica anidólica (*“nonimaging”*), donde el objetivo no es la formación de imagen, sino la transferencia eficiente de la luz. Dado que los sistemas anidólicos no son igual de sensibles a errores de fabricación como los de formación de imagen, muchos de los diseños ópticos basados en superficies anamórficas se han convertido en productos comerciales. De esa forma se ha hecho evidente el potencial de este tipo de superficies para proporcionar soluciones altamente compactas.

En las técnicas de optimización multiparamétrica, que suponen un elevado número de parámetros libres, el sistema inicial de partida es la clave para encontrar la solución óptima. En sistemas ópticos tradicionales como, por ejemplo, el objetivo de una cámara, existe una literatura inmensa que incluye muchas patentes y catálogos que pueden proporcionar el

sistema inicial de partida. Sin embargo, en el caso de nuevos sistemas como los de realidad virtual (o aumentada), conocidos como sistemas HMD (del inglés “*head-mounted-display*”), no existe mucha literatura que permita proporcionar sistemas de partida. Para tales sistemas, los métodos directos representan una útil herramienta, especialmente en diseños que incluyen un elevado número de parámetros que especifican la superficie. Cuanto más alto sea el número de parámetros, más difícil será encontrar una solución utilizando los métodos basados en la optimización multiparamétrica, por lo que la elección del sistema de partida en esos casos resulta ser clave. El método de Superficies Múltiples Simultaneas (SMS), que nació como una herramienta de cálculo para aplicaciones anidólicas, es un método directo de diseño tanto de superficies esféricas como anamórficas. Recientemente se ha demostrado además como una herramienta para el diseño de sistemas ópticos de formación de imagen, y puede utilizarse para generar los sistemas de partida necesarios para la optimización multiparamétrica.

Esta tesis doctoral investiga nuevas estrategias de diseño de sistemas de superficies no-esféricas. Consta de 2 partes y 7 capítulos. Los conceptos básicos de la óptica geométrica se introducen concisamente en el primer capítulo. Los capítulos 3-7 se agrupan en 2 partes: la óptica esférica (capítulos 3 y 4) y la óptica anamórfica (capítulos 5-7). El capítulo 2 está relacionado con ambas partes de la tesis, ya que proporciona una nueva percepción del método SMS en general.

En el capítulo 2 se presenta un nuevo desarrollo del método SMS que cambia la forma en la que el método ha sido considerado hasta ahora. Se demuestra que para ciertos casos específicos de diseño con la apertura alejada de las superficies se pueden crear imágenes estigmáticas de múltiples puntos de objeto utilizando un menor número de superficies. Avanzando paso a paso desde las bases del método, pasamos a la formación de imagen de dos puntos objeto a través de una única superficie. Posteriormente, aumentando de dos a tres puntos objeto incrementando el paralelismo, llegamos a un diseño de seis superficies cumpliendo una formación de imagen estigmática para ocho puntos objeto. Se presenta un criterio aproximado que proporciona una estimación del número de puntos objeto que se pueden enfocar en el plano imagen (estigmáticamente), que utiliza para ello el tamaño de la apertura del sistema y de las superficies. En el proceso de maximización de ese número se intentan agotar los grados de libertad de las superficies esféricas y anamórficas. Al final se deduce que los diseños SMS con los grados de libertad agotados están muy cerca de la solución óptima, por tanto usarlos como puntos de partida para la optimización resultaría en una convergencia rápida hacia el sistema óptimo.

El capítulo 3 presenta un estudio del método SMS2D como el punto de partida para un problema de diseño de dos lentes esféricas. Se analizan dos ejemplos de diseño en los cuales se compara el método SMS (con la optimización posterior) y las técnicas tradicionales, tanto como la optimización global. Los métodos de diseño tradicionales utilizan un punto de partida esférico con la optimización de un paso único, de dos pasos y el método paso-a-paso, conocido como “*stepwise*”, mientras el sistema SMS se optimiza con un único paso. Se demuestra que, dado

que el método SMS aporta un punto de partida superior, la optimización con un único paso resulta ser suficiente para llegar a un diseño óptimo.

El capítulo 4 introduce un concepto novedoso en el diseño de óptica de formación de imagen conocido como "*foveated imaging*". *Foveated imaging* o formación de imagen adaptada es un método que emula el comportamiento del ojo humano mediante la generación de imágenes con resolución espacial variable. La primera parte de este capítulo investiga la posibilidad de utilizar el método SMS con discretización en pupila para diseñar sistemas ópticos adaptados. Utilizando la discretización en pupila el problema no está limitado por el número de puntos objeto necesarios para conseguir el mapeo deseado. Partiendo de un sistema de objetivo de microscopio ya existente, la parte asférica del sistema se ha rediseñado utilizando el método mencionado. Se optimizan y analizan los diversos conjuntos de superficies obtenidos para evaluar el nivel de complejidad de las superficies asféricas comparado con las superficies del diseño inicial. El mismo enfoque se aplica también a un diseño HMD, donde se diseñan cuatro superficies asféricas para cumplir con el requisito de adaptación (*foveation*). Aparte de la adaptación, este diseño tiene en cuenta las rotaciones del ojo, un enfoque comúnmente utilizado en el diseño de lentes oftálmicas. Combinando estos dos métodos, se diseña una lente HMD adaptada a la resolución del ojo humano. Adicionalmente, se presenta un análisis de tolerancias para este sistema, así como las medidas experimentales de la resolución del prototipo fabricado.

El capítulo 5 investiga un sistema afocal de dos espejos anamórficos utilizando óptica de primer orden. Las ecuaciones diferenciales parciales de primer orden resultantes tienen una solución analítica, imponiendo la única condición de que los ejes x - y y x' - y' son paralelos. Se presentan cuatro soluciones seleccionadas. Dos de ellas son semi-aplanáticas (cumpliendo la condición de aplanatismo sólo para la coordenada x), mientras que las otras dos son, de acuerdo a nuestro conocimiento, los primeros ejemplos de sistemas aplanáticos (*full aplanat*) que constan de dos espejos anamórficos. El problema presentado también puede considerarse como un caso límite de un problema SMS, como se comenta en el texto. Adicionalmente se han investigado geometrías de sistema alternativas. Sin embargo, en estos casos las ecuaciones de mapeo rayo-a-rayo no están necesariamente desacopladas, como en el primer caso discutido.

Otra contribución al método SMS se encuentra en el capítulo 6. Se presenta un algoritmo para diseñar 3 superficies anamórficas plano-simétricas utilizando el método SMS3D. Se presenta también un ejemplo de un sistema de este tipo, que forma imagen estigmática para los campos de diseño. Por otro lado se demuestra que en el caso límite en el que los tres frentes de onda de diseño coinciden, el ejemplo presentado es efectivamente un sistema semi-aplanático, donde se cumple el aplanatismo hasta el segundo orden en sólo una dirección.

El capítulo 7 compara diferentes polinomios ortogonales como representaciones de las superficies anamórficas en el caso de sistemas ópticos con aperturas rectangulares. Los diseños SMS3D que están destinados a aplicaciones de formación de imagen necesitan ser optimizados para el campo entero. Para ello es necesario elegir una representación de las superficies apropiada. Estudios recientes demuestran que la mejor representación en el caso de los

sistemas con aperturas circulares es la introducida por Forbes, y conocida como Q-polinomios. Sin embargo, las superficies anamórficas se usan a menudo en sistemas ópticos con aperturas no circulares (sistemas de alta relación de aspecto, sistemas de pantallas digitales). Se proporciona un análisis comparativo realizado en 2 sistemas de espejos. Se comparan 4 representaciones diferentes, incluyendo polinomios con diferentes áreas de ortogonalización (rectangulares y circulares) y diferentes métricas (ortogonalidad en el sag y en el gradiente). En los ejemplos considerados, los polinomios ortogonales dentro de un rectángulo convergen más rápidamente o hacia un mejor mínimo en comparación con los polinomios ortogonales dentro de un círculo. Esto es, lo más probable, debido a la pérdida de las buenas propiedades de la ortogonalidad cuando el área de ortogonalidad no coincide con la forma de la apertura de la superficie utilizada.

ABSTRACT

The design of imaging optical systems has a long and illustrious history. From the first magnifiers in the ancient Egypt, to the telescopes of Galileo and Newton at the commencement of modern science, to the omnipresence of modern optics, the process of optical design was considered both an art and a science. Optical design has evolved from paraxial imaging theory in Newton's age to modern computer-based multi-parametric optimization techniques. Although these methods can provide elegant solutions to many traditional design problems, the increasing demand for novel optical systems with non-spherical surfaces and off-axis geometry take the design problems to the level where they are no longer easily solved with these traditional methods.

Rapid development of the manufacturing techniques led to the ubiquity of aspheric surface shapes in modern optical designs. However, with the increasing complexity of imaging design problems, even more degrees of freedom are needed. They can be achieved either by increasing the number of surfaces in the system, which often leads to a bulky solution not appropriate for many applications, or by abandoning traditional rotational symmetry of the system, which allows for new non-rotationally symmetric surface shapes, known as freeforms.

Freeform surfaces were first used in non-imaging design problems, where the aim is not to form an image, but to maximize the energy transfer efficiency. Since non-imaging systems are not as sensitive to the manufacturing errors as the imaging ones, many freeform optical systems became commercial products. The potential of these surfaces in providing highly compact solutions thus became clear.

In multi-parametric optimization techniques with the large number of free parameters, starting system proves to be the key in finding the optimal solution. In traditional optical systems, such as for instance camera objectives, many starting systems are available in patent literature. However, in the case of new optical systems, such as head mounted display systems, no such literature exists. For such systems direct design methods present a useful tool, especially in designs with elevated number of surface parameters. The higher the number of parameters, the more difficult it is to find a solution using traditional multi parameter optimization techniques, and the choice of the starting point plays the key role. The Simultaneous Multiple Surface (SMS) design method, initially developed for non-imaging

applications, presents a direct method for designing both aspheric and freeform optical surfaces and can be used to generate starting points for multi-parametric optimization.

This thesis explores new design strategies for non-spherical imaging optical systems. It comprises of 2 parts and 7 chapters. The basic concepts in geometrical optics are shortly introduced in the first chapter. Chapters 3-7 are grouped into two parts: the aspheric optics (chapters 3 and 4) and the freeform optics (chapters 5-7). Chapter 2 is related to both parts of the thesis, since it gives a new insight to the SMS method in general.

In Chapter 2 we present new development of the SMS method which changes the way the method is regarded. We show that for specific design cases where the aperture stop is placed further away from the surfaces, multiple object points can be sharply imaged using lesser number of surfaces. Moving step-by-step from the basics of the method, to imaging two object points through one surface and transitioning from two to three object points by increasing the parallelism, we get to the designs of six surfaces imaging up to eight object points. A rule of thumb which estimates the number of object points that can be sharply imaged given the size of the aperture and the surface is presented. In the process of maximizing the object points to sharply image, we try to exhaust the degrees of freedom of aspherics and freeforms. We conjecture that the SMS designs with exhausted degrees of freedom are very close to a good solution; hence, using them as a starting point for the optimization will lead us faster to the optimal optical system.

Chapter 3 presents a study of the SMS2D method as a starting point for a two-lens design problem. We investigate two design examples and compare them with similar designs obtained via traditional design approaches, as well as global optimization. Traditional design approaches use a spherical starting point with a single-step, a two-step and a stepwise optimization, while SMS starting point is optimized using only a single-step optimization. It is shown that, since SMS method gives a better starting point, a single-step optimization is sufficient to get to the optimal design.

Chapter 4 introduces a novel concept in imaging optical design known as foveated imaging. Foveated imaging is a method in which the optical systems mimic the behaviour of the human eye in some cases, or use that behaviour to adapt the optics performance to it, by generating images with varying resolution. The first part of this chapter explores the possibility of using the SMS method with pupil discretization to design foveated optical systems. Using pupil discretization we are not limited by the number of object points needed to achieve foveation. Starting from an existing microscope objective lens, the aspheric part of the system is redesigned using the method. Several obtained sets of surfaces are then optimized and analysed to assess their asphericity compared with the initial surfaces in the design. One of the obtained sets shows better characteristics than the starting one. The foveated imaging approach is then applied to a virtual reality head mounted display (HMD) lens design, where 4 aspheric surfaces are designed to satisfy the foveation requirement. Apart from foveation, this design adopts an eye-rotation approach commonly used in spectacle lens design. An HMD lens adapted to the human eye resolution is designed combining these two techniques. The

tolerance analysis for this system is also presented, as well as the experimental resolution measurements of the manufactured prototype.

Chapter 5 investigates an afocal two freeform mirror design problem in first order optics. The resulting first-order partial differential equations for the freeform two mirror system have an analytic solution with the sole condition that the x - y and x' - y' axes are parallel. Four selected solutions are presented. Two of them are semi-aplanatic (fulfilling the aplanatic condition only for the x -coordinates), while the other two are, to our knowledge, the first examples of aplanatic two-mirror systems without rotational symmetry. Presented problem can also be seen as a limit case of an SMS design problem, as discussed in the text. Alternative system geometries have been investigated as well. However, in these cases ray-to-ray mapping equations are not necessarily decoupled.

Another contribution to the SMS method is given in chapter 6. An algorithm for designing 3 plane-symmetric freeform surfaces with the SMS3D method is presented. An example of such a system, fulfilling stigmatic imaging at the design points is given as well. Moreover, it is demonstrated that at the limit in which the three design wavefronts coincide, the design example presented is a semi-aplanatic system, where the aplanatism up to the second order is obtained only in one direction.

Chapter 7 compares different orthogonal polynomial surface representations for optical systems with rectangular apertures. SMS3D designs that are intended to be used in imaging systems need to be optimized for the whole field. For that purpose a convenient representation has to be chosen. Recent studies prove that for the case of circular apertures, the best surface representation from the manufacturability estimates and optimization convergence point of view is Forbes Q-polynomial representation. However, freeform optical surfaces are often used in optical systems with non-circular apertures. Several applications of freeform optics (high-aspect ratio systems, systems with digital displays) call for deeper analysis of systems with rectangular apertures. Two reflective systems with rectangular apertures are studied and compared using 4 different surface representations. We compare polynomials with different orthogonal areas (rectangular and circular) and different metrics (sag and gradient orthogonality). Polynomials orthogonal inside a rectangle converge faster or to a better local minimum than the ones orthogonal inside a circle in the examples considered. This is most likely due to the loss of the good properties of orthogonality when the orthogonality area does not coincide with the surface area used.

TABLE OF CONTENTS

Chapter 1. Basic concepts and definitions	1
1.1. Introduction to Geometrical Optics.....	1
1.2. Basic parameters of imaging optical systems.....	3
1.3. Generalized Cartesian oval and the SMS method	4
1.3.1.SMS2D method – The SMS chain.....	6
1.3.2.SMS3D method – The SMS surfaces	7
1.3.3.SMS2D method with pupil discretization	8
1.4. Aberration theory of imaging optics.....	10
1.5. Image evaluation	13
1.5.1. Geometrical spot diagram and RMS spot size	13
1.5.2. Transverse ray aberration curves	14
1.5.3. The modulation transfer Function	15
1.6. Degrees of freedom of aspheric surfaces.....	16
1.7. Freeform surfaces.....	18
1.8. Optimization of optical systems	19
1.8.1. Damped least squares.....	20
1.8.2. Global optimization.....	20
1.8.3. Merit function construction.....	20
Chapter 2. Conditions for perfect focusing multiple point sources with SMS design method	25
2.1. Introduction	25
2.2. Series vs. Parallel designs	26
2.3. Designing with more than two object points	29
2.3.1.Increasing parallelism in the compound Cartesian oval design	29
2.3.2.Series-parallel combination in SMS designs	30
2.4. Applications to optimization	33
2.5. Conclusions	36
 PART 1 Aspheric optics	
Chapter 3. An analysis of an SMS2D design as a starting point for a two lens design problem	41
3.1. Introduction	41
3.2. Starting point with the SMS method	43
3.3. Comparing the results of different design approaches	44
3.4. An insight in optimization process	51
3.5. Conclusions	53

Chapter 4. Foveated imaging	57
4.1. Introduction	57
4.2. Starting point from SMS2D with pupil discretization	60
4.2.1. Foveated imaging in endoscopic lenses	60
4.2.2. Application of SMS method with pupil discretization to a foveated microscope objective lens	60
4.3. Super-resolution VR HMD optics	67
4.3.1. Introduction to VR HMD optics	67
4.3.2. Mapping function and focal lengths for HMD optics	68
4.3.3. Eye rotation design approach	73
4.3.4. Design example	74
4.3.5. Tolerance analysis	77
4.3.6. Experimental resolution measurements of the manufactured prototype	84
4.4. Conclusions	91

PART 2 Freeform Optics

Chapter 5. Analytical solution of an afocal two freeform mirror design problem	97
5.1. Introduction	97
5.2. Phase space considerations	98
5.3. A particular two-mirror solution	100
5.4. Design examples	102
5.5. Discussion	105
5.6. Connection with the SMS method	105
5.7. Analytic calculation for different mirror system geometries	106
5.8. An afocal refractive system with two freeform surfaces	108
5.9. Conclusions	110
 Chapter 6. SMS3D method for design of 3 freeform surfaces	 113
6.1. Introduction to SMS3D method	113
6.2. Design procedure	114
6.3. Design example	117
6.4. Aplanatism and its link to SMS design method	121
6.4.1. Semi-aplanatic as a limit case of the 2-point SMS3D method.....	123
6.4.2. Semi-aplanatic of second-order as a limit case of the 3-point-in-line SMS3D method	124
6.4.3. RMS3D distribution of semi-aplanatic systems	124
6.5. Conclusions	125

Chapter 7. Comparison of different orthogonal polynomial surface representations	129
7.1. Introduction	129
7.2. Orthogonal surface representations under investigation	131
7.2.1. New Q-Legendre polynomials	132
7.3. Optical design examples	134
7.3.1. One surface off-axis magnifier	135
7.3.2. Two mirror configuration	138
7.4. Conclusions	142
Conclusions and future work	145
Appendix	151
Publications	159

Chapter 1

BASIC CONCEPTS AND DEFINITIONS

1.1. Introduction to Geometrical optics

Scientists have observed that light energy can behave like a wave as it moves through space, or it can behave like a discrete particle with a discrete amount of energy (quantum) that can be absorbed and emitted. The particle-like nature of light is modelled with photons. A photon has no mass and no charge. It is a carrier of electromagnetic energy and interacts with other discrete particles (e.g., electrons, atoms, and molecules). Wave-like model must be used to describe effects such as interference and diffraction. The propagation of light or electromagnetic energy through space can be described in terms of a traveling wave motion. The wave moves energy, without moving mass, from one place to another at a speed independent of its intensity or wavelength. This wave nature of light is the basis of physical or wave optics.

When the wavelength of light is small enough so that the spatial variations of the electromagnetic field are much higher than the wavelength, geometrical optics can be used. Since light propagation is described in terms of rays, this branch of optics is sometimes called ray optics. The condition of small wavelength is usually satisfied in most real-life optical applications where the size of optical elements is much bigger than the wavelength. That is why traditional optical system design uses geometrical optics approach.

A ray of light is defined as a straight line normal to any surface of constant phase of light wave (in terms of wave optics) which is called a wavefront. In geometrical optics, rays follow the laws of reflection and refraction. When light is reflected from a smooth surface, the reflected ray forms the same angle (α_r) with the surface normal as the incident ray (α_i), as in Equation (1.1.1). Incident and reflected ray, as well as the surface normal are coplanar.

$$\alpha_r = \alpha_i \quad (1.1.1)$$

When a ray of light passes from a medium with one index of refraction n_1 to a medium of a different index of refraction n_2 , its direction changes according to Snell's law of refraction. This law states that the ratio of the sines of the angles of incidence and refraction is equivalent to the ratio of phase velocities in the two media, or equivalent to the opposite ratio of the indices of refraction (Equation (1.1.2) - (1.1.3)).

$$v_1 \sin \alpha_2 = v_2 \sin \alpha_1 \quad (1.1.2)$$

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2 \quad (1.1.3)$$

There are many equivalent geometrical optics formulations. One is a Fermat's principle that states that light travels between two points along the path that requires the least time, as compared to other nearby paths. According to Fermat, a ray is an extremal curve of the following curvilinear integral [1]:

$$L_{OPL} = \int_A^B n(x, y, z) dl \quad (1.1.4)$$

where, $n(x, y, z)$ is the refractive index of the medium at point (x, y, z) and dl is the differential length along the integration path between points A and B . Value L_{opt} is known as the optical path length between points A and B .

The other equivalent formulation is a comprehensive model known as Hamiltonian system [2]. It is very common to deduce the Hamiltonian system from Fermat's principle as a solution to a variation problem expressed in Cartesian coordinates. Suppose that we have a ray passing through a point (x, y, z) with a direction v represented as a unitary six-dimensional vector (x, y, z, p, q, r) where $(p, q, r) = n(x, y, z)$ and v are the optical direction cosines of the ray with respect to the x -axis, y -axis and z -axis, respectively. From the previous it is obvious that $p^2 + q^2 + r^2 = n^2(x, y, z)$. The Hamiltonian formulation states that the trajectories of the rays are given as the solution to the following system of the first-order ordinary differential equations [3]:

$$\begin{aligned} \frac{dx}{dt} &= H_p & \frac{dp}{dt} &= -H_x \\ \frac{dy}{dt} &= H_q & \frac{dq}{dt} &= -H_y \\ \frac{dz}{dt} &= H_r & \frac{dr}{dt} &= -H_z \end{aligned} \quad (1.1.5)$$

where $H = p^2 + q^2 + r^2 - n^2(x, y, z)$ and t is a parameter along the trajectories of the light rays. H_k defines partial derivative of function H with respect to parameter k . The solution should be consistent with $H = 0$, as it can be deduced from the definition of (p, q, r) .

Above presented Hamiltonian formulation implies six-dimensional space coordinates (x, y, z, p, q, r) for ray trajectories. If n is known and one of the direction cosines is positive, a ray can be defined by a point and two direction cosines as a five-parameter entity. The five-

dimensional space defined by a point and two direction cosines is called Extended Phase Space [4], [5].

In phase space each point depicts both the position of a light ray and its propagation direction at that point. Hence, the ray bundles are represented as a volume of positions. Ray trajectory in three dimensions (x, y, z) with the material index of distribution $n(x, y, z)$ is a curve in the phase space. Propagation of one initial ray-bundle through a certain material is similarly defined as the transformation from one volume of initial positions into another volume of final positions as a function of time. A ray-bundle $M_{4D}(x, y, p, q)$, also known as manifold, is defined as a closed set of points in the extended phase space where each point represents a different ray. For example, if all rays intersect a given surface at $z = f(x, y)$, these rays together can be fully described as a four-dimensional bundle $M_{4D}(x, y, p, q)$ with the condition that $r = \sqrt{n^2 - p^2 - q^2}$.

All these concepts are similarly defined in two-dimensional (2D) geometry. The extended phase space in this case is the three-dimensional (3D) sub-manifold, defined as $p^2 + q^2 = n^2(x, y)$ in the four-dimensional space (x, y, p, q) . Reference surfaces become curves in the xy -plane and a ray-bundle m_2 is a two-parameter entity.

Phase space treatment introduced above will be useful in a discussion presented in Chapter 5, where an analytic solution of an afocal two freeform mirror system is obtained.

1.2. Basic parameters of imaging optical systems

Imaging optical systems consist of three main components: the object, the optics and the image it forms. The light or part of it from each point on the object is collected by the optical system and converged to the corresponding point on the image. The objective of an imaging optical system is to resolve a specified minimum-sized object over a desired field of view.

A specification document that lists the requirements the system needs to satisfy is the first step in designing an optical system. It defines basic specifications such as object and image location, aperture, field of view and spectral range. It usually contains image quality measures as well. It should also include packaging and environmental requirements, such as diameter, overall system length and temperature.

The aperture specification determines the amount of light collected from an object. Three most common ways to specify an aperture in imaging optics (assuming the focal length is known) are: the f-number ($f/\#$), the entrance pupil diameter (EPD) and a numerical aperture (NA). The EPD defines the size of the beam in object space and it is usually encountered in computer-aided design. The ratio of the focal length of a system f to its entrance pupil diameter is called the relative pupil or $f/\#$ (Equation (1.2.1)). This specification is commonly used in systems with distant objects (e.g. camera lenses or telescopes). Numerical aperture is the index of refraction of the image space media n' times the sinus of the half-angle cone of light u'

(Equation (1.2.2)). NA is commonly used to specify systems that work at finite conjugates (e.g. microscope objectives). $f/\#$ and NA are related by $f/\# = 1/(2NA)$ and:

$$f / \# = \frac{f}{EPD} \quad (1.2.1)$$

$$NA = n' \sin u' \quad (1.2.2)$$

The field of view (FOV) of an optical system is specified as an object height, object angle or image height. The half field of view (HFOV) is the radius of the object height h or image height h' , whereas the full field of view (FFOV) indicates a diameter measure. Another common definition involves an angular HFOV θ , which, in the case of a rectilinear point-to-point mapping between object and image planes, is related with the image height h' by:

$$h' = f \tan \theta \quad (1.2.3)$$

where f is the focal length of the system.

The spectral range of an optical system is represented by a central wavelength λ_0 and a bandwidth $\Delta\lambda$. Narrow bandwidth systems are called monochromatic, whereas larger-bandwidth systems are called polychromatic.

1.3. Generalized Cartesian oval and the SMS method

In order to achieve a perfect image, optical path length from object point to image point must be the same along all rays belonging to the same wavefront. This problem of perfect imaging was solved first by René Descartes in the 17th century [6]. He discovered that one spherical wavefront can be transformed into another using a single surface. This surface is known as a Cartesian oval, and it represents a pioneer aspheric surface used to correct spherical aberration. Generalized Cartesian ovals were found in 1900 by Levi-Civita [7]. They are reflective or refractive optical surfaces that transform any input wavefront WF_{in} into another output wavefront WF_{out} , as shown in Figure 1.1. Resulting optical surfaces are usually free from rotational symmetry, and are known as freeform surfaces.

Figure 1.1. A generalized Cartesian oval that transforms WF_{in} into WF_{out} .

The limitation of generalized Cartesian oval consists in the fact that it can only couple one pair of input and output wavefronts. The extension that involves coupling a greater number of wavefront pairs is known as Simultaneous Multiple Surfaces (SMS) design method [8]–[14].

The SMS method is a direct design method based on the laws of geometrical optics. It was first developed in 1990's as a Non-imaging design method [8]. Later-on, the method was extended to the field of imaging optics as well, with many successful applications. [11], [15], [16]

The SMS method traditionally involves simultaneous calculation of N optical surfaces using the same number of wavefront pairs. However, new developments presented in Chapter 2 of this thesis show that when specific conditions are applied, it is possible to couple a greater number of wavefronts using smaller number of surfaces.

There are two versions of the method – SMS 2D used for designing rotationally symmetric aspheric surfaces, and SMS 3D used for designing freeform surfaces without rotational symmetry. Figure 1.2 presents an SMS 3D optical system coupling two wavefront pairs.

Figure 1.2. A schematic demonstration of SMS 3D optical system coupling two pairs of wavefronts.

1.3.1. SMS 2D method – The SMS chain

The basic construction of the SMS optics is the SMS chain. Initial conditions for its construction are chosen by the designer and they involve choosing two optical path lengths L_1 from the wavefront $WF1_{in}$ to $WF1_{out}$ and L_2 from $WF2_{in}$ to $WF2_{out}$, as well as point P_0 on one of the surfaces to design with its normal unit vector N_0 . An illustration of SMS chain calculation is shown in Figure 1.3.

Figure 1.3. The SMS chain calculation for two refractive surfaces.

The algorithm applied in order to calculate two sets of points and normal vectors that form the SMS chain includes the following:

- (1) Ray 1 that comes from wavefront **WF1_{in}** is deflected at the point P_0 . Next, point P_1 with its normal vector N_1 is calculated along the deflected ray trajectory using the constant optical path length L_1 from the wavefront **WF1_{in}** to **WF1_{out}**;
- (2) The procedure is now reversed. Ray 2 is traced from **WF2_{out}** and deflected at P_1 . Using the constant optical path length condition for the second pair of wavefronts (L_2) we calculate now point P_1 with its normal vector N_2 ;
- (3) This process is repeated to obtain the full SMS chain.

Using the above procedure a sequence of two sets of points and normal vectors is obtained. Next step involves interpolating the obtained set of points to get curves such that tangent vectors are perpendicular to normal vectors. The optical path lengths between two pairs of wavefronts should be carefully selected in order to make the chain integrable. Otherwise, curves interpolated through a calculated chain will not have the normal vectors required.

If the rays coming from the two selected wavefront pairs all belong to a single plane, two resulting SMS curves can be rotated about the optical axis (symmetry axis). By this process, SMS 2D surfaces are obtained. However, SMS 2D designs only couple rays belonging to a single plane, therefore not whole wavefronts will be perfectly coupled, but just their meridian rays.

For a resulting optical system that perfectly transforms **WF1_{in}** to **WF1_{out}** and **WF2_{in}** to **WF2_{out}**, we need to apply SMS 3D design procedure.

1.3.2. SMS 3D method – The SMS surfaces

SMS design procedure for two freeform surfaces includes three steps [17]:

- (1) The construction of the seed rib;
- (2) The construction of spine curves;
- (3) The skinning process through a spine curve.

After SMS chain calculation previously explained, one of the interpolated curves is chosen as the input data for the next stage of the SMS 3D design process. This curve is called a *seed rib*.

We now assume that SMS chain may be generated from any point P of the seed rib using previously described procedure for a chain construction. The set of points generated from all the points of the seed rib will form another curve on the second optical surface. Repeating the process of propagation and coupling of rays belonging to two pairs of bundles described previously, two 3D sets of points together with its normal vectors (point clouds) are calculated. The points belong to interpolated curves which are called *SMS ribs*.

Next step is called *SMS skinning*, and it involves interpolating *SMS ribs* to SMS surfaces using Non-uniform Rational B-Splines (NURBS). This representation gives more local control of a

freeform surface. However, if we intend to use the obtained system for imaging, fitting a point cloud to a polynomial-based representation may be more appropriate for representing a system in commercial software for further optimization and analysis [11], [16]. Figure 1.4 shows an RMS spot size for an optical system of 2 SMS surfaces intended for further optimization. Figure 1.4 (left) presents a resulting root mean square (RMS) 2D curve obtained using only meridian rays (just a 2D cross-section) in SMS 2D design method, while Figure 1.4 (right) presents full RMS profile for the whole FOV of an arbitrary SMS 3D design. An extension of SMS 3D method for designing three on-axis surfaces is given in Chapter 6.

Figure 1.4. A resulting RMS_{2D} distribution curve of a rotationally symmetric optical system designed using SMS 2D for two surfaces (left); An RMS distribution for a 2-surface freeform system designed with the SMS 3D method (right).

1.3.3. SMS 2D method with pupil discretization

An alternative to standard SMS 2D method presented in Section 1.3.1, where N pairs of wavefronts (including only rays in a plane) are perfectly coupled through N SMS surfaces, is a so-called pupil discretization approach presented in [18] (Figure 1.5). SMS method with pupil discretization traces rays through N points of the chosen pupil to design N aspheric surfaces. Chosen pupil could be the entrance or the exit pupil of the system, or a desired aperture stop. Unlike in a standard SMS method previously explained, no wavefronts are specified, thus a constant optical path length condition cannot be used.

Figure 1.5. Standard SMS method with object discretization, where 4 object points are perfectly focused using 4 surfaces (up) and SMS method with pupil discretization, where 4 surfaces are designed using 4 points in the pupil (down).

A ray coming from point y at the object passes through the chosen pupil point and arrives at the image at the point $y' = my$, where m is a previously chosen magnification. Therefore, instead of using a Cartesian oval for the seed patch, as done in the standard SMS 2D with object discretization [4], a so-called “distortion-free oval” is used. [18] A Cartesian oval, as previously explained, forms a stigmatic image of all the rays coming from a single point at the object, whereas a distortion-free oval forms an image for a single ray coming from all the points of the object, following a prescribed object-to-image mapping $y' = my$. A difference between Cartesian Oval and distortion-free oval is illustrated in Figure 1.6.

Figure 1.6. An illustration of a Cartesian oval and a distortion-free oval for a single point at the object and a single point at the pupil, respectively.

This duality between the Cartesian oval and distortion-free oval can be expanded to the case of the SMS method with object discretization and SMS method with pupil discretization. Analogue to the SMS with object discretization, where a resulting system has N rays passing through each point of the surface, in SMS with pupil discretization, N rays pass through each point of the object. If we, for example take $N=4$, as in Figure 1.5, for a design with object discretization 4 rays pass through each point of the surface, whereas for a design with pupil discretization 4 rays pass through each point of the object (belonging to the four design points at the pupil). This method will be used in Chapter 4, for designing 4 aspheric surfaces of a foveated microscope objective lens.

1.4. Aberration theory of imaging optics

In ideal imaging optical systems, rays from an object point are transferred exactly to its ideal Gaussian image point. However, in real imaging systems the transmission is never perfect, and the imperfections in the image formation are known as aberrations.

Figure 1.7 shows that each ray passing through a rotationally symmetric optical system can be expressed by three parameters (ρ, r, θ) , where (ρ, θ) are the polar coordinates of intersection point at the exit pupil plane, and r' is the height of the image on the image surface. Since the image point is conjugate to an object point with height r , we can also express that ray with (ρ, r, θ) .

Figure 1.7. Parametric representation of a ray.

Wave aberration function W measures the difference between the real image wavefront and the ideal Gaussian image wavefront. For a rotationally symmetric optical system, any ray passing through the lens can be defined with parameters (ρ, r, θ) . Therefore, wave aberration function W is also a function of (ρ, r, θ) . W can be expressed as a power series of previously mentioned parameters as follows [20], [21]:

$$\begin{aligned}
 W = & {}_0W_{20}\rho^2 && \text{Defocus} \\
 & + {}_1W_{11}r\rho\cos\theta && \text{Lateral image shift} \\
 & + {}_2W_{00}r^2 && \\
 & + {}_0W_{40}\rho^4 && \text{Spherical aberration} \\
 & + {}_1W_{31}r\rho^3\cos\theta && \text{Coma} \\
 & + {}_2W_{22}r^2\rho^2\cos^2\theta && \text{Astigmatism} \\
 & + {}_2W_{20}r^2\rho^2 && \text{Field curvature} \\
 & + {}_3W_{11}r^3\rho\cos\theta && \text{Distortion} \\
 & + {}_4W_{00}r^4 && \\
 & + {}_0W_{60}\rho^6 && \text{Spherical aberration} \\
 & + \dots && \\
 \end{aligned}
 \quad
 \left.
 \begin{array}{l}
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \end{array}
 \right\}
 \begin{array}{l}
 \\
 \\
 \\
 \text{Third order aberration} \\
 \\
 \\
 \\
 \\
 \end{array}
 \quad
 \begin{array}{l}
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \end{array}
 \quad
 \left.
 \begin{array}{l}
 \\
 \\
 \\
 \\
 \end{array}
 \right\}
 \begin{array}{l}
 \\
 \\
 \text{Higher order aberration} \\
 \end{array}
 \quad
 (1.4.1)$$

In third-order aberration theory the higher terms in the wave expansion are dropped. The value of each polynomial coefficient gives the magnitude of the aberration it represents and is a function of lens parameters (e.g. index of refraction, thickness, radius of curvature), and marginal/chief ray heights and angles.

The existence of wave aberrations leads to deviations between real rays' intersection points with the image plane and the ideal image position. This lateral distance is called the transverse ray error ε . (Figure 1.8) Wave aberration function W and transverse ray error ε are related by a partial derivative with respect to pupil coordinate ρ . This relation is shown in Equation (1.4.2), where u'_a is the angle of the marginal ray in image space, and indices x and y present partial derivatives with respect to x and y , respectively. [1]

Figure 1.8. Wave aberration and transverse ray aberration.

$$\begin{cases} \varepsilon'_y = \frac{1}{n'u'_a} \frac{\partial W}{\partial \rho_y} \\ \varepsilon'_x = \frac{1}{n'u'_a} \frac{\partial W}{\partial \rho_x} \end{cases} \quad (1.4.2)$$

Wave aberration function is commonly defined as the optical length measured along a ray, from the aberrated wavefront to the reference sphere. Although the reference sphere can be placed at an arbitrary position, it is convenient to place it at the exit pupil position with relative pupil coordinates [1].

In 1857, Seidel classified five primary monochromatic aberrations as embodiments of departure from the first order theory. [21] They are illustrated in Figure 1.9.

Figure 1.9. Third order Seidel aberrations (the subscripts P, M, S and T stand for paraxial, marginal, sagittal and tangential, respectively). (Source: ref [22])

Spherical aberration is a variation in focus position between the paraxial rays through the centre of the lens aperture and those through the edge of the aperture. It is the only third order aberration to appear on-axis, and it has the same magnitude for any field coordinate.

Coma is an off-axis aberration where annular zones of the aperture have different magnifications. This aberration typically appears in asymmetrical optical systems. For example, errors in tilt or decentre of lens components result in Coma. The spot diagram shows that the resulting image of a point object looks like a comet, hence the name Coma.

Astigmatism occurs when a tangential (meridian) fan of rays focus at a different location than a sagittal ray fan. There is no astigmatism on axis, and its magnitude increases with the field coordinate.

A symmetrical optical system with spherical surfaces naturally forms an image on a curved surface. Field Curvature is the aberration that describes the departure of a real image surface from a desired flat surface.

Distortion is an aberration in which the magnification varies over the image height. Object points are imaged to perfect points with no blur (unlike in the case of the other third order

aberrations), but straight lines in the object become curved lines in the image. Distortion is measured as a lateral displacement of the image point to its partial equivalent.

Chromatic aberration is another type of aberration that affects polychromatic optical systems. It occurs as a consequence of the material dispersion, i.e. change in the refractive index of the material with the wavelength of light. Therefore, different wavelengths of light coming from a single object point are not focused at the same point in the image (Figure 1.10). Chromatic aberrations are divided into two groups: longitudinal chromatic aberration and lateral chromatic aberration. Longitudinal chromatic aberration or longitudinal colour occurs when rays with different wavelengths are focused in different positions of the optical axis, whereas lateral chromatic aberration or lateral colour occurs when the rays with different wavelengths are focused in different positions of the focal plane.

Figure 1.10. Lateral and longitudinal chromatic aberration.

1.5. Image evaluation

There are many criteria to evaluate the performance of an imaging optical system. They relate in different ways to the image quality of the system. Image quality can be thought of as resolution, i.e. how close two objects can approach each other while still being resolved from one another. It can also be thought of as image sharpness or contrast. In this thesis we will use three different criteria: a geometrical spot diagram combined with root mean square (RMS) spot size, transverse ray aberration curves and modulation transfer function (MTF).

1.5.1. Geometrical spot diagram and RMS spot size

Spot diagrams are geometrical image blur formed by the lens when imaging a point object. [23] They are obtained by tracing a great number of rays from the point object through the entrance pupil. A “spot” is plotted where each ray hits the image plane with respect to the paraxial image point.

A convenient metric for the quality of the image, called the RMS spot size, can be derived from the spot diagram. First the centre of gravity or a centroid of all the image plane

intersections is determined. The RMS spot size is then calculated by taking the square root of the sum of the squares of each ray's distance from the centroid, divided by the total number of rays:

$$RMS = \sqrt{\frac{\sum_i R_i^2}{n}} \quad (1.5.1)$$

Where R_i is the radial distance of the i^{th} impact from the centroid and n is the number of rays traced. The RMS spot size gives intuitive information of how the image of the point object expands.

1.5.2. Transverse ray aberration curves

Even though RMS spot size presents robust representations of image performance, it shows overall net resulting image quality. However, optical designers are sometimes interested in obtaining information about the aberrations present in the optical system. Transverse ray aberration curves provide exactly that. They are similar to the spot diagram, but instead of tracing a 2D grid of rays in the pupil, they use a tangential (y) and a sagittal (x) slice of the pupil. A perfect lens has a straight line along the horizontal axis in both plots. The shape of the plots at different values of field height yield a lot of information about which aberrations are present in the system. Typical ray aberration curves are shown in Figure 1.11

Figure 1.11. Typical transverse ray aberration curves in two planes of interest.

1.5.3. The modulation transfer function

MTF is probably the most widely used performance metric for imaging optical systems. It is the magnitude of the complex optical transfer function (OTF) defined in Equation (1.5.2). OTF measures the optical resolution which the system is capable of achieving without considering the resolution of the sensor.

$$OTF = \frac{\text{Image Contrast}}{\text{Object Contrast}} \quad (1.5.2)$$

Contrast is calculated from the maximum and minimum object and image intensities as:

$$\text{Contrast} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (1.5.3)$$

Under Fourier transformation, the OTF is represented as a vector quantity that takes into account amplitude and phase variations, as:

$$OTF(\omega) = MTF(\omega)e^{iPTF(\omega)} \quad (1.5.4)$$

In practice, lens designers are most concerned about MTF value (the amplitude of the OTF), although the phase transfer function (PTF) can have a secondary effect. In lenses with large aberrations, the OTF can be negative, leading to MTF bounce, which represents a phase reversal of light and dark features.

MTF is the ratio of the image sine wave modulation to that of the object as a function of the spatial frequency, typically specified at the image in cycles/mm. A cycle, i.e. a line pair, is one full period (peak and trough) of the sine wave (Figure 1.12). A complete MTF plot includes performance curves for several different field points and a diffraction limited reference line (Figure 1.13). For off-axis field points, the MTF depends on the orientation of the “lines” in the object and is therefore plotted for both radial (sagittal) and azimuthal (tangential) modulation, where the modulation direction is orthogonal to the feature direction.

Figure 1.12. Contrast of a bar target and its image. (Source: Edmund Optics)

Figure 1.13. Typical MTF curves (Source: ref [23]). Curve marked as perfect system presents a diffraction limit.

Imaging performance of a system is normally quantified by specifying an MTF value at a specific spatial frequency. For digital sensors, this frequency is usually associated with the sensor's Nyquist frequency.

1.6. Degrees of freedom of aspheric surfaces

Aspheric surfaces are rotationally symmetric optical surfaces that are neither spherical nor plane. Unlike a sphere, which is described by only one parameter (the radius of curvature), an asphere may have an infinite number of shape parameters. Mathematically an asphere is generated, for example, by the rotation of an axisymmetric plane curve about its axis.

Great potential of aspheric surfaces was recognized by early scientists such as Kepler in 1611, Descartes in 1638, and Huygens in 1678. One of the first aspheric surfaces was found by Descartes, and it is a previously presented Cartesian oval that stigmatically images one point to another. The first aspheres used in practice date from the middle of the 18th century. [24] These surfaces, however, departed from spheres only by very small amounts and could be generally regarded as quadric surfaces. More complicated aspheric shapes (with high accuracies required for imaging systems) could neither be calculated nor made at that time. Progress in computer development in the last century, and especially progress in manufacturing techniques and metrology methods in the last few decades, made possible the treatment of more complicated aspheres.

According to Schulz [24], Wave aberration function presented in Eq. (1.4.1) can be rearranged in ascending powers of r or ρ as:

$$W(r, \rho, \theta) = f_{00}(\rho) + (f_{11}(\rho) \cos \theta) r + O(r^2) \quad (1.6.1)$$

$$W(r, \rho, \theta) = (g_{11}(r) \cos \theta) \rho + (g_{20}(r) + g_{22}(r) \cos^2 \theta) \rho^2 + O(\rho^3) \quad (1.6.2)$$

The functions f , in expansion (1.6.1) and g , in expansion (1.6.2) are known expressions [25], [26] containing the coefficients W of (1.4.2). By means of aspheric surfaces a limited number of coefficients W can be made equal to zero using significantly smaller number of surfaces than when only spheres are used. These coefficients, for example, may be some or all of the five Seidel aberration coefficients. Moreover, not only single coefficients W but also total functions f , or g , can be made equal to zero. For example, making

$$f_{00}(\rho) \equiv 0, \quad (1.6.3)$$

the spherical aberration of all orders is eliminated for all ρ (full pupil), thus obtaining axial stigmatism. Furthermore, If we make:

$$f_{00}(\rho) \equiv f_{11}(\rho) \equiv 0, \quad (1.6.4)$$

the system, apart from being free from spherical aberration also becomes free from circular coma of all orders, which results in an aplanatic system. The concept of aplanatism will be mentioned in Chapter 5 and 6.

On the other hand, if we switch to the case presented in Equation (1.6.2), we can eliminate for all r (total field) the distortion of all orders by making:

$$g_{00}(r) \equiv 0 \quad (1.6.5)$$

Additionally, by doing:

$$g_{20}(r) \equiv 0 \quad (1.6.6)$$

$$g_{22}(r) \equiv 0 \quad (1.6.7)$$

we can eliminate both field curvature (Equation (1.6.6)) and astigmatism (Equation (1.6.7)) of all orders.

Requirements containing $f_{00}(\rho) = 0$ (Equation (1.6.3) - (1.6.4)) can be useful in cases of small fields and wide pencils (i.e. large pupils), and $g_{00}(r) = 0$ (Equations (1.6.5) - (1.6.6)) can be useful in cases of large fields and thin pencils (i.e. small pupils). The second case (large field thin pencils) will be used in Chapter 4, for a design of an aspheric virtual reality head-mounted display (VR HMD) lens.

This theory gives an insight of how many aspheric surfaces are needed to design an optical system free from a specific aberration. For example, in ophthalmic lens design, where a point pupil is used, with a large field of view, we can obtain a system free from both field

curvature (Equation (1.6.6)) and astigmatism (Equation (1.6.7)) using two aspheric surfaces. An anastigmatic lens could be obtained using only one aspheric surface; however, other aberrations will remain in the system.

Aspheric surfaces have always used a traditional polynomial representation with a base conic. Main problem with this representation involved the lack of orthogonality, which led to uncontrolled change in the coefficients during the optimization. A great advancement in this field was made in 2007 when Forbes introduced a new formulation called Q-polynomial representation with constrained average slope [27]–[30] and a metric based on the gradient of the profiles. This formulation was successfully used to reduce assembly sensitivity and to improve the manufacturability of the surfaces. [31], [32] This representation will be used in rotationally symmetric systems presented in Chapter 3 and 4.

1.7. Freeform surfaces

In Section 1.6 we briefly explained historical journey of aspheric surfaces. More recently, surfaces that are free from rotational symmetry came into spotlight. These surfaces are commonly known as freeforms. Advances in fabrication and metrology techniques allow for their precise manufacturing and measurements. This provides new opportunities to explore novel system geometries where such surfaces can be applied.

Freeform surfaces introduce new degrees of freedom to optical systems, which were rarely available using traditional rotational symmetric optics. They allow for new modern system geometries that can contribute to obtaining more compact systems. These surfaces are already widely applied in nonimaging fields, such as photovoltaic concentration, illumination, biomedical engineering and they promise to have a significant impact in the areas of imaging optics, as well. [33], [34] Freeforms are very important in off-axis imaging systems which are in general more compact compared to the traditional on-axis ones. [35], [36] A diffraction limited off-axis system comprising of freeform surfaces was presented by Fuerschbach et al. in [37], which rose additional interest in exploring the possibilities these new surface forms provide.

However, lack of rotational symmetry introduces many difficulties for freeform optical designers. Namely, standard aberration theory no longer applies. An alternative to traditional Seidel aberrations is known as Nodal aberration theory. [38] A very promising advance in this field involves an extension of the Nodal aberration theory to freeform surfaces. [39], [40] Nevertheless, it is not yet widely accepted in designers' community. Multi-parametric optimization approach is the most common technique used amongst freeform optical designers. A particular difficulty involves the lack of existing freeform designs which could be used as starting points for optimization. Another difficulty is related to the number of parameters needed to describe a surface. Specifically, the number of surface variables is significantly increased, which complicates multi-parametric optimization design approach.

Direct design methods, such as previously presented SMS 3D method can be used to provide the starting point. [41], [42]

There are many different surface representations used for describing freeform surfaces. One of the extensively studied ones is splines representation. [43], [44] In ref [43] it is demonstrated that splines coefficients change linearly with surface deformations, so that optical path difference (OPD) errors are ideal for correction. B-Spline representation presented in ref [44] has the possibility to specify local and global structures simultaneously, which can be used to describe as-built optical surfaces including metrological data. Another very interesting representation of freeforms involves Non Rational B-Splines (NURBS) representation. This representation is particularly convenient when one has to transfer a system into a CAD software. [45]–[47]. In SMS 3D method NURBS representation is the one commonly used.

Nevertheless, the most widely used freeform surface representation is a polynomial one, for it is easy to represent in commercial design software. Using polynomials representation helps in optimization algorithms, since the number of variables is smaller than in the case of splines, or NURBS. A great advance was made in this field when Forbes presented an extension of his Q-type aspheric representation to freeform surfaces. [48], [49] This representation gives a rough interpretation of shape at a glance and it offers a range of estimates of manufacturability. Even though this representation proved very well in the case of freeform surfaces with circular apertures, different aperture shapes may be convenient for some applications. More details about polynomial surface representations can be found in Chapter 7, where a comparative study of different representations for systems with rectangular apertures is conducted.

1.8. Optimization of optical systems

Modern lens design programs use numerical optimization to find desired solution with improved performance by means of minimizing a predefined merit function. The merit function uses a single number to represent total system performance where, in general, smaller numbers represent better design solutions.

Traditionally, optical designers start with a working system, which can be obtained from patent literature, or directly from libraries provided within lens design programs. Next step involves identifying variables (e.g. thicknesses, air spaces, material parameters, surface shape) as parameters that will be varied during the optimization, while the rest of the system is kept frozen. Performance operands that depend on the system variables are then defined and used to construct the merit function. These operands usually include image quality measurements (MTF, spot size) and some first-order properties (focal length). Constraints can also be applied in order to limit the change in variable, or to make the system physically realisable. Optimization algorithms now search for a merit function minimum, not the exact solution. Commercial lens design software use proprietary local and global optimization algorithms to find minima in the merit function. Local optimization involves gradient methods to find the

nearest minimum, while global optimization searches a wider parameter space in search for more minima.

1.8.1. Damped least squares method

Damped least squares (DLS) is the most common method of local optimization which minimizes the merit function value using linear regression approximations. Each variable is changed by a small step size that balances good numerical accuracy with fast convergence time. The change in each merit function operand for each variable step change is calculated. This produces a derivative matrix and defines a local slope that allows the optimizer to progress “downhill” in system landscape in order to find a better solution.

1.8.2. Global optimization

While local optimization uses gradient search to find the nearest local minimum and moves “downhill”, global optimization attempts to find the global minimum by allowing both uphill and downhill movements. However, this optimization can require extensive computational time, which is why it is not convenient for optical systems with many variables (e.g. highly aspheric surfaces, freeform surfaces).

Global optimizers are most often used as global search algorithms to scan parameter space for multiple solutions. Even a simple starting design with plane parallel plates with well-defined merit function can yield good results. Nevertheless, there is no guarantee that global optimization will ever find the absolute global minimum for a given design problem.

1.8.3. Merit function construction

In optimization algorithms, a merit function is a single number that captures all aspects of desired optical system performance. Since it is computed a large amount of times during the optimization process, it must be kept as simple as possible. The merit function ψ is constructed by taking the RMS of all identified operands:

$$\psi = \sum_{i=1}^m W_i^2 (c_i - t_i)^2 \quad (1.9.1)$$

where, m is the number of operands, W_i is the weighting factor for operand i , c_i is the current value and t_i the target value of the operand i . Squaring each operand ensures that positive and negative operand values do not offset each other in the sum, and it serves to magnify the operands with the worst performance. Weighting factor W_i is used to emphasize the desired contribution of a specific operand to overall performance.

Most design programs have their default merit function which automatically defines a set of image quality operands and includes the appropriate pupil sampling, as well as weighting for

fields and wavelengths. Designs presented in this thesis will use both default merit function from CODE V® software and specific user defined ones.

REFERENCES

- [1] M. Born and E. Wolf, *Principles of Optics, 7th (expanded) edition*, vol. 32, no. 5. 2000.
- [2] R. K. Luneberg, *Mathematical Theory of Optics*. University of California Press, 1964.
- [3] J. A. Arnaud, *Beam and fiber optics*. New York: Academic Press, 1976.
- [4] J. Chaves, *Introduction to Nonimaging Optics*. CRC Press, Taylor & Francis Group, 2008.
- [5] R. Winston, J. C. Miñano, and P. Benítez, *Nonimaging Optics*, vol. 53, no. 9. 2005.
- [6] R. Descartes, *Le Livre Premier de La Géométrie*. Leyden, 1637.
- [7] T. Levi-Civita, “Complementi al teorema di Malus-Dupin, Nota I, Atti della Reale Accademia del Lincei. Rendiconti della Classe de scienze fisiche,” *Mat. e Nat.* 9, vol. 9, pp. 185–189, 1900.
- [8] J. C. Miñano and J. C. González, “New method of design of nonimaging concentrators,” *Appl. Opt.*, vol. 31, no. 16, pp. 3051–3060, 1992.
- [9] P. Benítez *et al.*, “Simultaneous multiple surface optical design method in three dimensions,” *Opt. En*, vol. 43, no. 7, pp. 1489–1502, 2004.
- [10] O. Dross, R. Mohedano, P. Benítez, J. C. Miñano, J. Chaves, and J. Blen, “Review of SMS Design Methods and Real World Applications,” in *Proc SPIE 5529*, 2004.
- [11] J. C. Miñano, P. Benítez, W. Lin, J. M. Infante, F. Muñoz, and A. Santamaría, “An application of the SMS method for imaging designs,” *Opt. Express*, vol. 17, no. 26, pp. 24036–24044, 2009.
- [12] D. Grabovickic, P. Benítez, and J. C. Miñano, “Free-form V-groove reflector design with the SMS method in three dimensions,” *Opt. Express*, vol. 19 Suppl 4, no. May, pp. A747–A756, 2011.
- [13] W. Lin, “Advances in the Simultaneous Multiple Surface optical design method for imaging and non-imaging applications,” Universidad Politécnica de Madrid, 2012.
- [14] J. M. Infante, “Optical Systems Design using the SMS method and Optimizations,” Universidad Politécnica de Madrid, 2013.
- [15] F. Muñoz, P. Benítez, and J. C. Miñano, “High-order aspherics : The SMS nonimaging design method applied to imaging optics,” in *Proc. SPIE*, 2008, vol. 7100, pp. 1–9.
- [16] W. Lin *et al.*, “SMS-based optimization strategy for ultra-compact SWIR telephoto lens design,” *Opt. Express*, vol. 20, no. 10, p. 11581, 2012.

- [17] J. Blen, "Design of multiple free-form optical surfaces in 3D," Universidad Politécnica de Madrid, 2007.
- [18] F. Muñoz, "Sistemas ópticos avanzados de gran compatibilidad con aplicaciones en formación de imagen y en iluminación," Universidad Politécnica de Madrid, 2004.
- [19] O. N. Stavroudis, *The Mathematics of Geometrical and Physical Optics: The k function and its ramification*. Wiley-VCH Berlin, 2006.
- [20] H. H. Hopkins, *Wave Theory of Aberrations*. Clarendon Press, 1950.
- [21] W. T. Welford, *Aberrations of optical systems*. 1986.
- [22] P. Bely, *The design and construction of large optical telescopes*. .
- [23] R. E. Fischer, Bi. Tadic-Galeb, and P. R. Yoder, *Optical System Design*. 2008.
- [24] G. Schulz, "Aspheric surfaces," in *Progress in Optics XXV*, 1988, pp. 349–416.
- [25] G. Schulz, "Higher order aplanatism," *Opt. Commun.*, vol. 41, no. 5, pp. 315–319, 1982.
- [26] G. Schulz, "Aberration-free imaging of large fields with thin pencils," *Opt. Acta (Lond)*., vol. 32, pp. 1361–1371, 1985.
- [27] G. W. Forbes, "Shape specification for axially symmetric optical surfaces.," *Opt. Express*, vol. 15, no. 8, pp. 5218–5226, 2007.
- [28] G. W. Forbes, "Robust, efficient computational methods for axially symmetric optical aspheres.," *Opt. Express*, vol. 18, no. 19, pp. 19700–19712, 2010.
- [29] G. W. Forbes, "Effective Characterization of the Nominal Shape of Aspheric Optics," *Opt. Des. Eng.*, 2010.
- [30] G. W. Forbes and C. P. Brophy, "Asphere, O Asphere, how shall we describe thee?," *Opt. Des. Eng.*, vol. 7100, no. 1, p. 710002, 2008.
- [31] G. W. Forbes, "Manufacturability estimates for optical aspheres.," *Opt. Express*, vol. 19, no. 10, pp. 9923–9941, 2011.
- [32] G. W. Forbes, "Better Ways to Specify Aspheric Shapes Can Facilitate Design, Fabrication and Testing Alike," in *Proceedings of the International Optical Design Conference and Optical Fabrication and Testing*, 2010, p. JMA1.
- [33] K. P. Thompson and J. P. Rolland, "Freeform optical surfaces: a revolution in imaging optical design," *Opt. Photonics news*, vol. 23, no. 6, pp. 30–35, 2012.
- [34] J. P. Rolland and K. P. Thompson, "Freeform optics : Evolution ? No , revolution !," *SPIE Newsroom*, pp. 10–12, 2012.
- [35] T. Nakano and Y. Tamagawa, "Configuration of an off-axis three-mirror system," *Appl. Opt.*, vol. 44, no. 3, pp. 776–783, 2005.

- [36] O. Cakmakci and J. Rolland, "Design and fabrication of a dual-element off-axis near-eye optical magnifier.," *Opt. Lett.*, vol. 32, no. 11, pp. 1363–1365, 2007.
- [37] K. Fuerschbach, J. P. Rolland, and K. P. Thompson, "A new family of optical systems employing ϕ - polynomial surfaces," *Opt. Express*, vol. 19, no. 22, pp. 21919–21928, 2011.
- [38] K. P. Thompson, K. Fuerschbach, T. Schmid, and J. P. Rolland, "Using Nodal Aberration Theory to understand the aberrations of multiple unobscured Three Mirror Anastigmatic (TMA) telescopes," in *Proc SPIE*, 2009, vol. 7433, pp. 1–8.
- [39] K. Fuerschbach, J. P. Rolland, and K. P. Thompson, "Extending Nodal Aberration Theory to include mount-induced aberrations with application to freeform surfaces," *Opt. Express*, vol. 20, no. 18, pp. 821–836, 2012.
- [40] K. Fuerschbach, J. P. Rolland, and K. P. Thompson, "Nodal Aberration Theory Applied to Freeform Surfaces," in *Proc SPIE*, 2014, vol. 9293, pp. 12–16.
- [41] P. Benítez *et al.*, "Simultaneous multiple surface optical design method in three dimensions," *Opt. Eng.*, vol. 43, no. 7, p. 1489, 2004.
- [42] B. Satzer, U. Richter, U. Lippmann, G. S. Metzner, G. Notni, and H. Gross, "Using the 3D-SMS for finding starting configurations in imaging systems with freeform surfaces," in *Proc SPIE*, 2015, vol. 9626, p. 96260Y.
- [43] J. E. Stacy, "Asymmetric spline surfaces : characteristics and applications," *Appl. Opt.*, vol. 23, pp. 2710–2714, 1984.
- [44] P. Jester, C. Menke, and K. Urban, "B-spline representation of optical surfaces and its accuracy in a ray trace algorithm.," *Appl. Opt.*, vol. 50, no. 6, pp. 822–828, 2011.
- [45] H. Chase, "Optical Design with Rotationally Symmetric NURBS," in *Proc SPIE*, 2002, vol. 4832, pp. 10–24.
- [46] P. Ott, "Optic design of head-up displays with freeform surfaces specified by NURBS," in *Proc SPIE*, 2008, vol. 7100, pp. 1–12.
- [47] M. P. Chrisp and W. Street, "New Freeform NURBS Imaging Design Code," in *Proc SPIE*, 2014, vol. 9293, pp. 1–6.
- [48] G. W. Forbes, "Characterizing the shape of freeform optics," *Opt. Express*, vol. 20, no. 3, p. 2483, 2012.
- [49] C. Menke and G. W. Forbes, "Optical design with orthogonal representations of rotationally symmetric and freeform aspheres," *Adv. Opt. Technol.*, vol. 2, no. 1, pp. 97–109, 2013.

Chapter 2

CONDITIONS FOR PERFECT FOCUSING MULTIPLE POINT SOURCES WITH SMS DESIGN METHOD

2.1. Introduction

The problem of perfect focusing, also known as sharp imaging without transverse ray aberrations on the image, is present for a long time. It was solved for a single object point with monochromatic light with a single refractive surface first by René Descartes in the 17th century, using a shape called at present Cartesian oval. [1] Such a shape coincides with a hyperbola, when the object point is located at infinity in the higher refractive index media. Early works of Welford and Winston investigate the problem of sharp imaging of multiple object points. [2], [3] Correcting in reference [3] their first conclusion made in reference [2], they found that there was no fundamental contradiction that prevents the possibility of sharp imaging of meridian rays from multiple object points using a finite number of rotationally symmetric surfaces. The only limitation factor was that this sharp imaging is only possible for meridian rays, not for the full pupil including the skew ones.

Further development of perfect focusing came with the Simultaneous Multiple Surfaces (SMS) design method [4], [5], [6] introduced in Chapter 1.

Even though SMS method is normally mentioned as a design procedure that couples N wave-front pairs with N surfaces, recent findings seem to contradict that assertion. [7] In case of optical systems designed for wide field of view and with at least one surface far from the aperture stop, it was shown that it is possible to couple multiple ray sets having lesser number of surfaces. In the following sections we discuss why and when this is possible, starting with a simple design problem with two object points imaged through two surfaces, and further transitioning to doing the same with just one surface. We then go to three object points to

show that they can be perfectly imaged under specific conditions using two surfaces. Looking at the design examples presented we are able to establish a rule of thumb that specifically defines the condition under which perfect focusing of multiple object points with lesser number of surfaces can be achieved. Additionally we show that these types of systems can be interesting for optimization for the purpose of imaging applications with balanced performance throughout the whole field of view. We demonstrate that perfectly focusing the number of object points obtained using the rule of thumb exhausts the degrees of freedom of the system. In other words, the typical RMS distribution ripple for meridian rays obtained in SMS designs can be flattened, decreasing the difference between the design fields and the ones in between.

2.2. Series vs. Parallel designs

SMS surfaces (in 2D geometry) transform input one-parameter ray sets into output one-parameter ray sets. [4] Traditionally, the method involves the simultaneous calculation of a certain number of optical surfaces using the same number of input ray sets and offers the flexibility to choose the input and output ray sets to satisfy the imaging design condition: the input ray sets are diverging from object points and their associated output ray sets are converging to their associated image points. This type of ray set selection is called object discretization. There is an alternative selection called pupil discretization, when the rays converge to discrete points of the pupil [8]. Both object and pupil discretization are shortly introduced in Chapter 1.

The calculation of SMS optics is done by applying a constant optical path length condition for each coupled ray set, which is determined by choosing one initial point on each lens profile. For instance, for two object points at infinity, that is, two design angles of opposite sign, the overall symmetry implies an identical optical path length for both ray sets. In the first step the actual result of this design procedure is a set of discrete points and correspondent normal vectors, called *SMS chain*, which is later converted to complete optical surfaces through the *SMS skinning* process. It is done by choosing a seed patch which is a segment that interpolates two adjacent points of one of the surfaces of the SMS chain, matching point coordinates and normal vectors at the mentioned selected points. The *SMS skinning* is done by building SMS chains starting from the interpolated points.

First SMS design was a thin lens, where two surfaces shared an edge (Figure 2.1 (a)). This design proved that it is possible to sharply image two object points, having two surfaces. The problem was solved both in 2D and in 3D (Figure 2.1 (b)), where the 3D solution was done using freeform surfaces and is known as SMS 3D method. [9] [10] This freeform lens solved, by breaking the symmetry, the problem of perfect focusing skew rays that Welford and Winston encountered in their rotational symmetric designs, mentioned in the section 2.1 of this chapter. Hence, a 3D solution perfectly focuses the whole bundle, as can be seen in Figure 2.1(b).

Figure 2.1. SMS two-surface designs that sharply image two object points: in 2D (a), and 3D (b).

These thin SMS lenses have a convergence point. Assume that we know one point on one of the surfaces. Applying the SMS algorithm, doing the ray trace and using the constant optical path length condition, we make the SMS chain. Both sequences converge to a single point (point D in Figure 2.2). As we previously said, after having the SMS chain of points and normal vectors, we use skinning process to form surfaces. We interpolate the seed patch, making sure that it matches initial point coordinates and its normal vectors. Even though, excluding the boundary condition, the seed patch seems to be free, continuing the skinning process and getting close to the convergence point, we can notice a certain ripple in the optical surfaces. This is a result of calculation no longer being stable, nor optically valid. The fact is that there is a single seed patch that is able to produce a smooth convergence to the point D, which leads to profiles that are analytical (in the sense of expandable in a Taylor series), as was demonstrated on the example of aspheric V-groove reflector design in reference [11].

Figure 2.2. Convergence point of SMS two-surface design, showing the error in the calculation if choosing the initial segment badly. (Source: ref. [11])

Comparing this SMS design that perfectly focuses two object points with the hyperbola mentioned in the section 2.1 (and shown in Figure 2.3(a)), we get to the following conclusions:

both designs have surfaces positioned at the aperture stop, both are unique solutions and the number of object points that we can sharply image equals the number of surfaces.

Now, let us choose two symmetrical object points at infinity, and a surface far enough from the aperture stop such that the ray sets do not overlap on the surface (see Figure 2.3(b)). This means that we can build a Cartesian oval on the upper part of the optical axes that perfectly focuses the off-axis beam $\mathbf{v}_2$ into the image point A_2 . By symmetry, the mirror image of that hyperbola segment will focus $\mathbf{v}_1$ into A_1 , as shown in the Figure 2.3(b). This is a trivial solution built using two Cartesian ovals, and this possibility of sharp imaging of two off-axis object points with a remote aperture stop was already mentioned in the early work of Welford and Winston. [2], [3] In fact, the design in Figure 2.3(b), which will be referred here as a “compound Cartesian oval lens”, resembles the best known non-imaging device, the Compound-Parabolic-Concentrator, or CPC, shown in Figure 2.4.

Figure 2.3. Transition from one (a) to two (b) perfectly imaged ray bundles in the case of Cartesian oval design.

Figure 2.4. Non-imaging device called Compound Parabolic Concentrator (CPC).

Let us compare now in Figure 2.5 (a) the thin SMS lens previously shown in Figure 2.1 (a) and the compound Cartesian oval lens shown in Figure 2.3 (b) (see). In the case of SMS design with two surfaces, each surface is of the size of a beam print, unlike the compound Cartesian oval design, where surface size equals the half of the beam print (Figure 2.5 (b)). Another difference is that in the first case we have two refractions, while in the second one we have

only one. Both solutions are unique once the optical path length is fixed. We can visualize the SMS case as two surfaces working in series (since the rays cross them sequentially), while the compound Cartesian oval can be thought of as two surfaces working in parallel. Therefore, we can set the analogy with the series and parallel connected circuits. Additionally, both designs can be seen as two surfaces sharing an edge, where the only difference is the type of connection between them.

Figure 2.5. Beam print sizes of two-surface SMS design (a) and a compound Cartesian oval design (b).

2.3. Design with more than two object points

This section explains the transition from two to three and more sharply imaged object points on the already used examples of Cartesian oval design and SMS lens with two surfaces.

2.3.1. Increasing parallelism in the compound Cartesian oval design

Starting from the trivial case of compound Cartesian oval sharply imaging two object points, we will see next how to introduce a third object point to be sharply imaged by increasing the parallelism. The compound Cartesian oval is moved further away from the aperture stop in order to obtain the additional freedom to complete the surface at the centre with a free patch, as shown in Figure 2.6 (b). Moving the surface far enough, so that its size is approximately three times the size of the aperture stop, we can choose the free patch to perfectly focus a third on-axis bundle (Figure 2.6 (c)). Therefore, using only one surface it is possible to sharply image three object points. Visualizing this as three surfaces in parallel we can see that the parallelism has been increased. In this case, as in the previous one, we have a unique solution.

Figure 2.6. Transition from two to three sharply imaged object points for Cartesian oval design.

Repeating this procedure, we can increase the parallelism further. In Figure 2.7 we can see up to nine sharply imaged object points through one surface, which was designed in reference [12]. In this compound Cartesian oval lens the size of aperture stop is about 1/9 of the size of the lens. Figure 2.7 (b) shows the result of the ray trace on the lens, shown in the blue line that the angular RMS spot size of this design is null for the selected field angles, and shows ripples in between the nulls. This is a typical RMS spot size distribution graph of meridian rays (SMS 2D) for SMS designs with object discretization. As a curiosity, if instead of designing the lens with object discretization, as done so far, the pupil discretization is considered, the angular RMS_{2D} is averaged (red line in Figure 2.7(b)), as deeply discussed in reference [12].

Figure 2.7. Compound Cartesian oval design (a) and calculated RMS_{2D} spot sizes for both types of discretization (b). [12]

2.3.2. Series-parallel combinations in SMS designs

We have just proven that increasing parallelism in the compound Cartesian oval design leads to perfect focusing of multiple point sources. It has also been shown that in the case of thin SMS lens we have a fixed seed patch (i.e, with no degrees of freedom to choose it) and a

unique convergence point. In that particular design, we sharply imaged two object points. In order to introduce the third ray bundle to be perfectly focused, we need to find another degree of freedom which is obtained by moving one surface away from the aperture stop and then choosing the seed patch. This occurs because by increasing the thickness of SMS lens (Figure 2.8 (b)), we no longer have the convergence point, and therefore the seed patch is nearly free. [13]

Figure 2.8. Two-surface SMS lens with fixed seed patch (a) and nearly free seed patch (b).

If the second surface is far enough from the aperture stop, this new degree of freedom of the initial segment can be used to couple an additional on-axis ray set. [14] Starting with determining the convergence points and deriving the full lens profile from them using Fermat's principle, the unique analytical solution was obtained in [14] in the form of Taylor series expansion (Figure 2.9). This type of design is analogue to a combination of series plus parallel connection circuit. The problem was also solved in 3D in ref [15], simultaneously obtaining two freeform surfaces.

Figure 2.9. The convergence points in this design are characterized by the special case that the on- and off-axis rays passing through share identical points and normal vectors. (Source: ref. [14])

Two previously described transition processes from two to three object points (shown in Figure 2.6, Figure 2.8 and Figure 2.9) have certain similarities. The first one is that in both designs we started from the unique solution. By moving one surface away from the aperture stop, we gained additional degrees of freedom, shown in the form of *free patch*. This free patch

was then chosen to enable introducing the third ray bundle. The results are again two unique solutions, with no free patches, that sharply image three object points.

We can conjecture a general rule of thumb to calculate the number of points we can perfectly image depending on the number of surfaces we have in our system and how far they are from the aperture stop. This can be formulated in the following way: we can measure each surface size in units of the beam print on it, and add the surface sizes in those units, that is, in 2D:

$$\text{Number of object points (2D)} \leq \sum E \left(\frac{\text{Extension of surface}}{\text{Extension of beam print}} \right) \quad (2.1)$$

and applied to 3D round field:

$$\text{Number of object points (3D)} \leq \sum E \left(\frac{\text{Extension of surface}}{\text{Extension of beam print}} \right)^2 \quad (2.2)$$

Let us look at a system of 6 rotationally symmetric surfaces shown in Figure 2.10 (a) obtained by traditional SMS 2D design method that uses 6 object points. Resulting RMS distribution for meridian rays (RMS_{2D}) is shown in Figure 2.10 (b). [16] However, if we measure the extension of the beam print on each surface, as shown in Figure 2.11, we obtain that we could design a system with the same surface distribution using 8 object points. The resulting 6 surface system designed with 8 object points is shown in Figure 2.12 (a). It has an $f/3$ objective with focal length of 30 mm and a field of view of 28 degrees. The RMS_{2D} spot size distribution is presented in Figure 2.12 (b). Therefore, at least in the examples considered the rule of thumb seems valid.

Figure 2.10. An SMS 2D system with 6 aspheric surfaces designed using 6 object points (a); an RMS_{2D} spot size distribution graph for meridian rays (b).

Figure 2.11. Application of the rule of thumb on SMS 2D 6 surface system from Figure 2.10.

Figure 2.12. Six surface SMS design perfectly focusing 8 ray bundles (a) and an RMS_{2D} spot size distribution graph (b).

2.4. Application to optimization

Maximizing the number of sharply imaged object points we are trying to exhaust the degrees of freedom of aspheric and freeform lenses. The indication of this exhaustion is the flatness of the RMS_{2D} curve and its fast divergence outside the field of view shown in Figure 2.12 (b). It is in general considered that SMS designs that use all the degrees of freedom are close to an optimum solution, that is, close to a good local minimum in optimization nomenclature (Figure 2.13).

Figure 2.13. The illustration of the path followed by the merit function from SMS starting point to an optimum solution (a); an estimation of the expected optimum RMS_{2D} distribution that could be obtained from that SMS design (b).

On our path to finding a good design we can follow different optimization strategies. One is combining SMS surfaces with the standard surfaces and optimizing the free parameters of SMS as well (for example, the position of the vertices, free patches, wavefronts). [18] An alternative is to use the SMS design as a starting point for a standard optimization. Reference [19] shows an example of such an optimization for the case of imaging with high aspect ratio using freeform lens of two surfaces that perfectly focuses three ray bundles. The results from using a 3D design method were shown in ref. [19]. The authors fitted the surfaces with an x-y polynomial and introduced them to CODE V® software where the default optimization was done (Figure 2.14). The further compared in reference [19] this optimized design with an optimized rotational symmetric lens, demonstrating significantly smaller RMS spot diameter in the freeform case (Figure 2.15).

Figure 2.14. RMS spot diameter for the contour plots of the SMS freeform lens with high aspect ratio before and after the optimization (Source: Ref. [19]).

Figure 2.15. Comparison of the optimized rotational symmetric lens and the optimized freeform SMS 3D lens (Source: Ref [19]).

Different benefits of using SMS 2D design as a starting point for further optimization were discussed in more detail in Chapter 3, not only for meridian rays but for the full pupil (i.e, using the RMS_{3D} distributions). However, the great potential of SMS method lies in designing freeform surfaces, since that scientific field is rather new and a direct method that generates starting systems could make a big difference. Reference [20] discusses in detail using SMS 3D method to generate starting configurations for further optimization. The authors compare using this approach with classical starting configuration demonstrating the superiority of the first one in both RMS spot size values as well as the convergence of the merit function during the optimization. Additional studies of similar character are needed to prove this claim.

2.5. Conclusions

This chapter demonstrates that sharp imaging of multiple object points is possible, both on the example of compound Cartesian oval lens, as well as on a standard two surfaces SMS design. It was shown that traditional way of looking at the SMS method as a method that couples a certain number of wavefronts with the same number of surfaces does not necessarily have to apply to all systems. The conditions for coupling a greater number of wavefronts than the number of surfaces we want to design are that the surfaces are not at the aperture stop, making it feasible to perfectly focus additional ray bundles by increasing the parallelism. The general rule of thumb that can help in the initial assessment of the number of point sources (i.e. wavefronts) that a given system can perfectly image was also suggested. We presented a design example of six SMS surfaces perfectly imaging eight object points, to demonstrate the accuracy of the rule of thumb. The analysis of more designs to test the correctness of the rule of thumb will be desirable in the future.

Furthermore, it was indicated that the SMS designs of exhausted degrees of freedom can be used as good starting points for optimization, and that the method itself gives a direct algorithm to the solution. This particular feature is potentially very important in the case of systems that contain many aspheric or freeform surfaces. Such systems were not used extensively for imaging applications in the past due to inability to manufacture them with good accuracy. However, manufacturing techniques are developing quickly, and aspheric and even freeform surfaces nowadays can be fabricated with high accuracy. Nonetheless, designing imaging systems with many aspheres or freeforms is still quite challenging. It is mostly done optimizing a specific merit function in a commercial software. Since such surfaces have many degrees of freedom the number of variables used for optimization can be rather high, which makes it particularly difficult for the designer. Systems like these are easily stuck in a bad minimum, and the starting design is typically the key to getting to the desired solution. Yet, finding starting points is difficult due to the lack of patented systems of this type. SMS designs with exhausted degrees of freedom like the one presented in Figure 2.12 present a good starting point. This assertion was more deeply investigated in Chapter 3 where the traditional design approach of aspheric optics and SMS 2D method combined with local optimization were compared. A study conducted in [20] for SMS 3D method further demonstrates that SMS method can be a valuable tool for designing freeform optical systems for imaging applications.

REFERENCES

- [1] R. Descartes, *Le Livre Premier de La Géométrie*. Leyden, 1637.
- [2] W. T. Welford and R. Winston, "On the problem of ideal flux concentrators," *J. Opt. Soc. Am.*, pp. 531–534, 1978.

- [3] W. T. Welford and R. Winston, "On the problem of ideal flux concentrators : Addendum," *J. Opt. Soc. Am.*, vol. 69, no. 2, p. 20367, 1979.
- [4] J. Chaves, *Introduction to Nonimaging Optics*. CRC Press, Taylor & Francis Group, 2008.
- [5] J. C. Miñano, R. Mohedano, and P. Benítez, *Nonimaging Optics*. 2015.
- [6] J. C. Miñano, P. Benítez, W. Lin, J. M. Infante, F. Muñoz, and A. Santamaría, "An application of the SMS method for imaging designs.," *Opt. Express*, vol. 17, no. 26, pp. 24036–24044, 2009.
- [7] P. Benítez, J. C. Miñano, M. Nikolic, J. Liu, J. M. Infante, and F. Duerr, "Conditions for perfect focusing multiple point sources with the SMS design method," in *Proc. SPIE*, 2014, vol. 9191, p. 919102.
- [8] F. Muñoz, "Sistemas ópticos avanzados de gran compatibilidad con aplicaciones en formación de imagen y en iluminación," Universidad Politécnica de Madrid, 2004.
- [9] P. Benítez *et al.*, "Simultaneous multiple surface optical design method in three dimensions," *Opt. En*, vol. 43, no. 7, pp. 1489–1502, 2004.
- [10] P. Benítez, R. Mohedano, and J. C. Miñano, "Design in 3D geometry with the Simultaneous Multiple Surface design method of Nonimaging Optics," in *Proc. SPIE*, 1999, vol. 3781, no. July.
- [11] D. Grabovickic, P. Benítez, and J. C. Miñano, "Aspheric V-groove reflector design with the SMS method in two dimensions.," *Opt. Express*, vol. 19 Suppl 4, no. May, pp. A747–A756, 2011.
- [12] J. Liu, J. C. Miñano, P. Benítez, and W. Lin, "Single optical surface imaging designs with unconstrained object to image mapping," in *Proc. SPIE*, 2012, vol. 8550, pp. 1–7.
- [13] F. Muñoz, P. Benítez, and J. C. Miñano, "High-order aspherics : The SMS nonimaging design method applied to imaging optics," in *Proc. SPIE*, 2008, vol. 7100, pp. 1–9.
- [14] F. Duerr, P. Benítez, J. C. Miñano, Y. Meuret, and H. Thienpont, "Analytic design method for optimal imaging: coupling three ray sets using two free-form lens profiles," *Opt. Express*, vol. 20, no. 10, p. 10839, 2012.
- [15] F. Duerr, P. Benítez, J. C. Miñano, Y. Meuret, and H. Thienpont, "Analytic free-form lens design in 3D: coupling three ray sets using two lens surfaces," *Opt. Express*, vol. 20, no. 10, p. 10839, 2012.
- [16] J. M. Infante, "Optical Systems Design using the SMS method and Optimizations," Universidad Politécnica de Madrid, 2013.
- [17] J. C. Minaño, J. Liu, P. Benitez, M. Nikolic, and B. Narasimhan, "Anastigmatic Imaging with Unconstrained Object to Image Mapping," in *Classical Optics 2014, OSA Technical digest (online) (Optical Society of America 2014), paper ITh3A.1*.
- [18] W. Lin, P. Benítez, J. C. Miñano, J. M. Infante, and G. Biot, "SMS-based optimization strategy for ultra- compact SWIR telephoto lens design," *Opt. Express*, vol. 20, no. 9, pp.

- 9726–9735, 2012.
- [19] F. Duerr, Y. Meuret, and H. Thienpont, “Potential benefits of free-form optics in on-axis imaging applications with high aspect ratio,” *Opt. Express*, vol. 21, no. 25, pp. 31072–31081, 2013.
- [20] B. Satzer, U. Richter, U. Lippmann, G. S. Metzner, G. Notni, and H. Gross, “Using the 3D-SMS for finding starting configurations in imaging systems with freeform surfaces,” in *Proc SPIE*, 2015, vol. 9626, p. 96260Y.

PART 1

ASPHERIC OPTICS

Chapter 3

AN ANALYSIS OF AN SMS2D DESIGN AS A STARTING POINT FOR A TWO-LENS OPTIMIZATION PROBLEM

3.1. Introduction

Classic imaging design methods are based on aberration theory and multi-parametric optimization techniques [1]–[3]. With the development of precise manufacturing techniques, aspheric surfaces are no longer a luxury and therefore can be encountered quite often in modern designs. For optical systems with few mild aspheric surfaces described by a small polynomial order, the traditional techniques may provide many useful solutions. However, when more complex systems with many aspheric surfaces with a great number of free parameters are considered, the optimization approach can easily get trapped in a bad local minimum (like M_2 in Figure 3.1 (a)) when the starting point (S_2 in Figure 3.1 (a)) is far from the good local minimum of interest (M_1 in Figure 3.1 (a)). In such cases, the designer's experience proves to be a crucial factor when trying to get out of a bad minimum in search for a good solution. For instance, stepwise optimizations use a trajectory in the parameter space initially constrained to a lower dimension subspace (by optimizing only the low order coefficients). When a local minimum in that constrained subspace is reached (like M in Figure 3.1 (b)), the constrained subspace dimension is increased (by unfreezing some high order coefficients). In that larger dimension subspace, that design is no longer a minimum and it can progress further, as illustrated in Figure 3.1 (b). This change of the parameter subspace dimension can be done only once (two-step optimization) or more times, to the stepwise limit in which only one parameter is added in each step. However, even though this method can reach a good local minimum (M_1 from S_3 in Figure 3.1 (a)) this process can be very time consuming, and is very dependent on the designer's skills.

Figure 3.1. An illustration of an optical design landscape in 2D (a) and in 3D (b).

Other methods such as genetic algorithms [4] and simulated annealing [5] have been proposed to avoid getting trapped in a bad local minimum. However, these methods are based on generally applicable mathematical models and designers understand little about the optical systems behind them. Another method known as saddle point construction (SPC) method uses a systematic approach to find solutions via the design network, which helps to get out of a bad local minimum. [6], [7]

The Simultaneous Multiple Surfaces (SMS) method constitutes an alternative approach to this problem since it is a direct method that allows to find a better starting point (like S_1 in Figure 3.1 (a)) [8] [9] and progress from it with a single step optimization (varying all parameters). In this chapter, we are interested in rotationally symmetric designs, for which SMS 2D is used [10]. SMS 2D may use rays in the meridian plane or skew rays, but in any case obtaining a 2D profile of the surfaces. The method involves simultaneous calculation of N optical surfaces using N one-parameter bundles of rays for which specific conditions are imposed. The number of ray bundles can be greater than the number of surfaces to design when the footprints of the design bundles do not occupy the full SMS surfaces, as demonstrated in [11][12]. However, these cases were not discussed in this chapter.

The work presented here compares the designs obtained with an SMS starting point (described in section 3.2) and a single-step optimization of all parameters, with other designs obtained using spherical starting point design and single-step, two-step, stepwise and global optimizations. Two designs are considered, a simple one and a slightly more complex design as well (with a larger aperture and wider field). A detailed insight into the optimization process showing an issue with balancing is given in section 3.4.

3.2. Starting point with the SMS method

Standard SMS procedure consists of 2 steps: selection of central segments of the surfaces and recursive generalized Cartesian oval calculation. [14], [15] We will focus on SMS designs using only meridian rays, for which skew rays are typically controlled in a subsequent optimization.

Here we study two designs that consist of 4 aspheric surfaces, therefore we use SMS-4M design algorithm. The four meridian ray bundles selected for the SMS design correspond to the rays emitted from 4 object points placed at infinity symmetrically about the optical axis, and so are the image points located to match the specified focal length. The directions associated with the object points are selected so that their RMS_{2D} distribution curves (defined as the RMS spot diameter calculated using only meridian rays in the field) present a constant ripple over the field of view. [9]

Applying a standard SMS 2D procedure, for the design directions we simultaneously calculate the set of points and normal vectors. SMS profiles are then fitted into a Forbes Q-con polynomial [13] to be introduced to CODE V® software. The system is now ready to be optimized for the whole field of view, as the example shown in Figure 3.2.

Figure 3.2. SMS-4M design after fitting the surfaces to Forbes Q-con polynomials and introducing the system into CODE V®.

3.3. Comparing the results of different design approaches

Two lens systems were designed and optimized using SMS method and other approaches, which have been mentioned in the section 3.1. Both systems have four aspheric surfaces. Their specifications are given in Table 3.1. In the case of the smaller pupil size and field of view (FOV), it will be shown later in the paper that system 1 is a simple design task, and system 2 is slightly more difficult to design than system 1.

Table 3.1. Systems specifications

	Image F/Number	Effective Focal Length (EFL)	Maximum Half Field of View (MHFOV)
System 1	2.24	8.60 mm	7.50°
System 2	1.77	10.60 mm	11.25°

Three different design approaches were used for the comparison. They are: 1) SMS design with a single-step subsequent optimization; 2) optimization from a spherical system while gradually adding aspheric coefficients, either in a single step, two steps or one-by-one (stepwise); 3) global optimization with Global Synthesis (GS) in CODE V®. In order to have a fair comparison, the systems designed all use Q-con polynomial series [13] as the representation for the aspheric surfaces. The variables are the curvatures, normalization radii (a normalizing parameter in Qcon polynomial), the conic constants and the higher order coefficients of Q-con up to the 16th order. The vertices are fixed for all cases. We use a default CODE V® transverse ray aberration merit function (MF), based on a damped least squares method. The focal length of our system is constrained, and the only variables are the surface parameters. Normalization radii of the surfaces are constrained as well, to a value exceeding the size of the surface aperture for 5%. Vignetting was readjusted every 20 cycles, since the surface shape may vary significantly during the optimization process. Iterations are stopped when the merit function drops less than $10^{-8}\%$ (so convergence is assumed then). For GS optimization, the same MF and constraints are applied. However, the vignetting can only be adjusted after the GS is finished.

After the system was designed with SMS method, it was imported into CODE V® for further analysis and optimization. Using this design as a starting point, we used CODE V® to optimize the system for a better performance throughout the whole field of view, instead of stigmatic imaging for SMS design fields. The merit function for the optimization is the default CODE V® merit function based on transverse ray aberration. The left flow chart of Figure 3.3 shows the design flow with SMS.

Figure 3.3. Flow chart for different design approaches.

The second approach is conventional and intuitive. We started from a good spherical system. This spherical system was obtained with Global Synthesis (GS) of CODE V®. System with four spherical surfaces is simple enough so that we can trust that GS is able to find the best solution. Next, the aspheric coefficients were added. There are different ways for adding the freedom to the system. The abrupt way is to add all the variables at once and optimize. In our case, it means to free all 40 variables together and optimize. For an optimization landscape with multiple minima existing, designer will not have control of the optimization route and it is easy to get trapped in a bad local minimum. Experienced designers would often use a different approach, where the variables are added step by step. Here we apply three different approaches: a single-step, a two-step and an extensive stepwise approach. A two-step way was initiated by adding all the conic coefficients at once and then optimizing to the local minimum with conic surfaces. Subsequently, the higher order coefficients of Q-con polynomial were added all together and optimized. A more laborious intensive approach involves adding one freedom each time and then optimizing. We started with adding conic constant to the surface 1, and then the conic constant on surface 2, etc. Each new freedom was added on the surface following the previous one, as shown in Figure 3.4. Surface 1 is the surface that follows after the surface 4.

Figure 3.4. The strategy for adding freedom step by step. The blue shape represents each optical surface. We can imagine a matrix, where the row starts with conic constant and ends with the highest order used for Q-con polynomial, the column starts from surface 1 to surface 4. As the dashed arrow points, the freedoms are added from the up left corner to the bottom right corner.

In addition to the three approaches mentioned above, a global optimization method based on CODE V®'s Global Synthesis (GS) was compared as well. To apply the global optimization, we first started from spherical solution, which was mentioned in the previous paragraph. Then all the freedoms were added at once to the system. With a constraint on the focal length, GS was executed. A time limit of 2 hours is set for the GS, which means the algorithm is forced to stop after 2 hours if the GS is not finished. The GS result will list the name of the found systems and their MF value. The one with the smallest MF value is chosen for the comparison below. However, the designs found had an extremely small edge thickness, or even a negative one in some cases. To correct that, an additional edge thickness constraint was introduced (edge thickness larger than 0.2 mm) and the GS result was re-optimized in order to obtain a physically realizable system. The working flow is shown in Figure 3.3 on the right side.

The resulting system shapes are presented in the Figure 3.5. The thickness of the second element of the GS result is still quite small. Figure 3.6 shows the root mean square (RMS) spot size calculated at each field. Notice that the SMS original design has significantly poorer RMS far from the on-axis field. This is due to the use of only meridian rays in its design, and thus the skew rays are degrading the imaging quality off axis. After optimization, we also observe that there is an issue with field balancing in some of the designs, which will be discussed in more details in section 3.4. We suspect it is the specific issue of this system. However, final GS result is quite balanced. In Figure 3.6, we can see that all five systems generated by three approaches show good performance under 5°. When the field increases, the spot size of the solution obtained by two-step optimization strategy grows rapidly (Figure 3.6 (a)). In Figure 3.6 (b), it can be seen that the result of one-step and stepwise approach demonstrate fast increasing

behaviour, as well. In comparison, the results of SMS+OPT and the GS are rather balanced through the whole field.

Figure 3.5. Plots of the systems obtained using different design techniques.

(a)

(b)

Figure 3.6. RMS spot size distribution curves for different design approaches considered - complete curves (a); enlarged section (b).

We can see from Figure 3.6 (b) that both SMS+OPT and GS achieve good results with GS being slightly better. The system obtained with the two steps strategy exhibits the worst

performance in this case. It cannot even achieve the quality of an un-optimized SMS design. This is due to the fact that we have used many variables in the design (40 in total with 4 constrained normalized radii). With only two steps to optimize the system, there are many possible paths for the optimizer to follow. Therefore, the designer may lose the control of the system and there is no guarantee for the final solution. To overcome this, a good practice is to use the stepwise strategy, by starting with the lower order aspheric coefficients and gradually adding higher order terms. A gradual decrease of the MF with stepwise strategy often helps to avoid poor solutions, resulting in a deeper minimum. Surprisingly, for system 1 a single-step optimization finds a better solution, even better than that obtained with the stepwise optimization. We think this result is a consequence of this being an “easy problem” (of 1.5 μm Airy disc diameter), where with the degrees of freedom provided, it is easy to achieve diffraction limited performance. In this situation, many equally good minima exist in the solution space. Even though the designer still has no control during a one-step optimization, the chance for the optimization reaching a good minimum is higher than in a more difficult design problem.

From Figure 3.5 by looking at the curvatures of the surfaces, we notice that each system has a distinguished shape. This is because different strategies lead the optimization to a different basin of attraction in the optimization landscape. SMS leads the optimization into a different basin of attraction producing result better than manually adjusting the variable and optimizing (one-step, two-steps and stepwise). One-step optimization, on the other hand, leads to another basin of attraction which results to be slightly better than stepwise approach. The GS method generated 147 results. By looking at the system plots of these results, we roughly recognized several different system shapes. However, the purpose here is not to discuss the details of these GS results, and it is a laborious work to check every result from GS. We chose the system with the smallest MF. As we have mentioned previously, an additional optimization with the edge thickness constraint was needed for this system, and the final system results in a different basin of attraction compared to the other four approaches.

Figure 3.7 shows the comparison of the efficiencies of different strategies considered in this manuscript. GS is not included since it is a very different approach compared to the others, and we do not have an insight of the amount of cycles the software used to generate the results. The time GS took for system 1 was one hour, and 147 results were generated. It took us some time to analyse the results, select a system and do a subsequent optimization. In Figure 3.7 the starting points are shown at cycle 0. The differences between the final MF values for different design approaches are quite small (the GS result’s MF value is 0.015), and a quick MTF analysis shows that all five systems are diffraction limited. This is consistent with our statement that the design problem is easy and multiple equally good minima are present. Therefore, in simple design problems like system 1 less robust methods can either get trapped in a bad minimum (two-steps) or by chance find a path to the optimal solution (single-step). However, from Figure 3.6 it is clear that the RMS spot sizes of the obtained solution show more differences amongst them than the pure MF value. For instance, the MF value for a two-step and a stepwise solution seem to coincide. Nonetheless their RMS spot size distribution is

obviously different. Apparently, the MF of transverse ray aberration does not indicate accurately about the performance in this case.

Figure 3.7. Comparison of the efficiencies of different design methods analyzed, evaluated as the MF value versus the number of cycles performed. The starting points are shown at 0 cycles and are connected with the straight lines with the results obtained using different design methods. The logarithmic scale was used for the vertical axis.

System 1 has a relatively small pupil, which reduces the difficulty of the design. In comparison, we used SMS method to construct system 2, which has a larger pupil and a larger field of view. The same approaches were used for system 2. Figure 3.8 shows the system plots of the obtained results. Only four plots are shown given the fact that two-steps optimization ended in ray-failure situation in the second step. GS found 363 results in 2 hours, and again we chose the one with the smallest MF value with an additional edge thickness optimization. In Figure 3.8, different plots still show distinguished system shapes. However, from Figure 3.9, the RMS spot diameter can be divided into two groups. Although one-step optimization did converge, the result is a bad local minimum clearly different from others. SMS with optimization, stepwise optimization and GS all obtain good solutions with close performance (Figure 3.9 (b)). Different from the curves in Figure 3.6 (b), which are separated through the whole fields, the curves of the three best results start to intersect for larger field angles (around 7.8°). Notice also that the Airy disk diameter is $1.2 \mu\text{m}$ in this case. Since the system shapes of the three solutions are different, we believe that they present three equally good minima located in different basins of attraction. Different from system 1, a one-step optimization results in a poor local minimum (MF value 10.46 in Figure 3.10) and two-steps optimization went into ray failure, which indicates that less robust methods start to fail in a more difficult design task. Another evidence is that from the 363 results of GS, large amount of them went into ray-failure after we readjusted the vignetting and post-optimized. More trial and error was needed than in the case of the system 1 to find a GS result that is practical.

Figure 3.8. Plots of the systems obtained using different design techniques.

(a)

(b)

Figure 3.9. RMS spot size distribution curves for second system using different design approaches: complete curves (a); enlarged section (b).

Figure 3.10 shows the comparison of the efficiency of different methods. GS result is not put into the plot due to its complicated procedure. We see a similar pattern as in Figure 3.7 the design obtained with stepwise optimization is positioned at the far right in the plot with its large number of optimization cycles (3440 cycles). With a similar merit function value, fewer number of optimization cycles is needed for SMS optimization, which indicates a closer distance to the minimum in the optimization landscape.

Figure 3.10. Comparison of the efficiencies of different design methods considered evaluated as the MF value vs. the number of cycles performed. The starting points are shown at 0 cycles and are connected with the straight lines with the results obtained using different design methods. A logarithmic scale is used for the vertical axis.

3.4. An insight in optimization process

The RMS spot size vs. field angle curve of the two-step optimization of system 1 shown in Figure 3.6 is very unbalanced over the field. This is not unexpected, since the merit function does not penalize such unbalancing explicitly. This has occurred also in other designs during the optimization process, in particular in the single-step optimization from the SMS starting point also of system 1. The evolution of the merit function values is shown in Figure 3.11.

Figure 3.11. Merit function (transverse ray aberration from CODE V®) progress during the optimization.

It can be noticed that after 50 cycles the merit function progresses rather slowly. We were interested in knowing how this change in merit function reflects on the system performance. Therefore, we evaluated the RMS spot sizes throughout the field for selected designs during the optimization in order to see if there was a noticeable change in system functioning. Maximum and minimum values of the RMS spot size were evaluated every 20 cycles (Figure 3.12). The RMS spot size distribution curves for some specific designs during the optimization are plotted in Figure 3.13. These curves explicitly demonstrate the change in the balancing during the optimization process.

Figure 3.12. Progress of maximum RMS spot size and minimum RMS spot size of all fields during the optimization (evaluated every 20 cycles).

Figure 3.13. RMS Spot Size distribution curves for systems in different stages of optimization.

We observe that around the 900th optimization cycle, merit function value makes a more significant decrease (a step in Figure 3.11). This results in the system being unbalanced, as can be observed in Figure 3.13. Namely, the maximum RMS spot size (typically appearing at the

edge of the field of view, as in Figure 3.13) starts to grow, whereas the minimum RMS spot size (on-axis direction) continues to decrease (Figure 3.12 and 3.13). The merit function value alone does not show us this peculiar behaviour. This problem of balancing can be fixed by using a merit function that recalculates the field weights during the optimization in order to balance the performance, as the one suggested by Cheng *et al.* in [16]. However, since no balancing method was applied here, even though the merit function value is dropping, the average RMS spot size does not necessarily do so, which is due to the peak in maximum RMS spot size, previously mentioned. This characteristic is best observed starting from the 900th cycle, where, while the merit function continues to drop, the average RMS spot size clearly increases. We believe this is due to the aperture weights applied in the construction of the default CODE V[®] merit function which puts more emphasis on the rays in the central part of the aperture, if not specified otherwise. [17] This, as we progress in optimization becomes more important due to the significant change in surface shapes. For the purpose of this study we chose a system obtained after 1400 cycles, since it gives a low RMS value throughout the field with a small peak at the edge.

3.5. Conclusions

Two different optical systems comprising of two lenses were analysed: a simple case and a slightly more difficult one with the increased field and pupil. A comparison of SMS 2D designs as starting points with traditionally used spherical systems was conducted. SMS 2D designs were optimized in a single-step, while the spherical starting points were optimized using a one-step, a two-step and a stepwise approach. Global optimization method from CODE V[®] software was compared as well. In the case of the simpler system analysed, all design techniques considered found solutions close to a diffraction limit. Nevertheless, the surface shapes of the final designs were different which suggested that their optimization paths ended in different basins of attractions. The fact that there are various minima with diffraction limited performance further proves that the solution to this problem is an easy one. Furthermore, a balancing issue during the optimization which puts an emphasis on lowering the MF value for the on-axis fields was encountered. Posterior SMS optimization and global synthesis approach were able to surpass this issue without any additional requirements put in the merit function, providing rather balanced solutions.

The second system analysed had a larger field and a bigger pupil. In this case only the SMS with subsequent optimization, a stepwise approach and GS found optimal solutions. Carefully analysing the three results obtained, we conjectured that they belong to three different minima, which is to be expected since different optimization strategies lead to different basins of attraction. Since designing system 2 is more demanding than system 1, less robust methods such as one-step optimization and two-steps optimization did not lead to good minima. Despite that three approaches found good solutions, stepwise optimization usually takes the largest

amount of optimization cycles. It also needs frequent interactions between the optical designer and designing tools. GS works well in both cases. However, the difficulties of the approach lay in choosing some systems among the generated results for the next design phase. Moreover, with large number of variables, GS tends to run for a long time and generate several hundreds of results. In this case, it is necessary to set a time limit constraint for GS execution. In both cases, SMS with posterior optimization can find good solutions. Therefore, the assertion that the SMS method provides a good starting point for optimization seems to be correct, at least in the examples considered.

REFERENCES

- [1] R. Kingslake and R. B. Johnson, *Lens Design Fundamentals*, no. 2nd edition. 1978.
- [2] W. J. Smith, "Modern Lens Design." McGraw-Hill, 2004.
- [3] R. E. Fischer, Bi. Tadic-Galeb, and P. R. Yoder, *Optical System Design*. 2008.
- [4] K. E. Moore, "Algorithm for global optimization of optical systems based on genetic competition," in *Proc. SPIE 3780*, 1999, pp. 40–47.
- [5] G. W. Forbes and A. E. Jones, "Towards global optimization with adaptive simulated annealing," in *Proc. SPIE 1354*, 1991, p. 144.
- [6] F. Bociort and M. van Turnhout, "Finding new local minima in lens design landscapes by constructing saddle points," *Opt. Eng.*, vol. 48, no. 6, p. 63001, 2009.
- [7] Z. Hou, I. Livshits, and F. Bociort, "One-dimensional searches for finding new lens design solutions efficiently," *Appl. Opt.*, vol. 55, no. 36, p. 10449, 2016.
- [8] W. Lin, "Advances in the Simultaneous Multiple Surface optical design method for imaging and non-imaging applications," *Universidad Politécnica de Madrid*, 2012.
- [9] W. Lin, P. Benítez, J. C. Miñano, J. M. Infante, and G. Biot, "SMS-based optimization strategy for ultra- compact SWIR telephoto lens design," *Opt. Express*, vol. 20, no. 9, pp. 9726–9735, 2012.
- [10] J. Chaves, *Introduction to Nonimaging Optics*. CRC Press, Taylor & Francis Group, 2008.
- [11] P. Benítez, J. C. Miñano, M. Nikolic, J. Liu, J. M. Infante, and F. Duerr, "Conditions for perfect focusing multiple point sources with the SMS design method," in *Proc. SPIE*, 2014, vol. 9191, p. 919102.
- [12] F. Duerr, P. Benítez, J. C. Miñano, Y. Meuret, and H. Thienpont, "Analytic free-form lens design in 3D: coupling three ray sets using two lens surfaces," *Opt. Express*, vol. 20, no. 10, p. 10839, 2012.
- [13] G. W. Forbes, "Shape specification for axially symmetric optical surfaces.," *Opt. Express*, vol. 15, no. 8, pp. 5218–5226, 2007.

- [14] J. C. Miñano, P. Benítez, W. Lin, J. M. Infante, F. Muñoz, and A. Santamaría, “An application of the SMS method for imaging designs.,” *Opt. Express*, vol. 17, no. 26, pp. 24036–24044, 2009.
- [15] W. Lin, P. Benítez, J. C. Miñano, J. Infante, and G. Biot, “Progress in the SMS design method for imaging optics,” in *Proc SPIE*, 2011.
- [16] D. Cheng, Y. Wang, and H. Hua, “Automatic image performance balancing in lens optimization,” *Opt. Express*, vol. 18, no. 11, pp. 11574–11588, 2010.
- [17] “CODE V®Reference Manuals”.

Chapter 4

FOVEATED IMAGING DESIGNS

4.1. Introduction

In optical design, there are typically two methods to increase the field of view of an imaging system while still maintaining high image quality. The first method is to reduce the entrance pupil, and thus increase the $f/\#$ of the system. The second approach is to add many optical elements in order to remove off-axis aberrations that appear when the field of view is increased. Adding elements undoubtedly increases the complexity, size, weight and cost of the system. However, this was traditionally the only way to maintain high image quality at the end of the field while keeping the desired entrance pupil size, and a large field of view. Nonetheless, there are many applications where size and weight are very important. For these applications, one of the alternatives to standard approach known as foveated imaging could be rather useful.

Foveated imaging approach refers to optical systems that create images with varying resolution. Some of these systems mimic the behaviour of the human eye (Figure 4.1 and Figure 4.2) while others, in which the human eye is the final viewer, use that behaviour to adapt the optics performance to it.

Figure 4.1. Structure of a human eye ball.

Figure 4.2. Angular eye resolution in arcmin as a function of a peripheral angle given in degrees.

Human eye separates the field of view into two resolution regions - one with high and one with low resolution. The region with high resolution is known as *fovea centralis*, and it is located at the centre of the field of view. Low resolution region is called a peripheral zone and the resolution decreases as we move further from the fovea. The advantage of this type of system is the ability to see over a wide FOV while maintaining high resolution in the central region.

A pioneer demonstration of a foveated imaging system was presented in 1989 by Spiegel *et al.* [1] The authors described an “artificial retina” as a custom-designed sensor array with spatially variant response. Following this work, many alternative approaches for customized on-chip foveated sensor arrays were presented. [2], [3]

The first attempt to achieve foveation with optical components was done by Suematsu *et al.* where the authors propose a door peephole lens together with the distortion correction software to obtain foveated images. [4] Figure 4.3 illustrates the concept of their vision sensor with fovea. They later presented a higher performance lens that had similar distortion characteristics over a larger image. [5] Further development of this approach to foveation was done by Wakamiya *et al.* [6] This group presented a wide-angle camera lens adapted to a standard CCD array format. A similar approach was used by Hagen and Tkaczyk to design a foveated endoscopic lens. [7] They use optical distortion within a customized objective lens, where they attempt to capture all of the benefits of foveation: wide field of view, high-resolution central field and easier optical design. In order to achieve all the aspects mentioned they use highly aspherical optical elements.

Figure 4.3. A sketch of the vision sensor with fovea presented by Suematsu *et al.*
(Source: Ref. [4])

Alternative approach to foveated imaging presented by Martinez and Wick uses a transmissive [8] or a reflective [9] liquid crystal spatial light modulator (SLM) to dynamically correct the aberrations in a localized region of the field of view. Another interesting design presented by Rolland *et al.* is a head-mounted display with high-resolution insets [10].

Ude *et al.* presented a pragmatic approach that combines two cameras viewing the same scene in order to obtain foveation. [11] One camera is a high-resolution low-FOV camera, and the other one is a low-resolution high-FOV camera. However, this method can produce parallax problems when viewing nearby objects. Hua and Liu adapted the previous method to use a beamsplitter to separate the objective lens optical path into two paths of different magnification, each of which is sampled by its own detector array. [12]

Two designs presented in this chapter both use foveation with optical components, where a distorted image generated by the optical system is corrected via software using image post processing.

The first part of this chapter is dedicated to exploring the benefits of using SMS method with pupil discretization in foveated imaging systems. We demonstrate how SMS method can be used as a starting point for designing optical systems with prescribed distortion. The method was applied to an existing foveated endoscopic optical system, where only the aspheric part of the system was redesigned. The objective of this study was to find milder aspheric surfaces that would be easier to manufacture than the ones initially placed in the design.

Afterwards we present the second design, which is a foveated imaging design for a virtual reality head mounted display (VR HMD) applications. We include a detailed explanation of a design approach used, as well as tolerance analysis and experimental resolution measurements.

4.2. Obtaining a starting point from SMS2D with pupil discretization

4.2.1. Foveated imaging in endoscopic lenses

Endomicroscopy is a technique for obtaining histology-like images from the inside of the human body in real-time. Current endomicroscopes lack the field of view needed to provide the proper context for the high-resolution measurements needed to detect lesions. [7] Clinicians first do a routine scan of the entire tissue section with low resolution, and then focus down on areas of interest with high magnification and resolution. High resolution images allow one to determine the health and morphology of individual cells. However, this information is only useful when the cells' location within the overall tissue is known. Using previously described foveation approach, endomicroscopes with both large field of view and high resolution in the central region of the image can be obtained. Some designs were already presented in the literature. [7] [13]

Foveated imaging design approach can be seen as designing a lens that produces a specific prescribed distortion, obtaining in that way a wider field of view with varying magnification. Distortion is afterwards corrected using image post processing algorithms. This approach can be observed as an intermediate point between an optical zoom and a digital zoom. Optical zoom gives high resolution but it involves many mechanical difficulties, since the focal length of the system is changed mechanically moving the elements. This results in slow and expensive solutions. Digital zoom, however is much faster and the whole zoom process is done by software, thus the price is significantly lower. Nonetheless, image obtained in this way has low resolution when high magnification is applied. Using foveated imaging approach, we are able to design systems that would have high resolution with the quality of the optical zoom at the centre of the image, which is the main region of interest, without any moving elements in the system.

4.2.2. Application of SMS method with pupil discretization to a foveated microscope objective lens

We started with a working microscope objective system presented in Figure 4.4. A mapping function and a change in magnification for this system are shown in Figure 4.5.

Figure 4.4. Starting system, with a zoomed region of interest containing two strong aspheres (marked with the red dashed circle).

Figure 4.5. Change in magnification and image height versus field height.

The presented system was designed for a visible spectrum, with a working distance of 93.3 mm, and an object to image total track of 335 mm. The high distortion of 35.05% (Figure 4.6) was achieved using two aspheric surfaces positioned around the intermediate image, as can be observed in Figure 4.4. However, the mentioned aspheres could not be manufactured with the desired precision. Our task was to redesign the aspheric part of the system, increasing the number of aspheres to 4, the goal being to find a solution with milder aspheric surfaces. An indirect goal was to see whether a direct design method, such as the SMS method could turn out to be useful in this type of problems.

Figure 4.6. Image showing the effect of distortion on the object.

Bearing in mind that we already have a functioning system, SMS method in this case is used as an optimization tool. Yet, we do need to design four new surfaces that would not alter the system specification.

Application of SMS method with object discretization would make it very difficult to obtain the prescribed mapping, since only two wavefronts would be used, due to the symmetry of the system. This would lead to an SMS solution far from our goal design, which is not desirable. Applying the theory presented in Chapter 2, which states that a greater number of fields can be used to design a smaller number of surfaces when the bundle size passing through the surface is effectively smaller than the surface aperture [14], could make the problem easier. The aspheres in the system are positioned around the intermediate image; therefore, the bundle size is quite small, which would allow for sharply imaging multiple object points. Applying the rule of thumb presented in Chapter 2 we could estimate the number of object points that can be sharply imaged through four SMS aspheric surfaces. Using this approach we could design four aspheres that follow the prescribed mapping.

An alternative to the method presented in Chapter 2 would be the SMS method with pupil discretization (Figure 4.7), introduced in Chapter 1, Section 1.3.3. More details regarding this version of SMS method can be found in ref. [15], [16]. Using pupil discretization we can use as many object points as we need to, making it possible to represent in detail the wanted object-to-image mapping. Instead of discretizing the object, we now discretize the pupil, choosing four specific points where the rays from the object will pass.

Figure 4.7. SMS method with pupil discretization. In the example illustrated, discretization is done at the exit pupil, although the entrance pupil or the aperture stop of the system can also be chosen.

First step in any SMS design involves fixing the initial conditions. That includes choosing the convenient location for the 4 surfaces that are to be designed. Due to the existence of the intermediate image in the system and high distortion prescription, the best location for aspheric surfaces is exactly around the intermediate image, where the initial two aspheres are already positioned. Hence, we positioned the two new vertices around the intermediate image. These new surfaces will form an additional aspheric lens, with thickness of 1 mm. The position of the other two vertices remains the same as in the initial design.

Next we choose the pupil where the 4 design points will be specified. Bearing in mind that this system has different NAs for different field heights on both object and image side, it does not have a unique entrance and exit pupil. Yet, the aperture stop of the system is unique for all the fields; thus, we specified our design points at the aperture stop. From all the points on the object 4 rays were traced, each passing through the chosen point at the aperture stop and going to the image point prescribed by the mapping function. The rest of the surfaces in the system were maintained as in the initial design.

The aspheric part of the system was redesigned using this approach. Changing the seed patch of the SMS surfaces, different design forms were obtained, three of which are shown in Figure 4.8.

Figure 4.8. Three different sets of aspheres (shown in the zoom on the right) obtained with the SMS method with pupil discretization.

Notice that, for the same position of vertices, different SMS designs were obtained. The object-to-image mapping for the design rays was identical to the prescribed one. The surfaces were then introduced into a commercial software for further optimization and analysis for the whole pupil. For that purpose, Forbes Q-con representation was used. [17], [18]

All presented designs showed similar behaviour. Monochromatic RMS spot size profiles plotted in Figure 4.9 demonstrate that. However, the shapes of the surfaces vary significantly. All presented solutions are found by changing the seed patch (which includes curvature and conic constant of the surfaces) and redoing the SMS process. In this way we are able to explore SMS landscape and search for different solutions.

Figure 4.9. Monochromatic RMS spot size for the whole pupil, evaluated for 580 nm.

These systems now present a starting point for further optimization. The surfaces need to be optimized for the whole pupil, and for the visible spectrum, since the SMS designs obtained here are monochromatic. The optimization is done in CODE V® software, where the only variables are the surface parameters of the 4 newly designed aspheres. Two optimized SMS designs are presented in Figure 4.10, since SMS design 1 and 2 during the optimization converge to a similar final design. Monochromatic and polychromatic RMS spot sizes show superior behaviour of new designs in comparison with the initial one (Figure 4.11). Yet, this is to be expected, since the number of surfaces is increased.

Figure 4.10. Two resulting optimized SMS surfaces (shown enlarged on the right).

Figure 4.11. Monochromatic (up), and polychromatic (down) RMS spot size curves calculated for two optimized SMS designs.

The goal of this study was to find milder surfaces, easier to manufacture. Two presented designs use Q-con representation, where the first design uses coefficients of up to the 16th order, while the second one uses only up to the 10th order. Initial design included aspheric coefficients of up to the 14th order. However, the order of the polynomial is not necessarily the indicator of the manufacturability of the surface. A more detailed analysis including slope and sag departure from the best fit sphere is conducted. Table 4.1 presented below shows that the second design obtained is milder than the initial design with two aspheric surfaces, but the first one turns out to be extremely aspheric. The asphericity, however, is not a consequence of the SMS design procedure, since the initial SMS designs are optimized further. Namely, the high asphericity in foveated designs is probably difficult to avoid due to the high prescribed distortion. For applications such as VR HMD optics presented in section 4.3 where the spatial frequencies of interest are in general lower (due to their large FOV and limited resolution of digital displays available), errors introduced with available manufacturing techniques for aspheric surfaces are in general acceptable. However, for more challenging applications such as endomicroscopy, where only very small manufacturing errors can be allowed the importance of asphericity increases.

Table 4.1 Peak and RMS sag departure from the best fit sphere (BFS), as well as Peak and RMS slope, evaluated for the initial aspheres, as well as the newly obtained surfaces using the SMS method with subsequent optimization.

	Surface	RMS sag dep from BFS	Peak sag dep from BFS	RMS slope	Peak slope
Initial	Surface 1	0.24	0.57	0.64	11.51
	Surface 2	1.55	-30.48	8.48	-150.57
System1	Surface 1	12.15	12.25	6.56	-102.43
	Surface 2	8.05	-130.76	33.06	-272.58
	Surface 3	136.70	138.29	32.87	-188.43
	Surface 4	201.9	-202.11	9.71	160.88
System2	Surface 1	0.04	0.14	0.13	-1.41
	Surface 2	4.01	-4.11	1.66	16.07
	Surface 3	0.28	3.15	1.28	10.8
	Surface 4	5.27	-5.33	0.72	8.43

An indirect goal, that included exploring the possibilities of SMS method in foveated imaging designs, was achieved. Using this method we were able to find different design geometries, changing just one input parameter. Downside of this approach is that we are not able to control surface asphericity a priori. We are obliged to do the whole design, interpolate the surfaces and then analyse them in order to evaluate the manufacturability. However, since these designs need additional optimization that includes the full pupil (not only the four rays per field) manufacturability constraints in surface shape could be added to the subsequent optimization merit function.

4.3. Super resolution VR HMD optics

4.3.1. Introduction to VR HMD optics

Head Mounted Display (HMD) technology is a rapidly developing field. [19] One aspect of HMD provides a full immersive visual environment, which is commonly referred to as virtual reality (VR). In VR HMD user only observes the images provided by one or more displays, while the outside environment is visually blocked. These devices have application in areas such as entertainment, gaming, military, medicine and industry.

An ideal VR HMD combines a high resolution, a large field of view, a low and well-distributed weight, and a structure with small dimensions. Although some technologies successfully achieve these desired features individually, so far no known technology has been able to combine all of them. That results in an incomplete or even uncomfortable experience

for the user. Problems may include low resolution or imaging quality, small FOV, which thus leads to inability to create an immersive environment, and excessive weight that creates a lot of pressure on the user's head.

Let us now introduce the concept of opixel and ipixel as presented in [20]. The opixel is the physical pixel of the digital display, while the ipixel is the virtual image produced by the optics of the opixel in the VR screen. This virtual image is preferably formed at a certain distance from the eye (from one meters to infinity).

Since VR HMD devices are intended for direct human use, optical systems that are adapted to the behaviour of the human eye are desirable. This is what foveated imaging is all about. However, since these systems contain a digital display, they will be limited by the resolution of the display, as well.

Eye resolution curve adapted to the VR HMD system is given in Figure 4.12. The eye resolution presented here is shown in cycles per degree (which is just a half of the reciprocal of the values in Figure 4.2) and is limited by the Nyquist frequency of the ipixels.

Figure 4.12. Human eye resolution in cycles per degree limited by the resolution of the display used in a VR HMD, given as the Nyquist frequency.

4.3.2. Mapping function and focal lengths for HMD optics

In reference [20] (ρ, ϕ) are defined as the polar coordinates of a point r on the digital display and θ and φ as the polar and azimuthal angles of the spherical coordinates on the virtual screen, respectively. The horizon line is defined by $\varphi = 0$ and $\varphi = \pi$ and $\theta = 0$ is the front direction. The mapping function is then given by

$$r(\theta, \varphi) = (\rho(\theta, \varphi) \cos(\phi(\theta, \varphi)), \rho(\theta, \varphi) \sin(\phi(\theta, \varphi))). \quad (4.1)$$

The radial and sagittal focal lengths (f_{rad} and f_{sag} , respectively) at the virtual screen for the direction (θ, φ) are

$$f_{rad} = |\mathbf{r}_\theta| \quad (4.2)$$

$$f_{sag} = \frac{|\mathbf{r}_\varphi|}{\sin \theta}, \quad (4.3)$$

where the sub-indices on the right present partial derivatives. For other directions different than radial or sagittal, the focal length is given by:

$$f_\alpha = \left| \mathbf{r}_\theta \cos \alpha + \frac{\mathbf{r}_\varphi \sin \alpha}{\sin \theta} \right|, \quad (4.4)$$

where α is the angle formed by the direction along which the focal distance is calculated with respect to the radial direction. The focal length gives information about the amount of expansion or shrinking of the mapping in a particular direction. When the mapping between the object and the image is conformal, then f_α is independent of α , i.e. the mapping expansion or shrinking is isotropic.

The angular extent of an ipixel along the direction α is the physical opixel diameter over the focal length:

$$D_{ip,\alpha} = \frac{D_{op}}{f_\alpha}, \quad (4.5)$$

where sub-indices ip and op present ipixel and opixel, respectively. Therefore, the ipixel size in the direction α is inversely proportional to the focal length f_α . For simplicity, circular opixels are considered, but the reasoning can easily be extended to the usual square opixels. Human eye resolution depends on the peripheral angle as shown in Figure 4.2 and Figure 4.12 but it is isotropic in good approximation, thus it does not depend on α . Hence, it is desirable that the angular extent of the ipixels $D_{ip,\alpha}$ be independent of α . Since the diameter of the opixels D_{op} is in general mostly constant with α then, from Eq. (4.5) we can deduce that f_α independent of α is desirable.

Assume now a rotationally symmetric imaging system where the optical axis is the direction $\theta = 0^\circ$. This direction is imaged on the digital display at the point $\rho = 0$. Because of the rotational symmetry, the mapping function is such that ρ only depends on θ , i.e. $\rho = \rho(\theta)$, and $d\rho/d\theta = 1$. The radial and the sagittal focal lengths are

$$f_{rad} = |\rho_\theta| \quad (4.6)$$

$$f_{sag} = \frac{\rho}{\sin \theta}. \quad (4.7)$$

The mapping of a standard imaging optical system usually presents a rectilinear projection described by

$$\rho(\theta) = f \tan \theta \quad (4.8)$$

where f is a constant. The radial and sagittal focal lengths now coincide

$$f_{rad} = f_{sag} = \frac{f}{\cos \theta} \quad (4.9)$$

It can be noticed from Eq. (4.9) that at the centre of the virtual screen, where $\theta = 0^\circ$, focal lengths are equal, i.e. $f = f_{rad} = f_{sag}$. It can also be observed that the focal lengths are increasing functions of θ . If the rectilinear mapping is used, ipixels would progressively shrink moving towards the periphery, instead of becoming larger. At the edge of the directly gazed region of the virtual screen (which is in general considered to be $\theta = 20^\circ$), the ipixels would be 6% smaller than at $\theta = 0^\circ$, and at the edge of a 100° field of view ($\theta = 50^\circ$) the ipixels would be 35% smaller than at $\theta = 0^\circ$. This is not useful in practice since the eye will never directly look at those edge pixels.

Single lens HMDs usually deviate from the rectilinear mapping, presenting some pincushion distortion, being closer to the linear mapping given as

$$\rho(\theta) = f\theta \quad (4.10)$$

where again f is a constant equal to $f_{rad} = f_{sag}$ at the centre of the virtual screen $\theta = 0^\circ$. [19] For this mapping function focal lengths are given as

$$f_{rad} = f \quad (4.11)$$

$$f_{sag} = \frac{f\theta}{\sin \theta}. \quad (4.12)$$

In this case f_{rad} is constant for all values of θ and f_{sag} is an increasing function of θ . Even though in this mapping ipixels grow less than with the rectilinear mapping, still at $\theta = 45^\circ$ the size of the ipixels in the radial direction is equal to the size of those at $\theta = 0^\circ$, whereas in the sagittal direction they are 11% smaller than the ipixels at $\theta = 0^\circ$. Therefore, the optics is still not taking advantage of the fact that the ipixels can be progressively larger from $\theta > 20^\circ$ without our vision system (the eye) detecting the increase.

Alternative to the above presented mapping functions is a foveated mapping (sometimes also referred to as adapted mapping), where the focal length is adapted to the resolution of the human eye. A radial focal length is in this case a decreasing function of θ with an approximately constant slope outside the gazed region of the virtual screen. In this way the ipixels become significantly larger in the radial direction.

Figure 4.13 presents the mapping function of a design with foveation, as well as the mapping functions of the rectilinear and linear mapping previously introduced as presented in ref. [20]. All of them operate with the same FOV of 100° and the same 5.7" digital display of 16:9 aspect ratio. Each eye uses one half of the display, thus the distance from the centre of the half display to the closer point of the edge is about 32 mm.

Figure 4.13. The mapping function $\rho = \rho(\theta)$ of the three different mappings: linear, rectilinear and foveated mapping. [20]

The three mapping curves plotted in Figure 4.13 start and end at the same points, where the starting point is the centre of the virtual screen ($\theta = 0^\circ$, $\rho = 0$ mm) and the end point is ($\theta_{\text{end}} = 50^\circ$, $\rho_{\text{end}} = 32$ mm). The curve that represents the foveated mapping starts at the origin with a derivative higher than the ratio $\rho_{\text{end}}/\theta_{\text{end}}$ but ends with a smaller derivative than the same ratio. Since these derivatives basically present the radial focal length, this indicates that the system that follows foveated mapping will have a much greater focal length at the centre of the virtual screen than its counterpart system that follows rectilinear or linear mapping. In other words, it will have smaller pixels in the central area of virtual screen. This is achieved by making radial focal length at the periphery smaller, thus the ipixels there are larger in the radial direction.

A graph with the radial focal length that corresponds to the three previously described mapping functions is shown in Figure 4.14.

Figure 4.14. The radial focal length change with θ for the same three systems as in Figure 4.13. [20]

The sagittal focal length is not reduced as much as the radial focal length, so the ipixels in the peripheral region are strongly elongated in the radial direction. Full advantage of a system with adapted foveated resolution can only be obtained by breaking the rotational symmetry as discussed in section 7 of ref. [20].

The system presented here is designed for a 2560 x 1440 display with the opixel pitch of 50 μm , and the pixel density of 518 ppi. Table 4.2 shows a comparison of focal length, ipixel size and density and Nyquist frequency value on the virtual screen (ipixels) in the centre and at the edge of the virtual screen. The values mentioned are calculated for the three types of mapping previously presented.

Table 4.2 Comparison of the resulting parameters for a 5.7" display with 2560 x 1440 pixels at the centre ($\theta = 0^\circ$) and at the edge of the virtual screen ($\theta = 50^\circ$). (Source: reference [20])

Centre of the virtual screen ($\theta = 0^\circ$)				
Mapping	Focal length (mm)	ipixel diameter (arcmin)	ipixel density (ipixels/deg)	Nyquist freq. on virtual screen (cpd)
Rectilinear	26.6	6.4	9.4	4.6
Linear	36.5	4.7	12.8	6.4
Foveated	57.5	3.0	20.2	10.1
Edge of the virtual screen ($\theta = 50^\circ$)				
Mapping	Radial focal length (mm)	ipixel radial size (arcmin)	ipixel density (ipixels/deg)	Nyquist freq. on virtual screen (cpd)
Rectilinear	40.8	4.2	14.4	7.2
Linear	36.5	4.7	12.8	6.4
Foveated	6.0	28.4	2.1	1.1

It can be observed that by using the foveation approach, the focal length can be significantly increased in the central region of the virtual screen, resulting in decreasing the size of the ipixels in the same region. Moreover, the ipixel density in the central region of the virtual screen is increased. Since the eye resolves up to 2 arcmin (Figure 4.2), the size of the ipixel of 3 arcmin (Table 4.2) is still distinguishable. However, the distinction is considerably smaller than in the case of the other two mappings. This was achieved by reducing the radial focal length at the peripheral regions far off axis. At the edge of the virtual screen, for a peripheral angle of $\theta = 50^\circ$, the radial size of the ipixel is 28 arcmin, which results in a very poor spatial resolution. Nevertheless, in Figure 4.2 we can see that human eye resolution for the same $\theta = 50^\circ$ angle is around 50 arcmin; therefore the bad resolution at the edge of the virtual screen will not be noticed. However, we are not merely interested in designing a system in which the eye only gazes frontward. Gazing directions of up to $\theta = 20^\circ$ should be considered. For the $\theta = 20^\circ$ gazing direction, previously evaluated 50° peripheral angle now becomes approximately 30° . Eye resolution for 30° is around 30 arcmin, which is close to the obtained ipixel size (28 arcmin). Therefore, the resolution obtained coincides with what the eye is able to resolve.

4.3.3. Eye rotation design approach

An important feature that should be taken into account when designing HMD systems involves the rotation of the eye, altering the directly gazed region. The extent of the usually gazed region of the virtual screen is generally considered to be up to $\theta = 20^\circ$, where θ presents the rotation of the eye. Specifying image quality considering the eye rotations is a common practice in spectacle lens design as explained in ref. [21] [22]. In spectacle lens design the rays are traced through lenses and eyes with finite pupil sizes, which are rotated about their centre of rotation so that the fovea stays aligned with off-axis points of interest. [23] Taking into account the eye rotations in VR HMD optical systems, the design is optimized for the sharpest image of the ipixels to be obtained when they are directly gazed, thus the eye images them on the fovea (as in spectacle lens design approach).

The difference between spectacle lens design and HMD optical design is that in the first case point pupil is normally considered, whereas in the case of HMD optical design, a real human pupil size needs to be considered. The normal pupil size in adults varies between 2 to 5 mm, depending on the lighting. In the design presented here, we considered a pupil size of 3.5 mm.

Using the above explained reasoning, for the HFOV of 50° , when the eye is gazing frontwards (central pupil position, $\theta = 0^\circ$), the image quality of these rays can be progressively relaxed when the peripheral angle increases, as permitted by the decreasing human eye resolution (Figure 4.12). By rotating the eye pupil for an angle θ with respect to its central position, the directly gazed region is moved, and overlaps with a peripheral region observed when the eye pupil is centered (Figure 4.15). This implies that we have to design simultaneously for the central pupil position and different rotated pupil positions, where the emphasis will always be on the directly gazed regions.

Figure 4.15. An explanatory illustration of a difference between a directly gazed and a peripheral region (in the image, the two coincide). The directly gazed virtual image region is limited by the resolution of the display (the Nyquist frequency), while the virtual image region observed peripherally is limited by the eye resolution at that specific angle.

4.3.4. Design example

The design we present here combines foveation and eye-rotation design techniques. Resulting variable magnification from foveated imaging approach makes ipixels denser where they can be directly gazed, and coarser in peripheral regions, whereas eye rotation makes emphasis on the image quality of those directly gazed ipixels, minimizing the aberrations in those regions compared to the peripheral ones.

The system was designed using the usual design through optimization procedure. A standard Transverse ray aberration merit function was used in combination with user defined parameters that include the mapping function and focal lengths given in Figure 4.13 and Figure 4.14. The system specification is presented in Table 4.3 and the system layout is presented in Figure 4.16.

Table 4.3 System specification for a foveated VR HMD design presented here.

Basic system parameters	
Eye relief	12 mm
Distance Eye-Display	65.9 mm
Virtual Image	1.5 m
Pupil	3.5 mm
Wavelength band	Green Visible Central wavelength 540nm FWHM 70nm
Eye rotation (gazing directions)	0° - 33°
Full field of view	Circular of 99°
Focal distance (for 0° field)	57.5 mm
Lens materials	PMMA Polystyrene
Digital display: AMOLED	
Size	5.7" (145 mm)
Resolution	2560 x 1440 (~518 ppi)
Opixel pitch	49.3 μ m
Nyquist frequency	10.1 cyc/mm

Figure 4.16. A system layout of a foveated two-lens design for VR HMD application.

Mapping function and the focal length graphs of the resulting system are shown in Figure 4.17. They resemble very closely the desired corresponding functions presented previously in Figure 4.13 and Figure 4.14.

Figure 4.17. Centroid position on the display (left) and focal length distribution (right) for different peripheral angles for the resulting design form presented in Figure 4.16.

Polychromatic RMS spot size distribution curves evaluated at the display are shown in Figure 4.18. Eye rotations (i.e. gazing directions) of up to 33° were evaluated, as well as the whole FOV (100° in total) for the fixed eye position on-axis (0°).

Figure 4.18. RMS spot size distribution curves for gazing directions of different eye rotations (left) and for a fixed on-axis eye position (right).

The polychromatic RMS spot diameter of the rays impinging on the display when the eye is gazing the ipixels is in the 10-50 μm range (not exceeding the 49.3 μm opixel pitch) for gazing directions of up to 20°, which is the extent of the usually gazed region of the virtual screen (Figure 4.18, left). Notice that the image quality for peripheral regions is significantly lower than for the directly gazed ones (Figure 4.18, right), as specified by the design requirements.

Normalized polychromatic 3D RMS spot size plots for gazing direction of $\pm 30^\circ$ of the resulting system for a pupil size of 3.5 mm, as well as the whole FOV of $\pm 50^\circ$ for the on-axis fixed pupil position are shown in Figure 4.19. These plots give an insight of how the image quality will degrade progressively when moving off-axis.

Figure 4.19. Normalized 3D RMS plot for a 3.5 mm pupil for $\pm 30^\circ$ gazing directions (left) and a fixed eye position evaluated for a FOV of $\pm 50^\circ$ (right). Numbers plotted represent a normalized RMS spot size value where 1 is the minimum RMS spot size calculated for the on-axis direction.

Difference between the image shown on the display for a conventional VR HMD lens with linear mapping and here presented foveated lens is illustrated in Figure 4.20. It is clearly observed that using the adapted foveated approach the amount of distortion is increased considerably, emphasizing the central gazed region. Thus, the variation of the magnification for the foveated lens is much greater than for the conventional HMD lens.

Figure 4.20. Difference between the image shown on the display for a conventional HMD lens (that uses linear mapping) (left), and a foveated lens presented here (right).

4.3.5. Tolerance analysis

Current design methodology typically consists of two stages that often interlace. The first stage is the design process and it involves obtaining the best on paper, i.e. nominal design form. The second stage is the assessment of tolerances to the system, and the corresponding analysis and verification that arise [24]. There are multiple tolerance analysis tools and methods in literature [25]–[29]. In the analysis that follows we use Wavefront differential tolerancing (TOR) and Primary aberration-based tolerancing (TOL) from CODE V® [30][31]. A more rigorous analysis, such as Monte Carlo simulations [32] would be desirable if the system is intended for large volume production and commercial use.

The Wavefront Differential Tolerancing (TOR) is a tolerancing procedure based on a wavefront differential ray trace that provides the derivative of the OPD with respect to the tolerances, whereas the Primary Aberration-Based Tolerancing (TOL) calculates changes in various performance measures due to a subset of changes in the constructional parameters of any lens for which first- and third-order parameters are valid. [30][31]

Values of specific tolerances depend on the allowable degrade in the chosen merit function from the on-paper (nominal) system to the real fabricated (as-built) one. Tolerance specification will also depend on what the manufacturer is able to fabricate and the price we are willing to pay for the final system. A rule - looser the tolerances, cheaper the optics is in general applied.

A set of tolerance values chosen for this system is shown in Table 4.4. These values were chosen after conducting various sensitivity analyses and consulting the tolerance limits present in literature as well as in different optical manufacturers' websites. Table 4.4 also presents the default limit values from CODE V® software, as well as values given by optical surface manufacturer Optimax® depending on the desired precision and price of the resulting system [33].

Table 4.4 Applied tolerances for allowable decrease in system performance.

Tolerance	Definition/Units	Default CODE V®		Applied values		Optimax®		
		Min	Max			Commercial	Precision	Limit
Component – Surface Tolerances								
Radius Delta	Delta radius in mm	0.02	-	S2	7.8 (1.7%)	0.20%	0.10%	0.05%
				S3	0.05 (0.18%)			
				S4	0.2 (0.45%)			
				S5	0.05 (0.15%)			
Irregularity 45°	Irregularity in fringes at the fringe λ over a clear aperture	0.5	3	S2..5	3	2 fr	0.5 fr	0.2 fr
Irregularity 0°		0.5	3	S2..5	3	2 fr	0.5 fr	0.2 fr
Component – Compound Tolerances								
Centre thickness	Distance along the axis (mm)	0.02	0.5	S2	0.5	0.15 mm	0.05 mm	0.025 mm
				S3..5	0.05			
Wedge	TIR (mm)	0.002	0.1	S2..5	0.05	0.05 mm	0.01 mm	0.002 mm
Material								
Mean index	Average index error %	0.0001	0.002	Lens1	0.001	0.001	0.0005	0.0002
				Lens2	0.002			
V-number	Abbe number error %	0.002	0.008	Lens1	0.008	0.80%	0.50%	0.30%
				Lens2	0.008			
Optomechanical (separate surfaces)								
Element tilt	° tilt with respect to axis	0.0003	0.005	S2, S4	0.003	0.91°	0.18°	0.036°
				S3, S5	0.001	0.016 rad	0.003 rad	0.00063 rad
Element decentre	decentre from the axis (mm)	0.02	0.5	S2..5	0.01	0.13 mm	0.05 mm	0.001 mm
Optomechanical (group tolerances)								
Barrel tilt	Tilt with respect to axis (rad)	0.0003	0.005	Lens1	0.0005	0.91°	0.18°	0.036°
				Lens2	0.001	0.016 rad	0.003 rad	0.00063 rad
Group decentre	from axis (mm)	0.02	0.5	Lens1,2	0.02	0.13 mm	0.05 mm	0.001 mm

The tolerance analysis was conducted for a similar display of 5.1", a slightly smaller pixel pitch of 44 μm , and a pixel density of 575 ppi. The Nyquist frequency of this display, which can be approximated by

$$f_{Nyquist} = \frac{1}{2op}, \quad (4.13)$$

where op presents a pixel pitch of the display, is 11.36 cyc/mm. In this way we can evaluate our system to work with both available displays. Two sensitivity analyses for two different merit functions were conducted using the tolerance values presented in Table 4.4. It is important to emphasize that the whole tolerance analysis was conducted only for the directly gazing directions. Two graphs showing the degradation in system performance for both merit functions are presented in Figure 4.21.

Figure 4.21. Sensitivity analysis graphs for two chosen metrics: RMS wavefront error on the left, and MTF value at the Nyquist frequency of the chosen display.

Performance drivers for the first merit function, RMS wavefront error, are refractive index and centre thicknesses, while in the case of MTF merit function those drivers are decentre tolerances, such as surface decentre and tilt, although refractive index and centre thickness still play an important role in performance degradation. The only compensator used in these analyses is the defocus position, which does not exceed the value of 0.4 mm. Table 4.5 presents the average decrease in system performance for the two merit functions selected.

Table 4.5 The average decrease in system performance for the two selected merit functions

	RMS wavefront error	MTF at Nyquist frequency
Design	0.64	0.51
Design + tolerance	0.71	0.45

In order to better understand the system's flaws, we evaluate three additional merit functions that represent the Seidel aberrations of our system. Since the distortion is intentionally introduced and will be adjusted through software correction, we are not interested in evaluating that aberration. Additionally, having in mind that this system is intended for a direct human use, field curvature will not be of a great importance, since human eye is capable to adapt to this aberration, and the conflict between accommodation and convergence of these VR systems will hide it. Therefore, we evaluate the 3rd order Spherical

aberration, Coma and Astigmatism. For this evaluation we define three separate Zernike polynomial-based merit functions where we include merely the coefficients that describe the aberrations we are interested in. Using the Fringe Zernike representation from CODE V® [30], given in Table 4.6., the first merit function (spherical aberration) is defined with C9 Zernike coefficient, the second one (Coma) with coefficients C7 and C8, and the third one (Astigmatism) with the coefficients C5 and C6. The resulting graphs are shown in Figure 4.22. As mentioned, distortion and field curvature are excluded due to specific characteristics of this system and its application.

Table 4.6 Fringe Zernike polynomials from CODE V®. Only the first 9 coefficients are shown.

Term	Fringe Zernike polynomials	Aberration type
1	1	Piston (constant)
2	$R \cos \theta$	Distortion – Tilt (x-axis)
3	$R \sin \theta$	Distortion – Tilt (y-axis)
4	$2R^2 - 1$	Defocus – Field curvature
5	$R^2 \cos(2\theta)$	Astigmatism, Primary (axis at 0° or 90°)
6	$R^2 \sin(2\theta)$	Astigmatism, Primary (axis at ±45°)
7	$(3R^3 - 2R)\cos \theta$	Coma, Primary (x-axis)
8	$(3R^3 - 2R)\sin \theta$	Coma, Primary (y-axis)
9	$6R^4 - 6R^2 + 1$	Spherical Aberration, Primary

Figure 4.22. Performance degradation evaluated in terms of specific Seidel aberrations obtained by applying the tolerances given in Table 4.4.

Performance degradation for the latter merit functions is evaluated as the average amount of aberration introduced with the tolerances. This information is presented in Table 4.7. We observe that the main aberration in our system is astigmatism, and therefore it is the aberration that will remain dominant in the manufactured system, as well. From the graphs presented in Figure 4.22 it can be noticed that the amount of astigmatism increases with the field, which is to be expected. Nevertheless, eye rotations of more than 20° are very rare, and in general quite uncomfortable for the user. The amount of astigmatism for gazing directions up to 20° is acceptable, therefore, this aberration, in practice, should not present a problem for the final manufactured system.

Table 4.7 Average decrease in performance for the aberrations presented in Figure 4.22.

	Spherical Aberration	Coma	Astigmatism
Design	0.0054	0.0199	0.2319
Design + tolerance	0.0062	0.0315	0.3265

We also introduce surface deformation tolerances as presented in Table 4.8. [34] We evaluate those separately in order to see how sensitive our system is to the change in the

surface shape. [35] Since RMS surface error tolerance cannot be applied to Zernike polynomial-based merit functions, we evaluate those using only the cosine ripple amplitude and slope surface deformation. The resulting performance degradation graphs are shown in Figure 4.23. The axis labels are excluded for simplicity and are equivalent to those from the Figure 4.21 and Figure 4.22.

Table 4.8 Surface deformation tolerances.

Tolerance	Definition/Units	Default Values CODE V®		Applied value	
		Min	Max		
Random Surface Error	RMS surface error (in waves)	0.01	1	S2	0.065
				S3	0.065
				S4	0.05
				S5	0.05
Cosine Ripple Amplitude surface deformation	Rotationally symmetric cosine sag variation - the amplitude (in waves)	0.01	1	S2	0.075
				S3	0.075
				S4	0.05
				S5	0.055
Cosine Ripple slope surface deformation	Rotationally symmetric cosine sag variation – maximum slope of the variation (in radians)	0.0001	0.5	S2	0.0002
				S3	0.0002
				S4	0.0001
				S5	0.0002

Figure 4.23. Performance degradation for different merit functions produced by the surface deformation tolerances from Table 4.8. The axis labels, excluded for simplicity, are equivalent to those from the Figure 4.21 and Figure 4.22. All horizontal axes are field angle in degrees, while the vertical axis present specific merit function values.

Furthermore, we analyse peak and RMS sag departure from the best fit sphere, as well as peak and RMS slope for the four surfaces in the design. The results are shown in Table 4.9. From these results we see that surface 4 may result problematic for manufacturing. However, since the performance of the system highly exceeds the spatial frequency of the chosen display, potential large manufacturing errors should not present an issue. If the resolution of the nominal system was not so high above the resolution of the display we would need to redesign the surfaces (or at least the fourth surface).

Table 4.9 Peak and RMS sag departure from the best fit sphere and Peak and RMS slope.

Surface number	RMS sag departure from BFS	Peak sag departure from BFS	RMS surface slope	Peak surface slope
1	0.56	-0.67	0.10	0.21
2	7.32	19.90	0.59	1.12
3	4.18	-5.03	0.24	1.15
4	186.17	198.35	2.46	-71.49

The system presented was then manufactured using a direct-cut technique. Since the lens materials used in the design were plastics (PMMA, Polystyrene), the system could also be manufactured by injection moulding. This technique is particularly appropriate for large volume production, for which the price of a single system becomes rather low. However, since we merely wanted to manufacture a prototype, we decided for a direct-cut technique.

4.3.6. Experimental resolution measurements of the manufactured prototype

The manufactured prototype has to be experimentally tested in order to check for any degradation in system performance due to the manufacturing errors. We decided for one of the most common resolution measurements using the 1951 US Air force target, or shortly USAF (Figure 4.24). This target consists of multiple groups of vertical and horizontal bars that represent different spatial resolution. This measurement is normally done by the experienced viewer, with the trained eye, who is able to tell with more certainty whether a certain element is resolvable or not. This is the most common experimental specification for visual systems, due to its simplicity and direct relation with the human eye. However, for the same reason it is highly subjective. For that reason we conduct a more objective Slanted Edge MTF measurement as well, which will be explained later in the text.

Figure 4.24. The 1951 USAF resolution test chart.

For the first measurement a USAF 1951 target including groups from 2 to 7 was used. The spatial frequencies distribution per group and element is shown in Table 4.10.

Table 4.10 Distribution of spatial frequencies in USAF resolving power test target 1951 per group in line pairs per mm. (Source: Edmund Optics)

Element Number	Group Number					
	2	3	4	5	6	7
1	4.00	8.00	16.00	32.00	64.00	128.0
2	4.49	8.98	17.95	36.00	71.8	144.0
3	5.04	10.10	20.16	40.3	80.6	161.0
4	5.66	11.30	22.62	45.3	90.5	181.0
5	6.35	12.70	25.39	50.8	102.0	203.0
6	7.13	14.30	28.50	57.0	114.0	228.0

A white back light in combination with a green glass filter was used to illuminate the target. The filter used has a central wavelength of 525 nm with a FWHM of 80 nm. Using 3D printed holders shown in Figure 4.25, the airspace between the lenses was kept fixed, as well as the position of the target and the illumination set (the back light and the colour filter). The USAF target was placed in a 3D-printed holder, in front of the colour filter that is shown in Figure 4.25 on the left (behind the lens case). On the other side of the same holder a white backlight was positioned. A case connected to the lens holder was printed with a rectangular window, so that the target can be moved laterally and still be seen through the lens.

Figure 4.25. A CAD representation of 3D printed holders used for the measurement.

This case determining the target position places the target at a 24.74 mm distance from the last lens surface, which is different from the nominal design in which the display is placed at 26.94 mm. The difference in display location is a consequence of the mechanical constraints of the final HMD prototype, which were modified after the design process. This change influences the system significantly, positioning the virtual image at 0.8 m instead of 1.5 m as in the nominal design. In order to adequately compare the nominal design and the prototype, we analysed the MTF of the theoretical design including the new display and virtual image locations. Virtual image is considered to be a sphere in this case, to exactly resemble the virtual image obtained in the experimental setup. The values obtained from the software were used for comparison with the experimental results.

For the on-axis position of the pupil, 10% MTF in the nominal design is obtained for the spatial frequency around 90 cycles per mm. This corresponds to 0.000044 rad at a nominal distance of 254 mm, which is around 6.8 times smaller than what the eye is able to resolve (1 arcmin = 0.0003 rad). Therefore, we need to magnify our image in order to see whether an element of interest of USAF target can be resolved. Instead of using a highly magnifying lens, we decided to use a combination of a machine vision camera and an objective lens with long focal length.

The camera chosen for the measurement contains a 1/1.2" Sony® CMOS Exmor sensor (IMX174LQ) with a 1920x1200 resolution and a pixel pitch of 5.86 μm . The detailed specification is given in the Appendix of this thesis. We need to evaluate the Nyquist frequency of the camera on the target location, in order to see whether it is valid for our measurement. We first calculate the real location of the sensor from the objective lens for the object placed at 0.8 m (object in this case is the virtual image obtained with the lens under test). Since we do not have the information regarding the material (and the refractive index) used for the objective lens, we used a thin lens formula, which is enough in this case for rough estimation. We obtain that the sensor is positioned at $S = 82.76$ mm from the objective lens. We can now calculate the ipixel size at the virtual image in arcmin as:

$$ip = \frac{\text{sensor pixel size}}{S} \frac{180}{\pi} 60, \quad (4.14)$$

Obtaining $ip = 0.24$ arcmin. Using now the focal length information of the design under test given in Figure 4.17, we can calculate the size of the opixel at the display. For the on-axis gazing direction the focal length of the system under test is 57.83 mm. The opixel pitch at the centre of the display is then 4.1 μm , and the Nyquist frequency of the camera evaluated at the display according the Eq. (4.13) is 122.1 cyc/mm.

This camera should be used to measure spatial frequencies up to the Nyquist frequency of the camera sensor (122.1 cyc/mm), in order to avoid any aliasing effect disrupting the measurement. Considering the degradation of the system performance due to the low-cost manufacturing technique, as well as high errors obtained in general with 3D printing technique used for holders presented in Figure 4.25, we expect that this is sufficiently high to measure the resolution for all gazing directions. Moreover, this frequency is still high above the Nyquist frequency of the display used in the design, which is 11.36 cyc/mm. In a full HMD prototype that uses this display, spatial frequencies above 11.36 cyc/mm will simply not be resolvable due to the limitations of the display, therefore, the information obtained for those frequencies will clearly not be useful.

Spectral response curve of this sensor is shown in Figure 4.26 (left). Using only the green channel of the sensor in combination with a green glass filter whose spectral response curve is shown in Figure 4.26 (right), we obtained a similar spectrum to the one used for the nominal design (Figure 4.26, down).

Figure 4.26. Spectral response curves of the sensor (left) and green glass filter (right) used in the experimental measurements. Spectral response curve for the display used in the nominal design is presented at the bottom.

The final laboratory setup is shown in Figure 4.27. Using this setup we could capture the images taken through the lens under test and estimate with more certainty the resolution of our system. Using the images captured with the camera we could also evaluate the MTF curves using the Slanted Edge technique which would give us more objective information about the performance of the fabricated lens. [36], [37]

Figure 4.27. Experimental setup for resolution measurements.

Once the experimental set-up is prepared and all the elements are aligned, we evaluate the real position of the virtual image of the prototype. The best focus is obtained around 0.8 to 0.85 m, as expected. The aperture of the objective lens is fixed to approximately 3.5 mm, as in the nominal design.

The images of the target for the on-axis direction are shown in Figure 4.28. The element resolved in this case is the fifth element of the group 5 (for vertical lines) and the fourth element of the same group for the horizontal lines. This corresponds to the spatial frequency of 50.8 cyc/mm. This value is below the estimated limit of 122 cyc/mm, therefore the chosen sensor is good enough for this measurement.

Figure 4.28. Images of the USAF target taken through the lens using a CMOS camera.

Now we compare the results obtained in the laboratory with the ones taken directly from the optical design software evaluating the MTF (Table 4.11). We notice that the experimental results differ significantly from the theoretical ones. As previously mentioned, this can be due to the lens manufacturing errors, errors in 3D printed holders, as well as some alignment errors during the measurements. The USAF technique is also quite subjective, and therefore should be taken with caution.

Table 4.11 Experimental resolution measurements obtained using the USAF 1951 test chart with spatial frequency values given in Table 4.10. Spatial frequencies evaluated at the 10% MTF for the nominal design are given as well for comparison.

Eye rotations (°)	Nominal design		Experimental measurements	
	Radial 0.1 MTF (cyc/mm)	Sagittal 0.1 MTF (cyc/mm)	Radial - bars perpendicular to the meridian plane (cyc/mm)	Sagittal - bars parallel to the meridian plane (cyc/mm)
0 °	90	90	50.8	45.3
5 °	71	90	40.3	57
10 °	76	89	36	57
15 °	79	36	45.3	57
20 °	39	18	16	22.6
25 °	17	12	11.3	14.3
30 °	9	9	10.1	11.3

As previously mentioned, even though this is the most common resolution measurement, it is highly subjective. In order to obtain a more objective measurement of our system performance, we used a Slanted Edge method. The Slanted Edge method [36], [37] of measuring the spatial frequency response (SFR) as an approximation of the MTF is a widely used image quality testing method. It is described in detail in the ISO 12233 standard [38]. The first step involves scanning the image of an edge feature in a direction perpendicular to the edge. An edge profile is then derived from the data as the edge-spread function (ESF). From this ESF, a point-spread function (PSF), which in 2D becomes a line-spread function (LSF) is computed either by a discrete first derivative or by a parametric fit to the data. The discrete Fourier transform of the LSF is then computed, with its modulus recorded. If the input edge feature used is of sufficiently high optical quality, in terms of its edge modulation, then the above measured modulus can be taken as estimating the MTF of the system under test.

In order to obtain a slanted edge target of sufficiently high optical quality, we used a 5° tilted USAF target, centred at the square positioned between the elements 2/1 and 3/1 (Figure 4.28). A region similar to that presented in Figure 4.29 is chosen for evaluation.

Figure 4.29. An image of a slanted edge.

Different open source software for Slanted Edge MTF measurement is available. The one used for this measurement is given in ref. [39]. The MTF curves obtained showed spatial frequencies in cycles per pixel. Using previously explained calculus to compute the opixel size at the target, final MTF curves shown in cycles per millimetre were obtained. In Figure 4.30 both nominal and experimental MTF curves are shown (nominal in dashed lines and experimental in full lines).

Experimentally obtained results for sagittal direction seem to exceed the nominal ones for some gazing directions. We suspect that this could be the consequence of some unaccounted errors in the alignment of the experimental setup, particularly in the alignment of the target. Surface measurement using interferometry could give us an insight of the departure of the fabricated prototype from the nominal design, which could further explain the difference in system behaviour obtained here.

(a)

(b)

Figure 4.30. Experimentally obtained MTF curves for up to 30° gazing direction (solid lines), compared with the MTF results obtained in CODE V® for the nominal design (dashed lines). The results obtained for the radial direction are shown in the upper graph (a), while the ones for the sagittal direction are shown in the bottom graph (b). The black vertical line at 11.36 cyc/mm represents the Nyquist frequency of the display used for the final VR HMD system.

4.4. Conclusions

In this chapter, a foveated imaging design approach for two different applications was demonstrated.

In section 4.2 SMS method with pupil discretization is used to generate different starting points to redesign the aspheric part of a specific endoscopic microscope objective lens. Changing only the seed patch of the surfaces, 3 different surface sets were obtained. This allowed for multiple starting points for further optimization. Since SMS designs are in general close to the final solutions, a short optimization run optimized the systems, where only aspheric surface parameters were used as variables. Two final design solutions were presented. Peak and RMS sag and slope deviation information, showed that one of the newly obtained surface sets should be easier to manufacture than the aspheres from the initial design, which was the goal of this study. Downside of this approach is that surface shape is not controlled during the SMS construction. However, since the surfaces need to be further optimized for the full pupil, this does not represent an issue. Additional constraints related to the surface shape may be included during the post SMS optimization.

In section 4.3 a VR HMD design was presented. Apart from using the foveation approach it includes eye rotation design technique commonly used in spectacle lens design.

Since human eye resolution is much coarser in peripheral vision than close to the gazing direction, the design conditions, such as focal length and image quality across the virtual screen, were adapted to match the optics. In this way the pixels of the virtual screen (ipixels) are not smaller than needed, since the eye will not appreciate it. By adapting the image quality across the virtual screen, we achieved a configuration in which the ipixels are approximately equally resolved by the optics. Additionally, including the eye rotations, a design that prioritises directly gazed ipixels was achieved. This implies that the image quality for the ipixel, when it is directly gazed is the best, while the quality of the peripheral regions falls gradually.

Meeting the performance requirement is not the only goal optical designers need to achieve. The final system needs to be manufacturable, and in general the cheaper the better. In order to make sure the as-built system would still meet the performance goals, tolerance analysis was conducted showing the performance drivers for this system, which predominantly included surface tilt and decentre. Astigmatism turned out to be the leading aberration of this system; yet its amount for the usually gazed parts of the virtual screen is rather small, even after including the tolerances. Since the design exceeds the resolution of the display considered, even looser tolerances than the ones considered in the analysis were used in manufacturing process in order to obtain a low-cost lens.

After manufacturing the prototype with the direct cut technique, the experimental measurements to show the resolution of the as-built system were conducted. Two methods were used to experimentally evaluate the spatial resolution, one being a highly subjective, yet very common, USAF resolution test chart, and another one being a more rigorous Slanted-Edge

MTF method. The obtained results show that the manufactured prototype satisfies the design goals, and shows very good spatial resolution for up to 20° gazing directions, as wanted.

One of the difficulties of foveated imaging systems lies in the strong aspheric surfaces needed in order to produce high distortion that gives the wanted varying magnification across the FOV. Even though the manufacturing techniques have been developing rapidly, strong aspheres present in foveated systems are still difficult to manufacture with high precision. In VR HMD optics where displays in general limit the resolution to low spatial frequencies, large manufacturing errors do not present an issue. This, however, can be a problem in other type of application, such as endomicroscopy.

REFERENCES

- [1] J. Van Der Spiegel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Diario, F. Fantini, P. Bellutti, and G. Soncini, "A foveated retina-like sensor using CCD technology," in *Analog VLSI Implementation of Neural Systems*, 1989, pp. 189–211.
- [2] R. Wodnicki, G. W. Roberts, and M. D. Levine, "Foveated Image Sensor in Standard CMOS Technology," in *Proc. of the IEEE Custom Integrated Circuits Conference*, 1995, pp. 357–360.
- [3] S. Xia, R. Sridhar, P. Scott, C. Bandera, and S. U. N. Y. Buffalo, "CMOS Foveal Image Sensor Chip," in *Proc. of the Eleventh Annual IEEE International ASIC Conference*, 1998, pp. 409–413.
- [4] Y. Suematsu and T. Hayase, "An advanced vision sensor with fovea," in *Proceedings of the IECON '90*, 1990, pp. 581–585.
- [5] Y. Suematu, H. Yamada, and T. Ueda, "A Wide Angle Vision Sensor with Fovea - Design of Distortion Lens and the Simulated Images," in *Proceedings of the IECON '93*, 1993, pp. 1770–1773.
- [6] K. Wakamiya, T. Senga, K. Isagi, N. Yamamura, Y. Ushio, and N. Kita, "A new foveated wide angle lens with high resolving power and without brightness loss in the periphery," in *Proc SPIE*, 2005, no. 6051.
- [7] N. Hagen and T. S. Tkaczyk, "Foveated endoscopic lens," *J. Biomed. Opt.*, vol. 17, no. 2, pp. 21104–6, 2012.
- [8] T. Martinez, D. V. Wick, and S. R. Restaino, "Foveated , wide field-of-view imaging system using a liquid crystal spatial light modulator," *Opt. Express*, vol. 8, no. 10, pp. 555–560, 2001.
- [9] G. Curatu, D. V. Wick, D. M. Payne, T. Martinez, J. Harriman, and J. E. Harvey, "Wide field-of-view imaging system using a liquid crystal spatial light modulator," in *Proc SPIE*, 2005, vol. 5874, pp. 587408-1–7.
- [10] J. P. Rolland, A. Yoshida, L. D. Davis, and J. H. Reif, "High-resolution inset head-mounted display," *Appl. Opt.*, vol. 37, no. 19, pp. 4183–4193, 1998.
- [11] A. Ude, C. Gaskett, and G. Cheng, "Foveated vision systems with two cameras per eye,"

- in *Proc. of the 20016 IEEE Internationa Conference on Robotics and Aut9omation*, 2006, no. May, pp. 3457–3462.
- [12] H. Hua and S. Liu, “Dual-sensor foveated imaging system,” *Appl. Opt.*, vol. 47, no. 3, pp. 317–327, 2008.
 - [13] J. D. Rogers, T. T. Tkaczyk, and M. R. Descour, “Foveated endoscope objective design to combine high resolution with wide field of view,” in *Proc Spie*, 2010, vol. 7558, p. 755808.
 - [14] P. Benítez, J. C. Miñano, M. Nikolic, J. Liu, J. M. Infante, and F. Duerr, “Conditions for perfect focusing multiple point sources with the SMS design method,” in *Proc. SPIE*, 2014, vol. 9191, p. 919102.
 - [15] F. Muñoz, P. Benítez, and J. C. Miñano, “High-order aspherics : The SMS nonimaging design method applied to imaging optics,” in *Proc. SPIE*, 2008, vol. 7100, pp. 1–9.
 - [16] F. Muñoz, “Sistemas ópticos avanzados de gran compatibilidad con aplicaciones en formación de imagen y en iluminación,” Universidad Politécnica de Madrid, 2004.
 - [17] G. W. Forbes, “Shape specification for axially symmetric optical surfaces.,” *Opt. Express*, vol. 15, no. 8, pp. 5218–5226, 2007.
 - [18] G. W. Forbes and C. P. Brophy, “Asphere, O Asphere, how shall we describe thee?,” *Opt. Des. Eng.*, vol. 7100, no. 1, p. 710002, 2008.
 - [19] J. P. Rolland and H. Hua, “Head-Mounted Display Systems,” *Encycl. Opt. Eng.*, pp. 1–14, 2007.
 - [20] P. Benítez and J. C. Miñano, “Immersive compact display glasses,” Internatioanal patent WO 2015/077718 A1, 2015.
 - [21] D. A. Atchison, “Spectacle lens design:,” *Appl. Opt.*, vol. 31, no. 19, pp. 3579–3585, 1992.
 - [22] G. L. Stephens and J. K. Davis, “Spectacle Lens Design, Chapter 51B.” [Online]. Available: <http://www.oculist.net/downaton502/prof/ebook/duanes/pages/v1/v1c051b.html#the>. [Accessed: 12-Jan-2017].
 - [23] D. A. Atchison, “Modern Optical Design Assessment and Spectacle Lenses,” *Opt. Acta Int. J. Opt.*, vol. 32, no. 5, pp. 607–634, 1985.
 - [24] J. D. Nelson, R. N. Youngworth, and M. Aikens, “The Cost of Tolerancing,” in *Proc. SPIE*, 2009, vol. 7433.
 - [25] R. N. Youngworth, “21 st Century optical tolerancing : a look at the past and improvements for the future cost yield,” in *Proc SPIE*, 2006, vol. 6342, pp. 1–15.
 - [26] G. Adams, “Selection of tolerances,” in *Proc SPIE*, 1988, vol. 892, pp. 173–185.
 - [27] W. J. Smith, “Fundamentals of the Optical Tolerance Budget,” in *Proc SPIE*, 1990, vol. 1354, pp. 474–481.
 - [28] S. Magarill, “Optomechanical sensitivity and tolerancing,” in *Proc SPIE*, 1999, vol. 3786, pp. 220–228.
 - [29] W. Smith, “Modern Lens Design.” 2004.
 - [30] “CODE V Reference Manuals.” .
 - [31] D. Hasenauer, “Optical design tolerancing: A key to product cost reduction.” CODE V, Synopsys, 2015.
 - [32] D. Forse, “Statistical tolerancing for optics,” in *Proc SPIE*, 1996, vol. 2775, pp. 18–27.

- [33] ["http://optimaxdev.webhost.io/innovation/optical-manufacturing-tolerance-chart/."](http://optimaxdev.webhost.io/innovation/optical-manufacturing-tolerance-chart/) .
- [34] J. R. Rogers, "Slope Error Tolerances for Optical Surfaces," in *Proc SPIE*, 2007, vol. TD04, p. TD0404.
- [35] U. Fuchs, S. Str, and D.- Jena, "Tolerancing aspheric surfaces in optical design," in *Proc SPIE*, 2015, vol. 9582, pp. 1–7.
- [36] J. K. M. Roland, "A study of slanted-edge MTF stability and repeatability," *SPIE Proc. Vol. 9396*, p. 93960L, 2015.
- [37] P. D. Burns, "Slanted-Edge MTF for Digital Camera and Scanner Analysis," *Proc. IS&T 2000 PICS Conf.*, pp. 135–138, 2000.
- [38] ISO/TC42/WG18, "Resolution and spatial frequency response." International Organization for Standardization (ISO), 2014.
- [39] "Slanted Edge MTF open source software." .

PART 2

FREEFORM OPTICS

Chapter 5

ANALYTICAL SOLUTION OF AN AFOCAL TWO FREEFORM MIRROR DESIGN PROBLEM

5.1. Introduction

Optical systems that are free from spherical aberrations and circular coma are known as aplanatic systems. [1] Aplanatic here refers to full aplanatic, i.e. those aberrations vanish for all orders, and therefore they satisfy Abbe sine condition. The first rotationally symmetric two mirror aplanatic systems were described by Schwarzschild in 1905 [2] and were later developed and analytically deduced for telescopes which required superior aberration-free imaging [3]–[6]. One particular aplanatic design is the afocal case, in which the null spherical aberration implies that the plane input wavefront WF_{in} is coupled to a plane output wavefront WF_{out} stigmatically. The Abbe sine condition in its general form [7] states that for this afocal case the mapping of the rays between the two wavefronts is not arbitrary, rather that the transformation between Cartesian coordinates x - y defined on WF_{in} and Cartesian coordinates x' - y' defined on WF_{out} , given by:

$$\begin{aligned}x' &= a_{00}(x, y) \\ y' &= b_{00}(x, y)\end{aligned}\tag{5.1}$$

must be linear (the sub-indices are used for consistency with equations below).

We will use here the term semi-aplanatic for optical systems in which only one of the functions in Eq. (5.1) is linear. Since the coordinate systems on the wavefronts can be rotated, without loss of generality it can be considered that the semi-aplanatic case corresponds to the case in which a_{00} is a linear function of only x , that is, $x' = m_x x + l_x$ (where m_x and l_x are real numbers), which can be viewed as a one-dimensional Abbe sine condition.

A recent publication by Miñano *et al.* [8] connects the general condition for aplanatism and the SMS 3D design method for freeform optical systems. From that reference it is deduced that a general afocal aplanatic system can be viewed as the limit case of an optical system that couples three flat input wavefronts with linearly independent normal vectors into other three flat output wavefronts with also linearly independent output vectors, when the three input and output normal vectors converge to the normal vectors of $\mathbf{WF}_{in}$ and $\mathbf{WF}_{out}$, respectively. An example of such a system is obtained by designing three freeform surfaces with the SMS 3D method. However, the necessity of three freeform surfaces for the general case does not impede that there may be particular cases in which fewer surfaces can provide an aplanatic design, as the aforementioned two-mirror rotational case, or the well know case of the aplanatic points of the refractive sphere, which uses a single surface. The semi-aplanats can be also viewed as the limit case of an optical system that couples only two flat input wavefronts into other two flat output wavefronts, when the wavefront pairs converge to $\mathbf{WF}_{in}$ and $\mathbf{WF}_{out}$, respectively. An example of semi-aplanatic system is obtained by designing two freeform surfaces with the SMS 3D method. A semi-aplanatic system with 3 freeform surfaces can be obtained in the limit case of SMS 3D method, which is discussed in more detail in Chapter 6 of this thesis.

In this chapter we will first address a more general afocal design problem in first order optics, in which the ray-to-ray mapping between $\mathbf{WF}_{in}$ and $\mathbf{WF}_{out}$ is such that a_{00} is a function of only x , not necessarily linear. Section 5.2 is devoted to deriving the first-order angular magnification of the afocal system for this case, using the étendue conservation theorem of bi-parametric ray bundles in the phase space. In section 5.3 the design problem of two afocal freeform mirrors fulfilling such a ray-to-ray mapping is stated and the resulting partial differential equations are solved in a closed-form. Four specific design examples are given in Section 5.4, two of them presenting, to our knowledge, first examples of aplanatic systems made of only two non-rotational symmetric mirrors. The apparent contradiction with the general rule that three freeform surfaces are needed [8][9] is discussed in section 5.5. In section 5.6 a connection with the SMS 3D design method is briefly considered. Alternative design geometries including two different mirror systems and one refractive case are shortly addressed in sections 5.7 and 5.8.

5.2. Phase space considerations

Let us consider the plane wavefronts, $\mathbf{WF}_{in}$ and $\mathbf{WF}_{out}$, which are in general non-parallel, and two Cartesian coordinate systems, x - y - z and x' - y' - z' , such that $\mathbf{WF}_{in}$ and $\mathbf{WF}_{out}$ coincide with the planes $z=0$ and $z'=0$, respectively and propagate towards positive z and z' values.

Let us consider the 4-dimensional phase space at the $z = 0$ and $z' = 0$ planes as reference surfaces. Let p and q be the optical direction cosines of a ray when crossing the plane $z=0$, and p' and q' those at the plane $z'=0$. The afocal optical system will transform a ray (x, y, p, q)

crossing the plane $z = 0$ onto the ray (x', y', p', q') crossing $z' = 0$, and such transformation can be expressed in power series of p and q as:

$$\begin{aligned} x' &= a(x, y, p, q) = a_{00}(x, y) + a_{10}(x, y)p + a_{01}(x, y)q + \dots \\ y' &= b(x, y, p, q) = b_{00}(x, y) + b_{10}(x, y)p + b_{01}(x, y)q + \dots \\ p' &= c(x, y, p, q) = c_{00}(x, y) + c_{10}(x, y)p + c_{01}(x, y)q + \dots \\ q' &= d(x, y, p, q) = d_{00}(x, y) + d_{10}(x, y)p + d_{01}(x, y)q + \dots \end{aligned} \quad (5.2)$$

The stigmatic condition (i.e., the rays of **WF_{in}** are coupled with those of **WF_{out}**) can be written as $p' = 0$ and $q' = 0$ when $p = 0$ and $q = 0$, which means that $c_{00} = d_{00} = 0$ in Eq. (5.2).

Applying the étendue conservation theorem [10] for bi-parametric bundles as done in ref. [8], we can deduce that the coefficients a_{00} , b_{00} , c_{10} , c_{01} , d_{10} and d_{01} fulfil:

$$\begin{pmatrix} a_{00x} & b_{00x} \\ a_{00y} & b_{00y} \end{pmatrix} \begin{pmatrix} c_{10} & c_{01} \\ d_{10} & d_{01} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (5.3)$$

where the sub-indices x and y denote partial derivation. Both matrices on the left hand side of Eq. (5.3) must therefore be non-singular, in particular:

$$\begin{vmatrix} a_{00x} & b_{00x} \\ a_{00y} & b_{00y} \end{vmatrix} \neq 0 \Rightarrow a_{00x}b_{00y} - a_{00y}b_{00x} \neq 0 \quad (5.4)$$

We are interested in analysing the implications of imposing the condition that a_{00} is a function of only x , i.e. $a_{00}(x, y) = a_{00}(x)$. From Eq. (5.4), we get $b_{00y} \neq 0$. Hence from the Eq. (5.3):

$$\begin{aligned} c_{10}(x, y) &= \frac{1}{\dot{a}_{00}(x)} & \dot{a}_{00}(x)c_{01}(x, y) + \frac{b_{00x}(x, y)}{b_{00y}(x, y)} &= 0 \\ d_{10}(x, y) &= 0 & d_{01}(x, y) &= \frac{1}{b_{00y}(x, y)} \end{aligned} \quad (5.5)$$

where the dot on top denotes derivation of a one-variable function. Eq. (5.2) now becomes:

$$\begin{aligned} x' &= a_{00}(x) + a_{10}(x, y)p + a_{01}(x, y)q + \dots \\ y' &= b_{00}(x, y) + b_{10}(x, y)p + b_{01}(x, y)q + \dots \\ p' &= \frac{1}{\dot{a}_{00}(x)}p + \frac{b_{00x}(x, y)}{\dot{a}_{00}(x)b_{00y}(x, y)}q + \dots \\ q' &= \frac{1}{b_{00y}(x, y)}q + \dots \end{aligned} \quad (5.6)$$

Particularly, we conclude that if a_{00} is a function of only x , $a_{00}(x)$, then q' must be independent on the first order of p (since $d_{10} = 0$), and that when $q = 0$, the ratio of the direction cosines p' and p (which indicates the first-order angular magnification) equals the inverse of the derivative of $a_{00}(x)$. The ray-to-ray mapping between **WF_{in}** and **WF_{out}** is visualized by making $p = q = 0$ in Eq. (5.6), so we obtain that also $p' = q' = 0$ and that:

$$\begin{aligned} x' &= a_{00}(x) \\ y' &= b_{00}(x, y) \end{aligned} \quad (5.7)$$

If we apply these results to the semi-aplanatic case, in which $a_{00} = m_x x + l_x$, we obtain, as expected, that such magnification is the constant $1/m_x$, independent of x . To simplify the nomenclature, in what follows we will call $a_{00}(x) = a(x)$, which is assumed to be known, and $b_{00}(x, y) = Y(x, y)$.

5.3. A particular two-mirror solution

Consider now the particular case defined in which the two coordinate systems have parallel axes as shown in Figure 5.1. We are going to design an afocal two-mirror system such that the ray of WF_{in} that impinge on mirror S_1 at the point P will be reflected towards point Q of mirror S_2 , and reflected there again to become a ray of WF_{out} , such that the x-coordinate of Q will only depend on the x-coordinate of P. Therefore, in the coordinate system x-y-z we can write the equations of the two mirror surfaces in parametric form as $P = (x, y, z(x, y))$ and $Q = (a(x), Y(x, y), Z(x, y))$, where functions z , Y and Z define the mirror surfaces and x and y act as parameters for both surfaces.

Figure 5.1. Two mirror afocal system that transforms one plane wavefront WF_{in} to another plane wavefront WF_{out} .

First, the optical path length between WF_{in} and WF_{out} must be constant, and thus:

$$z + L(x, z, Z, Y) + D - Z = C \quad (5.8)$$

where D is the constant distance between WF_{in} and WF_{out} in Figure 5.1 and

$$L(x, z, Z, Y) = \sqrt{(a(x) - x)^2 + (Y - y)^2 + (Z - z)^2} \quad (5.9)$$

Let us now apply a reflection law on the first surface. It can be computed in the standard vector form or, alternatively, as the derivative of the optical path length with respect to two independent parameters. The latter, with Herzberger notation [13], is written as:

$$\mathbf{r}'_u \cdot \hat{s}' n' - \mathbf{r}_u \cdot \hat{s} n = E_u \quad (5.10)$$

where vectors $\hat{s}$ are unitary, the primes denoting output space, n, n' are the refractive indices and r, r' are the position vectors, E is the optical path length between the position vectors, and u is the parameter. Thus, for the reflection on the first mirror, taking $\mathbf{r}' = (x, y, z)$, $\mathbf{r} = (x, y, 0)$, $E = z$ and parameter x first we have:

$$(1, 0, z_x) \cdot (a(x) - x, Y - y, Z - z) = z_x L(x, z, Z, Y) \quad (5.11)$$

Eliminating L from Eq. (5.8) and Eq. (5.11), the partial derivative of z with respect to x becomes:

$$z_x = \frac{1}{C - D} (a(x) - x) \quad (5.12)$$

The analog to Eq. (5.11) for y as a parameter is given by:

$$(0, 1, z_y) \cdot (a(x) - x, Y - y, Z - z) = z_y L(x, z, Z, Y) \quad (5.13)$$

Eliminating L again with Eq. (5.8), the partial derivative of z with respect to y is:

$$z_y = \frac{1}{C - D} (Y - y) \quad (5.14)$$

Eqs. (5.8), (5.12) and (5.14) form the system to solve with unknown functions z, Y and Z . Integrating Eq. (5.12):

$$z = \frac{1}{C - D} \left(A(x) - \frac{x^2}{2} \right) + z_0(y) \quad (5.15)$$

where $A(x) = \int_0^x a(t) dt$ and $z_0(y)$ is an arbitrary function. Note that $z(0, y) = z_0(y)$, so this function describes the profile of the cross section of mirror S_1 with the plane $x=0$, and can be seen as a boundary condition for the PDE system.

Substituting Eq. (5.15) in Eq. (5.14):

$$Y = (C - D) \dot{z}_0(y) + y \equiv b(y) \quad (5.16)$$

So $Y(x, y)$ is a function of y only, which we have called $b(y)$. Finally, from Eq. (5.8)-(5.9) we get:

$$Z = z + \frac{1}{2(C - D)} \left((a(x) - x)^2 + (b(y) - y)^2 - (C - D)^2 \right) \quad (5.17)$$

and substituting Eq. (5.16):

$$Z = \frac{A(x)}{C - D} + \frac{a(x)(a(x) - 2x)}{2(C - D)} + z_0(y) + \frac{C - D}{2} (\dot{z}_0^2(y) - 1) \quad (5.18)$$

Equations (5.15), (5.16) and (5.18) describe the two mirror analytical solution to the problem in which the ray mapping between WF_{in} and WF_{out} fulfils that $x' = a(x)$, independently

of y , with the only condition that the x - y axes and x' - y' axes are parallel. In the process we obtained that this solution also satisfies that $y' = b(y)$, independently of x (while in the general case of Eq. (5.7), y' may depend on x too). Therefore, this afocal two freeform mirror system has a decoupled ray-to-ray mapping equations, one related to x -variables and the other y -variables, with full symmetry in between them (although our initial condition imposed was asymmetric).

This resulting two-mirror system of equations (5.15), (5.16) and (5.18) has also relevant non-imaging properties. When an irradiance distribution is defined on the wavefronts, the energy transfer is controlled by the Jacobian matrix of the mapping function Eq. (5.1), which for these two mirrors is a diagonal matrix with $a'(x)$ and $b'(y)$ as diagonal elements. This non-imaging problem was in fact treated in [14] for irradiance beam shaping when the prescription has rectangular symmetry (although restricted to the case in which the diagonal elements of the Jacobian matrix have equal sign). Notice that the energy transfer problem leads to a second-order Monge-Ampère PDE, while the mapping-constrained problem stated here is of first order only.

For the semi-aplanatic case $a(x) = m_x x + l_x$, we obtain that Eq. (5.15), (5.16) and (5.18) become:

$$z = \frac{1}{C-D} \left(l_x x + (m_x - 1) \frac{x^2}{2} \right) + z_0(y) \quad (5.19)$$

$$Y = (C-D) z_0'(y) + y \quad (5.20)$$

$$Z = \frac{m_x l_x}{C-D} x + \frac{m_x x^2}{2(C-D)} (m_x - 1) + z_0(y) + \frac{l_x^2}{2(C-D)} + \frac{C-D}{2} (z_0''(y) - 1) \quad (5.21)$$

The parabolic dependence of Eq. (5.19) with respect to x implies that in this semi-aplanatic case S_1 is plane-symmetric with respect to the $x=\text{constant}$ plane passing through the parabola vertex (coinciding with $x=0$ when $l_x=0$), and it is easy to see that S_2 has the same symmetry.

It can be noticed that in this semi-aplanatic case $z(0, y) = z_0(y)$, so this function describes the profile of the cross section of S_1 with the plane $x=0$, and can be seen as the boundary condition for the PDE system.

5.4. Design examples

Figure 5.2 shows as a first example an afocal freeform semi-aplanatic system that performs the ray-to-ray mapping $X = 2x/3$ and $Y = 0.35y^3 + 2.2y - 2.4$. From Eq. (5.16) we get the function $z_0(y)$, which is our contour condition, and it must be as follows:

$$z_0(y) = \frac{1}{C-D} \left(\frac{0.35y^4}{4} + 1.1y^2 - 2.4y \right) + K \quad (5.22)$$

Surfaces S_1 and S_2 are given by Eq. (5.19) and (5.20). The specific design in Figure 5.2 has the parameter $C - D = 3$ and the arbitrary integration constant $K = -2$.

Figure 5.2. Semi-aplanatic afocal two-freeform mirror system obtained analytically presented in 3D (a) and as a 2D cross-section (b).

The following example demonstrates ray compression on the output wavefront, as can be observed in Figure 5.3. The ray-to-ray mapping for this design is: $X = 2x/3$ and $Y = -0.46y^3 + 2.2y - 4.2$, the parameter $C - D = 3$, and the arbitrary integration constant $K = -5$. Subsequently:

$$z_0(y) = \frac{1}{C-D} \left(-\frac{0.46y^4}{4} + 0.6y^2 - 4.2y \right) + K \quad (5.23)$$

Figure 5.3. Semi-aplanatic afocal two freeform mirror system showing the effects of ray compression at the edge of the pupil at WF_{out} presented in 3D (a) and as a 2D cross-section (b).

In the previous example, the boundary condition $z_0(y)$ given in Eq. (5.23) is a forth order polynomial, which leads through Eq. (5.16) to a cubic mapping of the y coordinates. If we consider next a parabolic boundary condition $z_0(y)$, Eq. (5.16) will imply a linear mapping of the

y coordinates. Since we found in section 5.3 that in this afocal two-mirror $y' = b(y)$, independently of x , hence the linear mapping is fulfilled not only along the cross section $x = 0$, but for all x and y values. As a consequence, the semi-aplanatic system with parabolic boundary condition is a full-aplanatic system.

Taking $y' = m_y y + l_y$ (where m_y and l_y are real numbers), then:

$$z_0(y) = \frac{1}{C-D} \left((m_y - 1)y + l_y \right) \Rightarrow z_0(y) = \frac{1}{C-D} \left((m_y - 1) \frac{y^2}{2} + l_y y \right) + K \quad (5.24)$$

When $m_x = m_y$, it is easy to see that the solutions are two confocal paraboloids of revolution with axis parallel to the z axis. Figure 5.4 presents such a system for $m_x = m_y = 1.5$ and parameters $C - D = 10$, $K = 12$.

Figure 5.4. Two freeform mirror aplanatic system for the same magnification $m_x = m_y = 1.5$ that represents two confocal paraboloids represented in 3D (a); and as a 2D cross section (b).

In the general case, when m_x and m_y are different, as was already discussed in [8] and in the design of afocal étendue-squeeze microlens arrays in [16], and the resulting mirrors are parabolic-type quadrics without rotational symmetry. From the parabolic dependence of $z_0(y)$ in Eq. (5.24) with respect to y it is easy to deduce that in this aplanatic case S_1 and S_2 are not only plane-symmetric with respect to a plane $x = \text{constant}$ but also with respect to a plane $y = \text{constant}$ (coinciding with $y = 0$ when $l_y = 0$). Plane-symmetric two-mirror systems were already studied in [17], restricted to off-axis rotationally symmetric conics with sharp imagery through second order in both spatial and angular coordinates.

Figure 5.5 shows a specific example of such an afocal aplanatic two mirror design for parameters $C - D = 10$, $K = 3$, $l_x = 0$, $l_y = 0$ and different magnifications $m_x = -2$ and $m_y = 2/3$. The resulting mirrors are hyperbolic paraboloids with axes parallel to the z -axis.

Figure 5.5. Analytically obtained two freeform mirror aplanatic system in 3D (a); and two 2D cross sections with planes $y = 0$ (b, up) and $x = 0$ (b, down).

5.5. Discussion

The third and fourth selected designs shown in section 5.4 (Figure 5.4 and Figure 5.5) are to our knowledge the first examples of aplanatic systems made of only two non-rotationally symmetric mirrors. This seems to contradict the claim that in general at least three freeform optical surfaces are necessary to design an aplanatic system which was found in reference [8]. However, the necessity of three freeform surfaces for the general case does not impede that there may be particular cases in which fewer surfaces can provide an aplanatic design. In fact, apart from the trivial example of the flat mirror, two other particular cases are already known: the two-mirror rotational case mentioned in the introduction, and the aplanatic points of the refractive sphere (which uses a single surface). [1]

Although in general forcing that an afocal system has a mapping such that x' is a function of only x does not impose any condition on y' , when the x - y axes are parallel to the x' - y' axes and two mirrors are used, y' results to be a function of only y . As a consequence, the two-mirror semi-aplanatic system was automatically aplanatic. If there are other semi-aplanatic configurations in which this also occurs it is an open problem that deserves further research.

5.6. Connection with the SMS3D method

The design problem described so far can be obtained as a limit case of an SMS design problem. The SMS 3D method can be used to design two mirrors that couple stigmatically not only two plane wavefronts WF_{in} and WF_{out} , but also two additional plane wavefronts WF_{in}^* and WF_{out}^* obtained by rotating WF_{in} an angle α around the y -axis and WF_{out} an angle $\beta = \text{asin}(m_x^{-1} \sin \alpha)$ also around the y -axis, respectively. When $\alpha \rightarrow 0$, the SMS design converges to

the semi-aplanatic case presented in Eq. (5.19) to (5.21). The more general case of a non-linear function $x' = a(x)$ given by Eq. (5.16) to (5.18) can be obtained as a limit case of the SMS when $\alpha \rightarrow 0$, just by prescribing a cylindrical wavefront WF_{out}^* with direction parallel to the y -axis such that its rays formed an angle $\beta(x) = \arcsin(\dot{a}^{-1}(x) \sin \alpha)$ with the z -axis.

The resulting system of partial differential equations described herein have the function $z_0(y)$ as boundary condition, so this function describes the profile of the cross section of mirror S_1 with the plane $x = 0$, which is equivalent to the seed rib of the SMS 3D design method. [12] The SMS 3D skinning process in this limit case is avoided (since the area of the skinning surfaces tends to zero).

5.7. Analytical calculations for different mirror system geometries

Let us now consider two different system geometries, with parallel and perpendicular WF_{in} and WF_{out} . Let us first evaluate the system shown in Figure 5.6, where the WF_{in} and WF_{out} are parallel, but the direction of the second wavefront is opposite to the previously discussed case in section 5.3.

Figure 5.6. An afocal two-freeform mirror system with parallel input (WF_{in}) and output (WF_{out}) wavefronts.

The optical path length between WF_{in} and WF_{out} for this case (Figure 5.6) is:

$$z + L(x, z, Z, Y) + Z = C \quad (5.25)$$

Equations (5.9)-(5.11) and (5.13) stay unchanged, since we are still considering mirror surfaces, therefore the reflection law still applies. However since the optical path length is different, partial derivatives of z with respect to x and y are now:

$$z_x = \frac{a(x) - x}{C - 2Z} \quad z_y = \frac{Y - y}{C - 2Z} \quad (5.26)$$

From Equations (5.25) and (5.9) we obtain:

$$4zZ + C^2 - 2Cz - 2CZ = (a(x) - x)^2 + (Y - y)^2 \quad (5.27)$$

Expressing from Equation (5.26) Z and Y as functions of z_x and z_y we get:

$$Z = \frac{1}{2} \left(C - \frac{a(x) - x}{z_x} \right) \quad (5.28)$$

$$Y = \frac{z_y}{z_x} (a(x) - x) + y \quad (5.29)$$

Changing Equations (5.28) and (5.29) into (5.27) we derive the partial differential equation of the z coordinate of $\mathbf{WF}_{in}$:

$$z_x^2 + z_y^2 + \frac{2z}{a(x) - x} z_x + \frac{C}{a(x) - x} z_x = 0 \quad (5.30)$$

Mirror surfaces S_1 and S_2 are completely defined with Equations (5.28)-(5.30). Notice that in this case if x -coordinate of the output wavefront $\mathbf{WF}_{out}$ depends only on the x -coordinate of the input wavefront $\mathbf{WF}_{in}$ ($X = a(x)$), y -coordinate of the output wavefront $\mathbf{WF}_{out}$ (Y) is not necessarily only a function of y -coordinate of the input wavefront $\mathbf{WF}_{in}$ (Equation (5.28)), unlike in system geometry discussed in section 5.3.

In a similar way we can consider a system presented in Figure 5.7, where the $\mathbf{WF}_{in}$ and $\mathbf{WF}_{out}$ are perpendicular.

Figure 5.7. An afocal two freeform mirror system with perpendicular input ($\mathbf{WF}_{in}$) and output ($\mathbf{WF}_{out}$) wavefronts.

The optical path length between the two wavefronts for this geometry is:

$$z + L(x, z, Z, Y) - Y = C \quad (5.31)$$

Next, Eq. (5.26) of the partial derivatives of z with respect to x and y for this geometry becomes:

$$z_x = \frac{a(x) - x}{C + Y - Z} \quad z_{2y} = \frac{Y - y}{C + Y - Z} \quad (5.32)$$

Analog to the previous derivation, for the system in Figure 5.7, from Equation (5.31) and (5.9) we get:

$$(C + Y - z)^2 = (a(x) - x)^2 + (Y - y)^2 + (Z - z)^2 \quad (5.33)$$

Equations (5.33) and (5.32) form a system to solve with the unknown functions z , Y and Z . Expressing from Equation (5.28) Z and Y as functions of z_x and z_y :

$$Z = C + \frac{a(x) - x}{z_x} (z_y - 1) + y \quad (5.34)$$

$$Y = \frac{z_y}{z_x} (a(x) - x) + y \quad (5.35)$$

Changing Equation (5.33) into (5.28) we calculate z coordinate of WF_{in} :

$$z_x^2 + z_y^2 + 2 \frac{z - y - C}{a(x) - x} z_x - 2z_y + 1 = 0 \quad (5.36)$$

Equations (5.34)-(5.36) define mirror surfaces S_1 and S_2 . However, same as for the system in Figure 5.6, in the case of perpendicular input and output wavefronts considered here, Y does not necessarily depend only on y , even though X depends only on x . Therefore, the only mirror configuration for which when $X=a(x)$, mapping is always decoupled, is the one treated in section 5.3 and shown in Figure 5.1.

5.8. An afocal refractive system with 2 freeform surfaces

Let us now consider a system presented in Figure 5.8, with two refractive surfaces in air ($n_1 = 1$, $n_2 \neq n_1$). We are interested in solving the PDEs for this system in order to see if more general conclusions regarding decoupled ray-to-ray mapping can be drawn.

Figure 5.8. An afocal refractive system with 2 freeform surfaces.

Optical path length for the afocal refractive system presented in Figure 5.8, when $n_1 = 1$ is:

$$z + n_2 L(x, z, Z, Y) + D - Z = C \quad (5.37)$$

Since we are no longer dealing with mirror surfaces, we have to apply law of refraction in vector form instead of the law of reflection as in sections 5.3 and 5.7. It is equivalent to the derivative of optical path length presented in Equation (5.10) with refractive index $n' \neq 1$ (in Figure 5.8 we use notation of n_1 and n_2 , where $n_1 = 1$ and $n_2 \neq 1$). Analogous to Equations (5.11) and (5.13) we obtain:

$$(1, 0, z_x) \cdot (a(x) - x, Y - y, Z - z) = \frac{1}{n_2} z_x L(x, z, Z, Y) \quad (5.38)$$

$$(0, 1, z_y) \cdot (a(x) - x, Y - y, Z - z) = \frac{1}{n_2} z_y L(x, z, Z, Y) \quad (5.39)$$

Eliminating L from Equations (5.37)-(5.39) and using the Equation (5.9) we obtain the following:

$$(a(x) - x)^2 + (Y - y)^2 + (Z - z)^2 = \frac{1}{n_2^2} (C - D + Z - z)^2 \quad (5.40)$$

$$a(x) - x + z_x (Z - z) = \frac{1}{n_2^2} z_x (C - D + Z - z) \quad (5.41)$$

$$Y - y + z_y (Z - z) = \frac{1}{n_2^2} z_y (C - D + Z - z) \quad (5.42)$$

Equations (5.40)-(5.42) form a system to solve with unknown functions z , Y and Z . The final partial differential equation to solve is:

$$z_y^2 + \left(1 + \frac{C - D}{(a(x) - x)^2} \left(\frac{C - D}{n_2^2 - 1} - 1 \right) \right) z_x^2 + \frac{1}{a(x) - x} \left(1 - \frac{2n_2^2}{n_2^2 - 1} (C - D) \right) z_x + \frac{n_2^4}{n_2^2 - 1} = 0 \quad (5.43)$$

Combining Equations (5.41) and (5.42) we can express Y in the following form:

$$Y = \frac{z_y}{z_x} (a(x) - x) + y \quad (5.44)$$

Last equation indicates that in the considered refractive afocal system, when X is a function of only x , Y does not have to be a function of only y , although such cases may exist. Therefore, in general the ray-to-ray mapping for this configuration is not necessarily decoupled.

5.9. Conclusions

An analytical solution of partial differential equations for an afocal two freeform mirror system was presented. It was proven that for a single imposed condition that coordinate systems are coincident this system has a decoupled ray-to-ray mapping equations, one relating x -variables and the other y -variables, with full symmetry between them. Furthermore, for a semi-aplanatic case, where Abbe sine condition is fulfilled in x -direction, choosing a specific contour condition a full freeform two-mirror aplanat can be achieved. This is the example of aplanatic system without rotational symmetry that uses fewer surfaces than the general aplanat in three dimensions, which requires three surfaces.

Presented problem can also be seen as an SMS design problem with $\alpha = \alpha(x)$ when α tends to zero. Aplanatic system is obtained choosing an appropriate contour condition which is equivalent to the seed rib of the SMS 3D design method. Optical path length chosen for solving a PDE system is equivalent to the average of the two optical path lengths of the SMS 3D and the skinning process is avoided.

Alternative system geometries have been investigated as well. We discussed an afocal two-freeform mirror system with parallel input and output wavefronts, but with opposite directions. Second alternative two freeform mirror geometry studied is the one in which output wavefront is perpendicular to the input one. Unlike in the case where the wavefronts coincide, in these two geometries, ray-to-ray mapping equations are not always decoupled. This means that even though x -coordinate of WF_{out} (X) depends purely on x -coordinate of WF_{in} , y -coordinate of WF_{out} (Y) can depend on both x and y variables. Resulting systems of partial differential equations for both geometries are presented, but we leave their resolution to future research.

With the purpose of finding out whether more general conclusions can be drawn, we considered a refractive case, where the input and the output wavefronts coincide (same as in the first mirror system geometry considered). In this case, as well as in alternative mirror system geometries we found that ray-to-ray mapping is not necessarily decoupled, and in these cases it is expected that Y is a function of both x and y when X is a function of only x .

REFERENCES

- [1] M. Born and E. Wolf, Principles of Optics, 7th (expanded) edition, vol. 32, no. 5. 2000.
- [2] K. Schwarzschild, "Untersuchungen zur geometrischen optik I-III," Abh. Konigl. Ges. Wis. Gottingen Mathphys., 1905.
- [3] A. K. Head, "The Two-Mirror Aplanat," *Proc. Phys. Soc. London Sec.*, vol. B 70, no. 10, pp.

945–949, 1957.

- [4] J. J. M. Braat and P. F. Greve, “Aplanatic optical system containing two aspheric surfaces,” *Appl. Opt.*, vol. 18, no. 13, pp. 2187–2191, 1979.
- [5] D. Lynden-Bell, “Exact optics : a unification of optical telescope design,” *Mon. Not. R. Astron. Soc.*, vol. 334, no. 4, pp. 787–796, 2002.
- [6] R. V Willstrop and D. Lynden-Bell, “Exact optics – II . Exploration of designs on- and off-axis,” *Mon. Not. R. Astron. Soc.*, vol. 342, no. 1, pp. 33–49, 2003.
- [7] W. T. Welford, “The most general isoplanatism theorem,” *Opt. Com.*, vol. 3, no. 1, 1971.
- [8] J. C. Miñano, P. Benítez, and B. Narasimhan, “Freeform aplanatic systems as a limiting case of SMS,” *Opt. Express*, vol. 24, no. 12, pp. 13173–13178, 2016.
- [9] B. Narasimhan *et al.*, “Design of freeform aplanatic systems,” in *Proc SPIE*, 2015, vol. 9629, p. 96290C.
- [10] J. C. Miñano, R. Mohedano, and P. Benítez, *Nonimaging Optics*. 2015.
- [11] J. Chaves, *Introduction to Nonimaging Optics*. CRC Press, Taylor & Francis Group, 2008.
- [12] P. Benítez *et al.*, “Simultaneous multiple surface optical design method in three dimensions,” *Opt. Eng.*, vol. 43, no. 7, p. 1489, 2004.
- [13] M. Herzberger, *Modern Geometrical Optics*. Interscience, 1958.
- [14] V. Oliker, “Optical design of freeform two-mirror beam-shaping systems,” *J. Opt. Soc. Am.*, vol. 24, no. 12, pp. 3741–3752, 2007.
- [15] P. Benítez *et al.*, “Simultaneous multiple surface optical design method in three dimensions,” *Opt. En*, vol. 43, no. 7, pp. 1489–1502, 2004.
- [16] R. J. Koschel, *Illumination Engineering: Design with Nonimaging Optics*.
- [17] B. D. Stone and G. W. Forbes, “Second-order design methods for definitive studies of plane-symmetric, two-mirror systems,” *J. Opt. Soc. Am.*, vol. 11, no. 12, pp. 3292–3307, 1994.

Chapter 6

SMS 3D METHOD FOR DESIGNING 3 FREEFORM SURFACES

6.1. Introduction to SMS 3D method

Designing freeform optical systems involves a specific difficulty related to the added freedom in the surface shape, which results in extremely high number of variables to be used in the optimization process. In optimization procedures such as damped least squares method, the key is in using the appropriate starting system, as explained in Chapter 1, section 1.8. In traditional optical design with rotational symmetry, there are many patents and catalogues where a starting system can be found. However, in the case of freeform optics, being this quite a new field, not many systems are available. A direct design method, such as SMS method can provide a needed starting point. It was shown that using SMS method as a starting point leads to better final designs and the convergence of the merit function is in general faster [1].

SMS method was developed as a non-imaging design technique in the 1990's and numerous non-imaging optical systems were designed with it [2]–[6]. In general, the method involves the simultaneous calculation of N optical surfaces (refractive or reflective) using N pairs of wavefronts. This explanation of the method is correct when the surfaces are close to the aperture stop of the system, otherwise a greater number of wavefronts can be used to design the same number of surfaces (N), as explained in Chapter 2 [7]–[9]. For each pair of wavefronts one condition is imposed. This condition states that the input and the output wavefront are perfectly coupled. Both 2D and 3D versions of the method exist, where the first one is used to design aspheric surfaces using only rays in a plane; and it is in some cases called SMS NM algorithm (where N represents a number of surfaces to be designed, or the number of meridian bundles M used) [10]. SMS 3D is used to design freeform surfaces, and it uses whole ray bundles (both meridian and skew rays). SMS 3D algorithm for designing two freeform surfaces was shortly introduced in Chapter 1, section 1.3, and can also be found in more detail in ref. [11].

With rising interest in freeform systems, and specifically in direct methods that give starting points for further optimization, the need to extend SMS 3D method to design a greater number of surfaces was obvious. The first step towards a generalization of SMS 3D method, analogue to the one presented in Chapter 5 of ref. [12] for SMS 2D, is a design of 3 freeform surfaces. In Figure 6.1 a schematic demonstration of such a problem is presented. Even though these surfaces and wavefronts could in general be freely chosen, here we will limit the problem to the geometry in which the surfaces are plane-symmetric with respect to two orthogonal planes ($x = 0$ and $y = 0$), and the input wavefronts are spherical and centered on three points on the x -axis, and the output wavefronts are also spherical and centered on other three points on a line parallel to the x -axis.

Rotationally symmetric two-surface SMS 2D designs are known to converge to aplanatic systems when the design wavefronts converge one to the other [14]. On the other hand, two surface SMS 3D designs have been proven to be semi-aplanatic at that limit [15]. We will show that at the limit in which the three wavefronts coincide the three-surface design example presented here converges to a second-order semi-aplanatic system.

Figure 6.1 A schematic demonstration of SMS 3D optical system coupling three pairs of freeform wavefronts

6.2. Design procedure

The SMS design procedure for 3 freeform surfaces is based on SMS 3D algorithm for two freeform surfaces given in ref. [13] and [14]. The refractive indices of the four media that separate the surfaces are assumed to be known, and the object points A, B and C and its corresponding image points A', B' and C' are also given (on lines parallel to the x -axis). The design procedure presented here is divided in two phases: (1) calculation of the seed patches

illustrated in Figure 6.2 and (2) the calculation of the rest of the SMS surfaces shown in Figure 6.3.

Figure 6.2 An illustration of Phase I of the SMS 3D design procedure for three on-axis freeform surfaces.

The first step in Phase I consists of choosing a continuous curve in a plane $x = 0$ (along with its normal vectors, which due to symmetry all belong to the plane $x = 0$). This curve is part of the middle surface denoted here as surface B . An additional smooth surface (symmetric with respect to the plane $x = 0$) is chosen as bottom surface C as well. However, even though this surface is chosen freely, its boundary is not. It is calculated in step 2 (Figure 6.2 (b)) by propagating a spherical wavefront from the point C through the surface (called here surface C)

and computing the boundary curve for which the rays coming from the point **C** all pass through the curve of the surface *B*. The boundary curve in the $x < 0$ part of the surface is obtained as well by the symmetry of the problem. In the step 3 of the Phase I (Figure 6.2 (c)), the same rays previously used to calculate the boundary of surface *C* are now refracted at the curve of the surface *B* using Snell's law, and the generalized Cartesian oval that focuses those refracted rays to a point **C'** on the image plane is calculated. This calculated curve belongs to the third top surface, denoted here as a surface *A*. In step 4 (Figure 6.2 (d)), a spherical wavefront coming from the point **B** is propagated and refracted first at the surface *C* and then at the central curve of the surface *B*. In the same way as in step 3, a generalized Cartesian oval that focuses the rays to a point **B'** is calculated. Thus, the second curve belonging to the surface *A* and contained in the plane $x = 0$ is obtained. The optical path length chosen for the wavefront pair **B-B'** and that of the wavefront pair **C-C'** must be conveniently chosen to avoid that the two curves obtained in steps 3 and 4 have a noticeable step, since we could encounter problems when interpolating them into a surface. The interpolation using NURBS is done as in ref. [10] to form the first half of the seed patch of the surface *A*. The other half is obtained via symmetry. The final result of the Phase I includes, therefore, two surfaces symmetric with respect to the $x = 0$ plane, and a middle curve belonging to this plane (Figure 6.2 (e)).

Using the seed patches obtained in Phase I, in Phase II we further grow the surfaces patch-by-patch in a positive x -direction (the growing of the surfaces can be done in the negative x -direction as well, in the same manner, although in this symmetric case it is not needed). We start by propagating a fan of rays coming from the point **B** and reaching the surface *C* at its point on the plane $x = \Delta x$ (where Δx is a constant of a small magnitude). After the fan is refracted at the seed patch of the surface *C*, a generalized Cartesian oval belonging to the surface *B* is calculated using the constant optical path length condition, focusing the fan on **B'** after refracting on the seed patch of surface *A*. The known curve of surface *B* in plane $x = 0$ and the newly calculated one are interpolated again to form the surface patch presented in green colour in Figure 6.3 (b). Now we can trace a fan of rays from point **A** through the known seed patch of surface *C* and through the newly calculated patch of the surface *B*, and calculate the next patch of the surface *A* (in positive x -direction) as shown in Figure 6.3 (c). In a similar way, tracing from point **C** through the known portions of surfaces *A* and *B*, a new patch of the surface *C* is calculated (Figure 6.3 (d)). Repeating this calculation, the three surfaces are grown simultaneously patch-by-patch, and the final result is shown in Figure 6.3 (e). The negative parts of the surfaces are obtained via symmetry.

Figure 6.3 An illustration of the Phase II of the SMS 3D design procedure for three freeform surfaces when the design points are distributed in a line.

6.3. Design example

An example of SMS 3D with three on-axis freeform surfaces is presented here. The system shown is a monochromatic design comprising two lenses, where the first surface is a prescribed plane, and the remaining three surfaces are SMS surfaces. Our system has a 10 mm focal length, and the aperture $f/2.6$. The object is placed at infinity, therefore the input wavefronts are planes. The design angles defined in a plane $x = 0$ are: -10° , 0° and 10° . The full system specification is presented in Table 6.1.

Table 6.1. System specifications for SMS3D design with 3 freeform surfaces

Back focal length	10 mm
f/#	f/2.6
Wavelength	580 nm
Index of refraction	1.49 (PMMA)
Object	Infinity
Design angles (defined in a plane x=0)	-10°, 0°, 10°

The output wavefronts are spheres with centres at the SMS image points defined as:

$$Y'_k = f \tan \theta_k \quad (6.1)$$

where f is the focal length and θ_k is the angle that the input ray forms with the optical axis. Index k refers to a particular wavefront pair, or image point.

The object is at infinity and the first surface is a prescribed plane. The thicknesses for both lenses are 1 mm, the airspace between them is 1.5 mm and the back focal length $BFL = 10$ mm. A resulting plane symmetric on-axis system with 3 freeform SMS surfaces is shown in Figure 6.4. Contour plots of the surfaces are given in Figure 6.5.

Figure 6.4 A CAD representation of the resulting SMS 3D design of 3 on-axis freeform surfaces.

Figure 6.5 A 3D representation of sag (top) and contour plots (bottom) of the three SMS surfaces. The range of the vertical axis presenting sag is the same for all surfaces.

We further analyse the RMS spot diameter of this system in order to see whether the RMS for the design fields is truly zero. The resulting RMS distribution plot is shown in Figure 6.6, and we clearly distinguish 3 zeros for 3 design fields. The RMS spot size at the SMS image points is of the order of magnitude of 10^{-9} mm.

Figure 6.6 RMS plot for the on-axis SMS3D design of 3 freeform surfaces. A 3D representation of the plot (left) and a top view of the plot (right) is shown. **V** stands for vertical y-direction, while **H** stands for horizontal x-direction

RMS spot diameter distribution curves for sagittal field angles in plane $y=0$ and tangential field angles in plane $x=0$ are shown in Figure 6.7. Since this is an on-axis design the second curve (tangential field angles) is analogue to the one that would be obtained if SMS 2D design algorithm for 3 aspheric surfaces was used (Figure 6.7, right) and has three clear zeros for the design angles.

Figure 6.7 RMS distribution curves obtained for field angles in x-direction, ($y = 0$), i.e. sagittal fields (right) and for field angles in y-direction ($x = 0$), i.e. tangential fields (left). Notice that the right graph is analogue to the RMS_{2D} plot that would be obtained in SMS 2D designs of aspheric surfaces with the same initial conditions, only in this case the RMS for the whole pupil is considered, not just the rays in a plane.

For certain applications we may want to optimize this system in order to balance the performance or change specific parameters which may be more suitable for manufacturing. For that we would need to introduce the system into commercial software. A suitable representation first has to be chosen in order to best represent the SMS surfaces and to allow fast convergence of the chosen merit function. An overview of different polynomial representations of freeform surfaces is presented in [15]. Since this is an on-axis system with circular apertures, the best surface representation would be a representation of Forbs Q-polynomial presented in Equation (6.2). [16] This polynomial representation has a specific characteristic related to manufacturability estimates, which is one of the reasons their popularity in freeform designers' community is rapidly increasing.

$$f(\rho, \theta) = \frac{c\rho^2}{1 + \sqrt{1 - c^2\rho^2}} + \frac{1}{\sqrt{1 - c^2\rho^2}} \left\{ u^2(1 - u^2) \sum_{n=0}^N a_n^0 Q_n^0(u^2) + \sum_{m=1}^M u^m \sum_{n=0}^N [a_n^m \cos(m\theta) + b_n^m \sin(m\theta)] Q_n^m(u^2) \right\} \quad (6.2)$$

In the above equation, $Q_n^m(v)$ is a polynomial of order n in v , c is the curvature of the *best-fit sphere*, and $\sqrt{1 - c^2\rho^2}$ is the cosine of the angle between the z-axis and the normal to the best-fit sphere. This cosine factor means that, to a first-order approximation, the entity within braces corresponds to a departure from the best-fit sphere measured along the local normal to that sphere [16]. This is particularly convenient for obtaining fast manufacturing estimates. It is important to emphasize that the sum presented on the first line of Equation (6.2) has the same form as rotationally symmetric Qbfs presented in [17], [18], except that the coefficients and the polynomials now have a superscript 0, in order to distinguish them from the non-rotationally symmetric components a_n^m , b_n^m and $Q_n^m(u^2)$, with $m > 0$. The crucial characteristic of this representation lies in the fact that the average over θ of $f(\rho_{max}, \theta)$ is exactly the sag of the best-fit sphere at the surface rim $\rho = \rho_{max}$, thus the component inside the

braces averages to 0. The orthogonalization of this representation is conducted in terms of the mean square gradient [16].

Since we have a system with traditional circular apertures, a polynomial orthogonal inside a circular domain would behave best during the optimization. Thus, bearing in mind that apart from orthogonality Q-polynomials take into account manufacturability of the resulting surfaces, this choice seems obvious. An optimization study comparing this and other surface representations with circular apertures presented in [19] also votes in favour for choosing Forbes Q-polynomial.

However, in different applications of freeform optical systems other aperture shapes may be more useful. Thus for different system geometries, with different surface apertures an alternative representation may prove to be more suitable. [15], [20] More details about choosing the suitable surface representation for systems with rectangular apertures can be found in Chapter 7 and in ref. [21].

6.4. Aplanatism and its link to SMS design method

As stated in Chapter 5 of this thesis, optical systems which satisfy the Abbe sine condition and are free from spherical aberrations are known as aplanatic systems. In order to show that the design previously presented is semi-aplanatic up to the second order, we will first introduce some basic concepts regarding the connection between aplanatism and SMS design method.

In ref. [22] a derivation of the sine condition under non-rotationally symmetric prescription between the source and the target based on the étendue-conservation theorem was presented. We will repeat some of the derivation from that reference in order to better understand the link between the SMS method and aplanatism, and to further demonstrate it on the example considered in this chapter.

Figure 6.8 Nomenclature used in establishing the link between SMS method and aplanatism. (Source: ref. [22]) A ray is characterized by the Cartesian coordinates (x, y) when intersecting the plane $z = 0$ (object plane) and by the optical direction cosines (p, q) at this point of intersection. (p, q, r) are optical direction cosines with respect to x , y and z axes, fulfilling $p^2 + q^2 + r^2 = n^2(x, y, z)$. n and n' are refractive indices of the medium in the object and image plane, respectively.

Let us consider the image space coordinates x' , y' , p' , q' as functions of the object space variables. Expanding the image coordinates in terms of x and y , we get:

$$\begin{aligned} x' &= A(x, y, p, q) = a_{00}(p, q) + a_{10}(p, q)x + a_{01}(p, q)y + \dots \\ y' &= B(x, y, p, q) = b_{00}(p, q) + b_{10}(p, q)x + b_{01}(p, q)y + \dots \\ p' &= C(x, y, p, q) = c_{00}(p, q) + c_{10}(p, q)x + c_{01}(p, q)y + \dots \\ q' &= D(x, y, p, q) = d_{00}(p, q) + d_{10}(p, q)x + d_{01}(p, q)y + \dots \end{aligned} \quad (6.3)$$

In order to achieve aplanatism, all rays have to satisfy the stigmatic condition between one object and one image point. Coordinate systems are chosen so these points are the origins, therefore:

$$\begin{aligned} x' &= A(0, 0, p, q) = a_{00}(p, q) = 0 \\ y' &= B(0, 0, p, q) = b_{00}(p, q) = 0 \end{aligned} \quad (6.4)$$

Using étendue conservation theorem for bi-parametric bundles in combination with Eq. (6.3) and Eq. (6.4) for the point $x = y = 0$ [22] we get:

$$\begin{pmatrix} a_{10} & b_{10} \\ a_{01} & b_{01} \end{pmatrix} \begin{pmatrix} c_{00p} & c_{00q} \\ d_{00p} & d_{00q} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (6.5)$$

We define semi-aplanatism along the x -direction when the system provides first order sharp imaging along the x -axis, which means that $a_{10p} = a_{10q} = b_{10p} = b_{10q} = 0$, i.e., a_{10} , b_{10} , must be constants. Additionally we can assume that the x - y plane is rotated around the z axis to get $b_{10} = 0$. The other constant is called lateral magnification along x (and x') axes, $a_{10} = M_x$. Thus, using Eq. (6.5) for the aplanatic designs we get, in particular:

$$\begin{aligned} c_{00p} &= M_x^{-1} \\ c_{00q} &= 0 \end{aligned} \quad (6.6)$$

from which: $c_{00} = p'_0 + p/M_x$, which is the generalized Abbe sine condition [22] along the x -axes, which states that the rays coming from the origin ($x=y=0$) apart from being sent to the other origin ($x'=y'=0$) must also fulfil

$$p' = p'_0 + \frac{p}{M_x} \quad (6.7)$$

where p'_0 is a constant that represents the variable p' corresponding to the ray originating from $x = y = 0$ with direction $p = q = 0$.

Updating Eq. (6.3), we have:

$$\begin{aligned} x' &= M_x x + a_{01}(p, q)y + \dots \\ y' &= b_{01}(p, q)y + \dots \\ p' &= p'_0 + \frac{p}{M_x} + \dots \\ q' &= d_{00}(p, q) + \dots \end{aligned} \quad (6.8)$$

where the higher order terms have been explicitly omitted .

We define second-order aplanatism along the x -direction when the system provides second order sharp imaging along the x -axis, which means that apart from being aplanatic, a_{20} and b_{20} from the expansion in Eq. (6.4) also have to be constants. This means that x' and y' functions of the second-order aplanatic system have the following form

$$\begin{aligned} x' &= M_x x + a_{01}(p, q)y + \dots \\ y' &= b_{01}(p, q)y + \dots \end{aligned} \quad (6.9)$$

6.4.1. Semi-aplanatic as a limit case of the 2-point SMS3D method

The general expression of the function $x' = A(x, y, p, q)$ for an SMS design stigmatically imaging 2 points ($\alpha = (\Delta, 0)$ and $\beta = (-\Delta, 0)$) onto 2 points ($\alpha' = (M_x \Delta, 0)$ and $\beta' = (-M_x \Delta, 0)$) can be written as:

$$\begin{aligned} x' &= M_x x + F(x, y, p, q, \Delta) \\ y' &= G(x, y, p, q, \Delta) \end{aligned} \quad (6.10)$$

where F and G are general continuous functions nulling at the points α and β , i.e they are of the form:

$$\begin{aligned} F(x, y, p, q, \Delta) &= A_0 y + A_1 (x^2 - \Delta^2) \\ G(x, y, p, q, \Delta) &= B_0 y + B_1 (x^2 - \Delta^2) \end{aligned} \quad (6.11)$$

where A_i and B_i are functions of x, y, p, q which take finite values at the points α and β . Note that

$$\begin{aligned} \left. \frac{\partial F}{\partial x} \right|_{\substack{x=\Delta \\ y=0}} &= 2A_1 \Delta \\ \left. \frac{\partial F}{\partial y} \right|_{\substack{x=\Delta \\ y=0}} &= A_0 \end{aligned} \quad (6.12)$$

At the limit when $\Delta \rightarrow 0$, the two points α and β become coincident with the origin ($x = y = 0$) and so are α' and β' (with $x' = y' = 0$).

It is clear from Eq. (6.12) that the quantity $\partial F / \partial x$ at the point $\alpha = (\Delta, 0)$ is also 0 in this limit case, but not $\partial F / \partial y$. Then, according to Eq. (6.9), for that limit case we obtain:

$$\left. \frac{\partial x'}{\partial x} \right|_{\substack{x=0 \\ y=0}} = M_x \quad \left. \frac{\partial x'}{\partial y} \right|_{\substack{x=0 \\ y=0}} = A_0(0, 0, p, q) \quad (6.13)$$

In a similar way for function G we obtain:

$$\left. \frac{\partial y'}{\partial x} \right|_{\substack{x=0 \\ y=0}} = 0 \quad \left. \frac{\partial y'}{\partial y} \right|_{\substack{x=0 \\ y=0}} = B_0(0, 0, p, q) \quad (6.14)$$

From Eqs. (6.9), (6.13) and (6.14), and their equivalents in the y' coordinate, we get that in this limit case

$$\begin{aligned} x' &= M_x x + \dots \\ y' &= 0 + \dots \end{aligned} \quad (6.15)$$

which when compared to the 2 first equations of (6.8) reveals that a two point SMS design at the limit (when $\Delta \rightarrow 0$) is semi-aplanatic.

6.4.2. Semi-aplanatic of second order as a limit case of the 3-point-in-line SMS 3D method

The general expression of the function $x' = A(x, y, p, q)$ for an SMS design stigmatically imaging 3 collinear object points ($\alpha = (\Delta, 0)$, $\beta = (0, 0)$ and $\gamma = (-\Delta, 0)$) onto 3 collinear image points ($\alpha' = (M_x \Delta', 0)$, $\beta' = (0, 0)$ and $\gamma' = (-M_x \Delta', 0)$) can be written as:

$$\begin{aligned} x' &= M_x x + H(x, y, p, q, \Delta) \\ y' &= I(x, y, p, q, \Delta) \end{aligned} \quad (6.16)$$

where H and I are general continuous functions nulling at the points α , β and γ , i.e., they have the form:

$$\begin{aligned} H(x, y, p, q, \Delta) &= C_0 y + C_1 x(x^2 - \Delta^2) \\ G(x, y, p, q, \Delta) &= D_0 y + D_1 x(x^2 - \Delta^2) \end{aligned} \quad (6.17)$$

where C_i and D_i are functions that take finite values the points α , β and γ .

Doing the same analysis as before, we easily deduce that, at the limit, not only the equations equivalent to Eqs. (6.13) and (6.14) apply, but also that $\partial^2 x' / \partial x^2 = 0$ at $x=y=0$, which proves the second-order semi-aplanatism.

6.4.3. RMS3D distribution of semi-aplanatic systems

Making $\Delta = 0$ in Eq. (6.11) and Eq. (6.10), a semi-aplanatic optics has:

$$\begin{aligned} x' &= M_x x + A_0 y + A_1 x^{n+1} \\ y' &= B_0 y + B_1 x^{n+1} \end{aligned} \quad (6.18)$$

where n is the order of aplanatism (we have used A and B instead of C and D letters for the second-order aplanatic case for simplicity). The centroid of the spot at the image from rays emitted from an object point (x, y) is calculated as:

$$\begin{aligned}
\langle x' \rangle &= \frac{1}{A_{pq}} \int_{(p,q) \in pupil} x' dpdq = \frac{1}{A_{pq}} \int_{(p,q) \in pupil} (M_x x + A_0 y + A_1 x^{n+1}) dpdq = \\
&= M_x x + \langle A_0 \rangle (x, y) y + \langle A_1 \rangle (x, y) x^{n+1} \\
\langle y' \rangle &= \frac{1}{A_{pq}} \int_{(p,q) \in pupil} y' dpdq = \frac{1}{A_{pq}} \int_{(p,q) \in pupil} (B_0 y + B_1 x^{n+1}) = \\
&= \langle B_0 \rangle (x, y) y + \langle B_1 \rangle (x, y) x^{n+1}
\end{aligned} \tag{6.19}$$

Therefore, the square of the standard deviation (i.e. of the RMS) is:

$$\begin{aligned}
RMS^2(x, y) &= \left(\frac{1}{A_{pq}} \int_{(p,q) \in pupil} (x'^2 + y'^2) dpdq \right) - (\langle x' \rangle^2 + \langle y' \rangle^2) = \\
&= \frac{1}{A_{pq}} \int_{(p,q) \in pupil} \left((M_x x + A_0 y + A_1 x^{n+1})^2 + (B_0 y + B_1 x^{n+1})^2 \right) dpdq - \\
&\quad - \left((M_x x + \langle A_0 \rangle y + \langle A_1 \rangle x^{n+1})^2 + (\langle B_0 \rangle y + \langle B_1 \rangle x^{n+1})^2 \right) \\
&= \frac{1}{A_{pq}} \int_{(p,q) \in pupil} \left((M_x^2 x^2 + A_0^2 y^2 + A_1^2 x^{2(n+1)} + 2M_x A_0 xy + O(x^{n+2}) + O(y^3) + O(x^{n+1})O(y)) + \right. \\
&\quad \left. (B_0^2 y^2 + B_1^2 x^{2(n+1)} + O(x^{n+1})O(y)) \right) dpdq - \\
&\quad - \left(M_x^2 x^2 + \langle A_0 \rangle^2 y^2 + \langle A_1 \rangle^2 x^{2(n+1)} + 2M_x \langle A_0 \rangle xy + O(x^{n+2}) + O(x^{n+1})O(y) + \right. \\
&\quad \left. + \langle B_0 \rangle^2 y^2 + \langle B_1 \rangle^2 x^{2(n+1)} + O(x^{n+1})O(y) \right) \\
&= M_x^2 x^2 - M_x^2 x^2 + 2M_x \langle A_0 \rangle xy - 2M_x \langle A_0 \rangle xy + \\
&\quad + (\langle A_0^2 \rangle - \langle A_0 \rangle^2 + \langle B_0^2 \rangle - \langle B_0 \rangle^2) y^2 + (\langle A_1^2 \rangle - \langle A_1 \rangle^2 + \langle B_1^2 \rangle - \langle B_1 \rangle^2) x^{2(n+1)} + \\
&\quad + O(x^{n+2}) + O(y^3) + O(x^{n+1})O(y) + O(x)O(y^2) \\
&= (\sigma_{A_0}^2 + \sigma_{B_0}^2) y^2 + (\sigma_{A_1}^2 + \sigma_{B_1}^2) x^{2(n+1)} + O(x^{n+2}) + O(y^3) + O(x^{n+1})O(y) + O(x)O(y^2)
\end{aligned} \tag{6.20}$$

Finally, the expressions for RMS for the fields belonging to the plane ($y = 0$), and the plane ($x = 0$) become:

$$\begin{aligned}
RMS(x, 0) &= \sqrt{(\sigma_{A_1}^2 + \sigma_{B_1}^2) x^{2(n+1)} + O(x^{n+2})} = \sqrt{(\sigma_{A_1}^2 + \sigma_{B_1}^2)} |x|^{n+1} + O(x^{(n+2)/2}) \\
RMS(0, y) &= \sqrt{(\sigma_{A_0}^2 + \sigma_{B_0}^2) y^2 + O(y^3)} = \sqrt{(\sigma_{A_0}^2 + \sigma_{B_0}^2)} |y| + O(y^{3/2})
\end{aligned} \tag{6.21}$$

This proves that the cross-section $x = 0$ of a semi-aplanatic in x (regardless its order) has non-zero slope (i.e. has a kink), and that the cross-section $y = 0$ is parabolic for classical aplanatism ($n = 1$) and cubic for second-order aplanatism ($n = 2$). This would be seen in Figure 6.7 (if the limit $\Delta \rightarrow 0$ were taken), where the RMS distribution plots for these two cross sections are shown.

6.5. Conclusions

SMS 3D method, as a direct method for designing freeform surfaces presents a great asset in the fast arising field of freeform imaging systems. There is an evident need for direct design methods that will provide starting points for further system optimization. With this in

mind, SMS 3D algorithm needs to be extended to design a greater number of surfaces, and eventually generalized into a single algorithm where the number of surfaces to be designed is just an input parameter. The first step towards this generalization is the design of three freeform surfaces distributed along the same line presented in this chapter.

The detailed design procedure was explained and a design example of such a system was shown. The system presented includes four surfaces, out of which one is a prescribed plane, and the remaining three are SMS surfaces. The RMS distribution plot of the final design proves that we obtained an SMS solution with three clear zeros for the design fields. Furthermore, we suggested a suitable surface representation for further optimization of the system which in our opinion is Forbes Q-polynomial representation, since we have a circular apertures and an on-axis system. An additional argument involves manufacturability estimates that this description provides.

Additionally we demonstrated that this system presents at the limit (when the points are coincident) a second order semi-aplanatic design. Deriving the RMS spot diameter for a semi-aplanatic system, we show that such systems have an RMS spot diameter distribution curve with non-zero slope in one cross-section, and depending on the order of the aplanatism, a parabolic curve for the classical aplanatism of the 1st order, or cubic for the 2nd order (which is the case presented in this chapter) in the other cross-section.

Future lines of work related to the SMS 3D method include the generalization to three non-collinear point designs (already being developed in another thesis in our group), an extension for a greater number of surfaces, randomly distributed in the system (not necessarily on-axis). A generalization of the algorithm would be very helpful as well. This would include choosing input parameters, such as surface vertices, input and output wavefronts and an aperture stop which could be positioned anywhere in a 3D system, and the SMS construction would be automatically conducted.

REFERENCES

- [1] B. Satzer, U. Richter, U. Lippmann, G. S. Metzner, G. Notni, and H. Gross, "Using the 3D-SMS for finding starting configurations in imaging systems with freeform surfaces," in *Proc SPIE*, 2015, vol. 9626, p. 96260Y.
- [2] J. C. Miñano and J. C. González, "New method of design of nonimaging concentrators," *Appl. Opt.*, vol. 31, no. 16, pp. 3051–3060, 1992.
- [3] D. Grabovickic, P. Benítez, and J. C. Miñano, "Aspheric V-groove reflector design with the SMS method in two dimensions.," *Opt. Express*, vol. 19 Suppl 4, no. May, pp. A747–A756, 2011.
- [4] D. Grabovickic, P. Benítez, and J. C. Miñano, "Free-form V-groove reflector design with the SMS method in three dimensions.," *Opt. Express*, vol. 19 Suppl 4, no. May, pp. A747–A756, 2011.

- [5] O. Dross, R. Mohedano, P. Benítez, J. C. Miñano, J. Chaves, and J. Blen, "Review of SMS Design Methods and Real World Applications," in *Proc SPIE 5529*, 2004.
- [6] M. Buljan, "Freeform optical systems for nonimaging applications," Universidad Politécnica de Madrid, 2014.
- [7] F. Duerr, P. Benítez, J. C. Miñano, Y. Meuret, and H. Thienpont, "Analytic free-form lens design in 3D: coupling three ray sets using two lens surfaces," *Opt. Express*, vol. 20, no. 10, p. 10839, 2012.
- [8] F. Duerr, P. Benítez, J. C. Miñano, Y. Meuret, and H. Thienpont, "Analytic design method for optimal imaging: coupling three ray sets using two free-form lens profiles," *Opt. Express*, vol. 20, no. 10, p. 10839, 2012.
- [9] P. Benítez, J. C. Miñano, M. Nikolic, J. Liu, J. M. Infante, and F. Duerr, "Conditions for perfect focusing multiple point sources with the SMS design method," in *Proc. SPIE*, 2014, vol. 9191, p. 919102.
- [10] L. Wang, "Advances in the Simultaneous Multiple Surface optical design method for imaging and non-imaging applications," Universidad Politécnica de Madrid, 2012.
- [11] J. Blen, "Design of multiple free-form optical surfaces in 3D," Universidad Politécnica de Madrid, 2007.
- [12] J. M. Infante, "Optical Systems Design using the SMS method and Optimizations," Universidad Politécnica de Madrid, 2013.
- [13] J. Chaves, *Introduction to Nonimaging Optics*. CRC Press, Taylor & Francis Group, 2008.
- [14] P. Benítez *et al.*, "Simultaneous multiple surface optical design method in three dimensions," *Opt. Eng.*, vol. 43, no. 7, p. 1489, 2004.
- [15] H. Gross *et al.*, "Overview o surface representations for freeform surfaces," *Proc SPIE*, vol. 9626, p. 96260U, 2015.
- [16] G. W. Forbes, "Characterizing the shape of freeform optics," *Opt. Express*, vol. 20, no. 3, p. 2483, 2012.
- [17] G. W. Forbes, "Manufacturability estimates for optical aspheres.," *Opt. Express*, vol. 19, no. 10, pp. 9923–9941, 2011.
- [18] G. W. Forbes, "Shape specification for axially symmetric optical surfaces.," *Opt. Express*, vol. 15, no. 8, pp. 5218–5226, 2007.
- [19] C. Menke and G. W. Forbes, "Optical design with orthogonal representations of rotationally symmetric and freeform aspheres," *Adv. Opt. Technol.*, vol. 2, no. 1, pp. 97–109, 2013.
- [20] A. Brömel *et al.*, "Performance comparison of polynomial representations for optimizing optical freeform systems," in *Proc. SPIE*, 2015, vol. 9626, p. 96260W.
- [21] M. Nikolic, P. Benítez, B. Narasimhan, D. Grabovickic, J. Liu, and J. C. Miñano, "Optical design through optimization for rectangular apertures using freeform orthogonal polynomials : a case study," *Opt. Eng.*, vol. 55, no. 7, p. 71204, 2016.
- [22] J. C. Miñano, P. Benítez, and B. Narasimhan, "Freeform aplanatic systems as a limiting case of SMS," *Opt. Express*, vol. 24, no. 12, pp. 13173–13178, 2016.

Chapter 7

COMPARISON OF DIFFERENT ORTHOGONAL POLYNOMIAL FREEFORM SURFACE REPRESENTATIONS

7.1. Introduction

New trends in optical design are highly influenced by the advances in fabrication and testing technologies. This has led to regular use of rotationally symmetric aspheric elements in optical system designs. More recently, progress in manufacturing has made feasible to use non-rotationally symmetric surfaces, which provide additional degrees of freedom to design novel systems for applications that require lightweight and compact solutions. These surfaces are known as freeforms.

SMS designs that are intended to be used in imaging systems need to be optimized for the whole field of view. In Chapter 3 rotationally symmetric SMS 2D designs for imaging systems were discussed. A comparison of different design approaches was used to demonstrate that SMS method generates a very good starting point. For the purpose of optimization in a commercial software, SMS surfaces were fitted to Forbes Q-con polynomials. Many studies in the past indicate the superiority of orthogonal Forbes representation of aspheres compared to the traditional non-orthogonal polynomial representation. [1] Hence, for SMS 2D case there are no doubts regarding which representation is to be chosen for further optimization in commercial software. However, in the case of freeform systems, additional variables such as aperture shape need to be considered when deciding the suitable surface representation. Even though Forbes Q-polynomials without rotational symmetry show superiority compared to other representations in the case of circular apertures [1], there are no studies that deeply investigate other aperture shapes. Aperture shapes that are of a particular importance for high aspect-ratio systems and head mounted display systems include, above all, rectangular shapes.

The traditional description in cylindrical polar coordinates for a rotationally symmetric aspheric surface is accepted to be

$$z = \frac{c\rho^2}{1 + \sqrt{1 - (1+k)c^2\rho^2}} + \sum_{n=0} a_{2n+4}\rho^{2n+4} \quad (7.1)$$

where c and k are the axial curvature and conic constant and $a_4, a_6 \dots$ are monomial degrees of freedom. For freeform surfaces that are not rotationally symmetric, Eq. (7.1) is usually generalized using Cartesian coordinates as

The traditional representation of freeform surfaces using Cartesian coordinates is:

$$z = \frac{(c_x x^2 + c_y y^2)}{1 + \sqrt{1 - (1+k_x)c_x^2 x^2 - (1+k_y)c_y^2 y^2}} + \sum_{j=0} \sum_{k=0} a_{jk} x^j y^k, \quad (jk \neq 0 \vee j > 2 \vee k > 2), \quad (7.2)$$

having now separate curvatures c_x and c_y and conic constants k_x and k_y in x and y directions, respectively, and a_{jk} are monomial degrees of freedom.

There are many representations of freeforms. Splines representations were deeply studied in the literature. [2], [3] Specifically, nonuniform rational B-splines (NURBS) were applied to the design of both rotationally symmetric aspheres [4] and freeform mirrors [5]. NURBS representation can be very useful when the data has to be exchanged between an optical design and computer-aided design (CAD) tools. Some solutions focus on a local representation as a sum of basis functions. For instance, Cakmakci *et al.* suggest the use of linear combination of Gaussians [6]. However, the most common representations in imaging optics design are based on polynomial expansions, but using different monomial bases than the one in Eq. (7.2). These representations are widely used mostly for its suitability for optimization and easy implementation in commercial software. An interesting overview of different freeform surface descriptions used by freeform optical designers is presented by Gross *et al.* in [7] (Table 7.1).

Table 7.1. An overview of global polynomial and local descriptions of freeform surfaces with some basic properties (Source: ref [7])

polynomials	global definition	local definition	sag/gradient orthogonal	slope orthogonal	polar based	cartesian based	boundary circular	boundary rectangular	remark
Zernike	x		x		x		x		
Zernike differences	x			x	x		x		
generalized Forbes	x			x	x		x		
off-axis aspheres	x				x				
Monomials	x					x			
Chebyshev	x		x			x		x	
Legendre	x		x			x		x	
Fourier	x		x			x		x	
New polynomials	x			x		x		x	under construction
Splines		x				x			
Radial basis functions		x			x	x			can be orthogonal
Wavelets		x	x			x			

7.2. Orthogonal surface representations under investigation

In this study we were interested in comparing the behaviour of the monomial terms with different bases, therefore we annulled the conic part, which is equivalent to limiting the best-fit curvature to 0. Our emphasis is on orthogonal polynomial surface representations. Circular domain of orthogonalization can be sufficient in many optical systems that use freeform optics, which is what was done in the case of Zernike polynomial representations. Also Forbes proposed a sum of Jacobi polynomials, now usually referred to as Q-polynomials, which was mentioned multiple times in this manuscript. Q-polynomials constitute an orthonormal basis of the unit circle using a gradient based metric. The key property of Forbes' representation is that they facilitate the enforcement of fabrication constraints e.g., the deviation of the normal vector from that of a best-fit sphere [8]. This representation was introduced in Chapter 6, at the end of the section 6.3. Some comparisons of these and other representations were already presented in ref. [1]. However, for specific applications such as systems that use rectangular displays, or high aspect ratio systems, optical systems with rather rectangular apertures are needed. Here we investigate the behaviour of such systems when various orthogonal polynomials surface representations are used in a design through optimization approach. We used a standard local optimization routine. We were particularly interested in performing a 2x2 matrix comparison of four different orthogonal polynomial bases, grouped by the area of orthogonality - circular and rectangular, as well as the orthogonality metrics – sag and gradient. Circular aperture group consists of Zernike and Forbes polynomials, while in the case of rectangular apertures we include the classical Legendre polynomials [9] and a new set of polynomials that we have developed so that they are orthogonal in gradient in a rectangular domain, analogue to Forbes, which we called Q-Legendre here (Table 7.2).

Table 7.2. Table representation of 2x2 matrix comparison of different orthogonal polynomial representations.

	Sag orthogonal	Gradient orthogonal
Rectangular area	Legendre	Q-Legendre*
Circular area	Zernike	Forbes

Other polynomials are also present in literature, like rectangular Zernike [10][11] or a well-known Chebyshev polynomials, which are orthogonal in sag inside a rectangle, as Legendre. However, for this case study we decided to include only one representative per group, in order to make the comparison more concise.

7.2.1. New Q-Legendre polynomials

The Q-Legendre polynomials are a set of polynomials orthogonal in gradient achieved through Gram-Schmidt orthogonalization. The representation of a freeform surface within a normalized interval $(-1 \leq x \leq 1, -1 \leq y \leq 1)$ is supposed to have the form of Eq. (7.2) with curvatures equal to 0.

The metric used in deciding whether two functions $a(x, y)$ and $b(x, y)$ are orthogonal in gradient is as follows:

$$\langle a, b \rangle = \int_{-1}^1 \int_{-1}^1 (\nabla a \cdot \nabla b) dx dy = 0. \quad (7.3)$$

The set of linearly independent polynomials:

$$V_{jk} = x^j y^k, \quad (0 < j + k \leq N), \quad (7.4)$$

are used to compute the standard Gram-Schmidt orthogonalization process [12] as:

$$\begin{aligned} U_{10} &= V_{10} & R_{10} &= \frac{U_{10}}{\|U_{10}\|} \\ U_{01} &= V_{01} - \mathbf{proj}_{U_{10}}(V_{01}) & R_{01} &= \frac{U_{01}}{\|U_{01}\|} \\ U_{20} &= V_{20} - \mathbf{proj}_{U_{10}}(V_{20}) - \mathbf{proj}_{U_{01}}(V_{20}) & R_{20} &= \frac{U_{20}}{\|U_{20}\|} \\ U_{11} &= V_{11} - \mathbf{proj}_{U_{10}}(V_{11}) - \mathbf{proj}_{U_{01}}(V_{11}) - \mathbf{proj}_{U_{20}}(V_{11}) & R_{11} &= \frac{U_{11}}{\|U_{11}\|} \\ &\vdots & &\vdots \\ U_{kl} &= V_{kl} - \sum_{i=0}^k \sum_{j=0}^l \mathbf{proj}_{U_{ij}}(V_{kl}) & R_{kl} &= \frac{U_{kl}}{\|U_{kl}\|}, \end{aligned} \quad (7.5)$$

where

$$\mathbf{proj}_U(V) = \frac{\langle V, U \rangle}{\langle U, U \rangle} U \quad (7.6)$$

and

$$\langle U, V \rangle = \int_{-1}^1 \int_{-1}^1 \nabla U \cdot \nabla V dx dy. \quad (7.7)$$

A sample of the resulting polynomials in a triangular representation is shown in the Table 7.3 and Figure 7.1.

Table 7.3. The Q-Legendre polynomials with $T \leq 5$.

	P_{j0}	P_{j1}	P_{j2}
0	0	$\frac{1}{2}y$	$\frac{\sqrt{3}}{4}y^2$
1	$\frac{1}{2}x$	$\frac{\sqrt{6}}{4}xy$	$\frac{\sqrt{30}(-x+3xy^2)}{24}$
2	$\frac{\sqrt{3}}{4}x^2$	$\frac{\sqrt{30}(-y+3x^2y)}{24}$	$\frac{\sqrt{30}(-y^2-x^2+3x^2y^2)}{16}$
3	$\frac{\sqrt{5}(-x+x^3)}{4}$	$\frac{\sqrt{21}xy(-4+5x^2)}{24}$	$\frac{5\sqrt{7}x(-2+3x^2)(-1+3y^2)}{4\sqrt{402}}$
4	$\frac{\sqrt{7}(-6x^2+5x^4)}{16}$	$\frac{(19-120x^2+105x^4)y}{8\sqrt{94}}$	
5	$\frac{3x(x^2-1)(-3+7x^2)}{16}$		
	P_{j3}	P_{j4}	P_{j5}
0	$\frac{\sqrt{5}(-y+y^3)}{4}$	$\frac{\sqrt{7}(-6y^2+5y^4)}{16}$	$\frac{3y(y^2-1)(7y^2-3)}{16}$
1	$\frac{\sqrt{3486}(-4xy+5xy^3)}{332}$	$\frac{3x(471-70x^2-2820y^2+210x^2y^2+2345y^4)}{32\sqrt{26197}}$	
2	$\frac{5\sqrt{7}x(-2+3y^2)(-1+3x^2)}{4\sqrt{402}}$

Figure 7.1. Plots of basis function associated to polynomials shown in triangular representation in Table 7.3. Rows correspond to $j = 0,1,2,3,4,5$, and columns to $k = 0,1,2,3,4,5$.

7.3. Optical design examples

Freeform surfaces are proven to be very useful in off-axis imaging systems. They in general make the overall optical systems more compact (lighter, thinner) than their on-axis counterparts. Such systems, with tilted and decentred elements offer increased field of view and speed as compared to the on axis systems. [13][14] A demonstration of an off-axis reflective system comprising of freeform surfaces with diffraction limited performance presented by Fuerschbach et al. in [15] additionally highlighted the role of freeform surfaces in off axis imaging systems.

A specific application of of-axis freeform optical systems is for a head mounted displays [14], [16], [17]. These systems find application in both virtual and augmented reality (VR and AR), where for the second application some see-through elements must be used. An ideal solution for an HMD system would be a one surface off-axis magnifier which is the first design

example we choose for comparing the polynomials. Mirror elements, that do not show dispersion, are commonly used in these systems, since no colour correction is needed. Second design under comparison is an extended two mirror configuration.

Both designs are symmetric with the respect to the plane $x = 0$, which means that a_{jk} from Eq. (7.2) is 0 when j is odd. The basic shape is a plane ($c_x = c_y = 0$) which makes the first term in Eq. (7.2) equal to 0, leaving only the monomials part. We use surface representations satisfying the relation $j + k \leq T$, where $T = 10$ is the polynomial order. Polynomial values at the origin and their first derivatives are fixed, which is equivalent to fixing surface tilt and decentre. All the polynomials are of the same order and for each of the surfaces 34 parameters were varied during the optimization. Table 7.4 presents all surface variables.

Table 7.4. Polynomial representation to $T = 10$.

		j					
		0	2	4	6	8	10
k	0		a_{20}	a_{40}	a_{60}	a_{80}	a_{100}
	1		a_{21}	a_{41}	a_{61}	a_{81}	
	2	a_{02}	a_{22}	a_{42}	a_{62}	a_{82}	
	3	a_{03}	a_{23}	a_{43}	a_{63}		
	4	a_{04}	a_{24}	a_{44}	a_{64}		
	5	a_{05}	a_{25}	a_{45}			
	6	a_{06}	a_{26}	a_{46}			
	7	a_{07}	a_{27}				
	8	a_{08}	a_{28}				
	9	a_{09}					
	10	a_{010}					

Our starting designs use a toroid surface representation, a standard simple anamorphic shape widely used in literature. We start by fitting our systems to a desired orthogonal polynomial representation and run a local optimization in CODE V® software. In the interest of being able to confirm our conclusions with complete certainty, we keep constant field weights throughout the entire field of view.

7.3.1. One surface off-axis magnifier

We chose a simple one surface off-axis magnifier similar to the ones designed by Cakmakci [8] and Liu [18]. Starting point for all designs under comparison is the same, with the only difference being the representation of the surface. Merit function used is a pure Root Mean Square (RMS) spot size (a user defined merit function). The behaviours of 4 previously mentioned polynomial types are compared. The aspect ratio of our system is 4:3, with the field of view $9.6^\circ \times$ semi-field and 7.2° y semi-field and a pupil of 3 mm. The image plane has a rectangular aperture of 4.8 mm by 3.6 mm, in x and y dimensions respectively. We constrained

the focal length for a 0 degrees field to 14.5 mm. System layout is shown in Figure 7.2 and system specification in Table 7.5.

Figure 7.2. Optical system layout - One surface off-axis magnifier with rectangular apertures

Table 7.5. System specifications for a one-freeform mirror configuration.

Aspect ratio	4:3
Pupil	3 mm
Focal length	14.5 mm
Wavelength	550 nm
Field of view	9.6°, X semi-field 7.2°, Y semi-field

During the optimization in each system we used 17 field points, and traced 216 rays across pupil. Our free parameters, apart from freeform surface coefficients, were mirror tilt and image tilt and defocus. Optimizations for a fixed mirror tilt and for a completely fixed system were also done leading to the same conclusions. We included additional structural constraints that make sure that our system is physically realisable.

It is important to emphasize that as previously stated, we were interested in merely analysing the monomial terms of different representations, and for that reason a curvature was set to zero in the fitting process. However, in the case of Forbes Q-polynomials, this restriction resulted in an extremely bad fitting, which led to a significantly worse system with a normalized merit function (MF) value 1.8, as opposed to the initial system of the MF value 1. This characteristic of Forbes polynomials is commented with more detail further in the text. Nonetheless we conducted the analysis, which resulted in a very poor system with an MF value still exceeding the value of the initial toroid design (the optimized MF value is 1.05). Therefore, we decided to use Forbes representation as a deviation from the best fit sphere for this design example. We refer to this representation as Forbes BFS.

Once the optimization process was done, we analysed our resulting designs. Both the average Root Mean Square (RMS) spot size as the average MTF values that are displayed in Table 7.6 show the superiority of the rectangular polynomials. The optimization progress for four systems is shown in Figure 7.3. We conjecture that polynomials orthogonal inside a unit rectangle behave better than the ones orthogonal inside a unit circle. Their RMS spot size is around 8 μm smaller than in the case of Zernike or Forbes BFS, and the average MTF performance for a Legendre polynomial is 48.3% compared to the MTF performance of circular Zernike polynomials of about 32,3% (Figure 7.4).

Table 7.6. Average MTF at 24 cycles per mm and RMS Spot size values for Initial Toroid system as well as for all four optimized systems (where Forbes polynomials are compared using a best fit sphere (BFS) basis).

Surface representations	Average MTF at 24 cyc/mm [%]	Average RMS Spot size [μm]
Initial Toroid Design	20.9816	50.1192
Legendre	48.2632	26.4535
Q-Legendre	45.4500	27.8417
Zernike	32.3342	35.3547
Forbes (BFS)	32.0842	34.9402

Figure 7.3. Optimization progress for four systems under comparison.

Figure 7.4. MTF evaluated at $\lambda=550\text{nm}$ for an optimized one surface off-axis magnifier when describing a surface using Zernike polynomials (up) and when using Legendre polynomials (down).

Looking at both Figure 7.3 and Table 7.6 we can clearly distinguish two different orthogonal families of solutions obtained starting from the same initial toroid design. Regardless whether they find better or worst local minima, all designs converge very fast. We suspect that the reason for this is the simplicity of our design which allows all our systems to get to a local minimum in just a few optimization cycles. In order to make the system more complicated we introduced the second mirror and additional parameters that would make it more difficult to get to the desired design performance.

7.3.2. Two mirror configuration

In the interest of obtaining a more complicated system we changed to a two mirror configuration (Figure 7.5). The aspect ratio of our system is 16:9, with the 51.4° diagonal full field field of view, 22.4° in x semi-field and 12.5° in y semi-field. Footprint plots on the surfaces of some of the fields used for optimization are shown in the Figure 7.6. A 3 mm pupil was used and a focal length of 50 mm is maintained throughout the optimization for 16 chosen field points in order to control the distortion of our system. This is done using constraints and is the

same for all 4 designs. The constraints are controlled separately, not forming part of the merit function. The first mirror is tilted around 33.5° (clockwise from the optical z-axis), while the second mirror has 11.6° tilt and 23.7 mm decentre along y-direction. Image plane has y-decentre of 32.7 mm and a tilt of 20° , counter clockwise from z-axis. Design wavelength is 550 nm and the overall length of the system is 44.25 mm. Design specifications are shown in Table 7.7.

Figure 7.5. Two-mirror optical system layout.

Figure 7.6. Footprint plots of the selected fields on mirror 1 (left) and mirror 2 (right).

Table 7.7. System specifications for a two-freeform mirror configuration.

Aspect ratio	16:9
Pupil	3 mm
Focal length	50 mm
Wavelength	550 nm
Overall length	44.25 mm
Field of view	22.4°, X semi-field 12.5°, Y semi-field

The behaviours of four different polynomial types, Zernike, Forbes, Legendre and Q-Legendre polynomials are compared. The merit function used for the comparison is a user defined RMS spot size, for all polynomial representations. Hence the local minimum we aim to find through the optimization exists for all our designs. However, numerical calculation of the gradient of the merit function is not exact which can lead to a different value in each design, finally getting to a different local minimum [19].

Giving the complexity of our design, during the optimization in each system we use 43 field points, and trace 216 rays across the pupil. The only variables in the system are freeform coefficients, 34 per surface. We have included additional structural constraints that are preventing obscuration and therefore make sure that our system is physically realisable.

The optimization progress for all four systems is shown in Figure 7.7. Between cycles 2 and 6 a temporary increment of the merit function occurs, due to the structural constraints. However, the increment is very small and does not influence the optimization progress itself. Therefore it can be disregarded. The stopping criterion for the optimization is considered to be the stagnation of the merit function during 10 cycles. Additionally, we have confirmed that increasing the polynomial order once the local minimum is reached does not influence the differences between the resulting optimized designs.

Figure 7.7. Optimization progress for four systems under comparison.

Once the optimization process is done, we analyse our resulting designs. If our criterion is the same as the merit function used – RMS spot size, and we are only comparing polynomials with the same basis (a plane) we can get to the conclusion that the polynomials which are orthogonal inside a rectangle converge faster and get to better local minimum (Table 7.8). Resulting MTF graphs for optimized designs are shown in Figure 7.8. Average MTF values shown in Table 7.8 are evaluated at 7 cycles per mm that corresponds to Nyquist frequency of a display of pixel size of 60 μm , which is rather representative of high-resolution LCD and OLED displays of low-cost backplane technologies. In this particular design, our new Q-Legendre polynomials show the best performance, but only slightly better than the Legendre polynomials (RMS spot size of 68.3 μm and an average MTF of 60.9% at 7 cycles per mm). This most likely occurs because the good properties of orthogonality can be lost if the surface area used does not

coincide with the orthogonality area. The good properties are mainly based on the fact that the cancelation between different monomial terms is avoided using orthogonal polynomials [20]. We suspect that the reason is the mismatch between the area of orthogonality of the polynomials and that of the aperture, but in order to be able to claim that with complete certainty we need additional studies which will be conducted in the future.

Table 7.8. Average MTF and RMS Spot size values.

Surface representations	Average MTF at 7 cycles/mm [%]	Average RMS Spot size [μm]
Initial Toroid Design	8.4	195.1
Legendre	55.5	75.5
Q-Legendre	60.9	68.3
Zernike	44.1	89.3
Forbes	61.3	77.8

Figure 7.8. MTF curves for four optimized designs for a central 0° field, 70% in both +y and -y directions, and a full field. The legend is presented on the left.

The starting value of the merit function for Forbes for a plane basis is not equal to the ones of the rest of the polynomials. This is due to the inability to achieve a good fit from the starting toroid design, which is related to using a best-fit curvature of 0. Forbes polynomials seem to be particularly sensitive to the value of the best-fit curvature. Notice that Forbes polynomials are the only ones considered which are not directly transformable into the other polynomial types, because the rotationally symmetric terms (with index $m = 0$ in Forbes nomenclature, which corresponds to the aspheric Qbfs) are constrained by definition to a zero sag at the rim of the best-fit sphere [20]. This means that the best-fit curvature for rotationally symmetric surfaces cannot be freely chosen and its value is unique, therefore the value selected

for freeforms may be somewhat constrained as well. In other words, in the case of Forbes polynomials, monomial terms are not being able to compensate for the lack of not using the best fit sphere approach as in the case of other 3 polynomials analysed. Nevertheless, in this design example, bad starting value does not prevent them to get to a very good local minimum, with the RMS spot size of 77.8 μm , which is extremely close to the RMS obtained by Legendre polynomials. The average MTF value for Forbes polynomials (61.3%) even exceeds the MTF of Legendre and Q-Legendre polynomials (55.5% and 60.9%, respectively). However, they do not converge as fast as the rectangular polynomials, and this could be related to having bad initial fitting and not the fact that the orthogonality area is not matched. Calculated peak and RMS sag departure from the best-fit sphere along with a peak and RMS surface slope for each of the resulting four design forms is presented in Table 7.9. It is important to emphasize that the best-fit sphere used for sag departure calculation is calculated separately for each of the designs.

Table 7.9. Peak and RMS sag departure from the best-fit sphere (BFS) and peak and RMS surface slope for Mirror 1 (top) and Mirror 2 (bottom)

Mirror 1	Peak sag departure from BFS	RMS sag departure from BFS	Peak surface slope in X (dZ/dX)	Peak surface slope in Y (dZ/dY)	RMS surface slope in X (dZ/dX)	RMS surface slope in Y (dZ/dY)
Legendre	0.955	0.520	-0.550	0.248	0.273	0.117
Q-Legendre	0.766	0.393	-0.520	0.227	0.275	0.116
Zernike	1.175	0.638	-0.532	0.298	0.268	0.120
Forbes	0.745	0.362	-0.500	0.231	0.273	0.119
Mirror 2	Peak sag departure from BFS	RMS sag departure from BFS	Peak surface slope in X (dZ/dX)	Peak surface slope in Y (dZ/dY)	RMS surface slope in X (dZ/dX)	RMS surface slope in Y (dZ/dY)
Legendre	-1.612	0.737	-0.442	-0.160	0.213	0.076
Q-Legendre	-1.710	0.804	-0.454	-0.163	0.240	0.084
Zernike	-1.178	0.465	-0.361	0.199	0.186	0.098
Forbes	-0.903	0.435	-0.495	0.293	0.201	0.106

7.4. Conclusions

Although there is no single freeform surface representation expected to be suited to all optical systems, the optimizations done in this study show superior behaviour of the polynomials with the rectangular base in optical systems of rectangular apertures. The reason may be that the orthogonal polynomials can lose their good properties when the area of

orthogonality does not coincide with the aperture of the surface they represent. When polynomial representations are analogue, such as the triangular distribution of the coefficients and the sole power series expansion representation (without the conic part of the surface) we see in the examples analysed better conduct of the polynomials orthogonal inside a rectangle. Between the two polynomials with rectangular basis considered, the new presented Q-Legendre polynomials with slope orthogonality behave slightly better than the sag orthogonal Legendre. It is somewhat analog to the case of circular aperture where it was proven that Forbes polynomials orthogonal in gradient (i.e. slope) show better behaviour compared to Zernike polynomials that are orthogonal in sag.

However, it is not clear if the results can be generalised, and additional studies for other design cases will be needed to assert it. Investigation of cases where plane symmetry is excluded could offer a valuable supplement to the study performed. Moreover, doing the comparison using polynomials represented as a deviation from a base conic, as well as a deviation from a best fit sphere base would also enrich the analysis.

REFERENCES

- [1] C. Menke and G. W. Forbes, "Optical design with orthogonal representations of rotationally symmetric and freeform aspheres," *Adv. Opt. Technol.*, vol. 2, no. 1, pp. 97–109, 2013.
- [2] J. E. Stacy, "Asymmetric spline surfaces : characteristics and applications," *Appl. Opt.*, vol. 23, pp. 2710–2714, 1984.
- [3] P. Jester, C. Menke, and K. Urban, "B-spline representation of optical surfaces and its accuracy in a ray trace algorithm.," *Appl. Opt.*, vol. 50, no. 6, pp. 822–828, 2011.
- [4] H. Chase, "Optical Design with Rotationally Symmetric NURBS," in *Proc SPIE*, 2002, vol. 4832, pp. 10–24.
- [5] P. Ott, "Optic design of head-up displays with freeform surfaces specified by NURBS," in *Proc SPIE*, 2008, vol. 7100, pp. 1–12.
- [6] M. P. Chrisp and W. Street, "New Freeform NURBS Imaging Design Code," in *Proc SPIE*, 2014, vol. 9293, pp. 1–6.
- [7] H. Gross, A. Brömel, M. Beieer, R. Steinkopf, J. Hartung, Y. Zhong, M. Oleszko, and D. Ochse, "Overview o surface representations for freeform surfaces," *Proc SPIE*, vol. 9626, pp. 1–14, 2015.
- [8] O. Cakmakci, B. Moore, H. Foroosh, and J. P. Rolland, "Optimal local shape description for rotationally non-symmetric optical surface design and analysis.," *Opt. Express*, vol.

- 16, no. 3, pp. 1583–1589, 2008.
- [9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, 1972.
- [10] V. N. Mahajan and G. Dai, “Orthonormal polynomials in wavefront analysis : analytical solution,” *J. Opt. Soc. Am.*, vol. 24, no. 9, pp. 2994–3016, 2007.
- [11] A. H. Al-Hamdani and S. Y. Hasan, “Zernike polynomials for optical systems with rectangular and square apertures of area equal to π ,” *Indian J. Pure Ap. Phys.*, vol. 51, pp. 837–843, 2013.
- [12] W. Cheney and D. Kincaid, “Chapter 7,” in *Linear Algebra: Theory and Applications*, Jones and Bartlett Learning, 2008.
- [13] T. Nakano and Y. Tamagawa, “Configuration of an off-axis three-mirror system,” *Appl. Opt.*, vol. 44, no. 3, pp. 776–783, 2005.
- [14] O. Cakmakci and J. Rolland, “Design and fabrication of a dual-element off-axis near-eye optical magnifier,” *Opt. Lett.*, vol. 32, no. 11, pp. 1363–1365, 2007.
- [15] K. Fuerschbach, J. P. Rolland, and K. P. Thompson, “A new family of optical systems employing ϕ - polynomial surfaces,” *Opt. Express*, vol. 19, no. 22, pp. 21919–21928, 2011.
- [16] O. Cakmakci and J. Rolland, “Head-worn displays: A review,” *IEEE/OSA J. Disp. Technol.*, vol. 2, no. 3, pp. 199–216, 2006.
- [17] O. Cakmakci, G. E. Fasshauer, H. Foroosh, K. P. Thompson, and J. P. Rolland, “Meshfree approximation methods for free-form surface representation in optical design with applications to head-worn displays,” in *Proceedings of SPIE*, 2008, vol. 7061, pp. 1–15.
- [18] J. Liu, J. C. Miñano, and P. Benítez, “Single freeform surface imaging design with unconstrained object to image mapping,” *Opt. Express*, vol. 23, no. 10, p. 13370, 2015.
- [19] W. Ulrich, “Freeform Surfaces Systems for Imaging System,” *Free. Opt. Incubator*, 2011.
- [20] G. W. Forbes, “Characterizing the shape of freeform optics,” *Opt. Express*, vol. 20, no. 3, p. 2483, 2012.

CONCLUSIONS AND FUTURE WORK

1. Conclusions

The majority of this thesis deals with different contributions to the SMS method for imaging optical systems both in two dimensions (aspheric optics) and three dimensions (freeform optics). Many interesting conclusions were reached along the way.

In Chapter 2 a demonstration of sharp imaging of multiple point sources was given. Starting from the example of a generalized Cartesian oval, the reasoning was extended to the standard two-surface SMS design. It was shown that traditional way of seeing the SMS method as a method that perfectly couples a chosen number of wavefront pairs with the same number of surfaces does not necessarily have to apply to all the systems. The condition for achieving sharp imaging of multiple point sources was that the aperture stop of the system is placed further away from the surfaces. In that way an additional ray bundle can be coupled increasing the parallelism. A general rule of thumb that gives an estimation of the number of object points that can be sharply imaged with the given surface size and aperture was suggested. Additionally, it was indicated that the SMS designs with maximized number of object points sharply imaged are the designs with exhausted degrees of freedom. This conclusion is important for imaging applications of the method. Namely, it is implied that SMS designs with exhausted degrees of freedom present a good starting point for posterior optimization. This assertion was further studied in Chapter 3 on a two-lens design problem.

In Chapter 3 an analysis of SMS 2D as a starting point for a two-lens design problem was conducted. Two different systems were analysed, the second being slightly more difficult. SMS 2D designs were optimized in a single-step, while the spherical starting points were optimized using a one-step, a two-step and a stepwise approach. Global optimization method from CODE V® software was compared as well. In the case of the simpler system analysed, all design techniques considered found solutions close to diffraction limit. Nevertheless, surface shapes of the final designs were different which suggested that their optimization paths ended in different basins of attractions. The fact that there are various minima with diffraction limited performance further proves that the solution to this problem is an easy one. The second system analysed had a larger field and a bigger pupil. In this case only the SMS with a subsequent optimization, a stepwise approach and a global optimization found optimal solutions. The one-step and two-step design techniques either went into ray failure (two-step) or got trapped in a bad minimum (one-step). Carefully analysing the two results obtained, we conjectured that

they belong to different minima, which is to be expected since different optimization strategies lead to different basins of attraction. However, the stepwise approach took around 6 times longer than the posterior SMS optimization. Therefore, the assertion that the SMS method provides a good starting point for optimization seems to be correct, at least in case of the examples considered.

In Chapter 4 foveated imaging design approach for two different applications was studied. First, the SMS method with pupil discretization was used to generate different starting points to redesign an aspheric part of the endoscopic microscope objective lens with high distortion needed for foveation. Changing only the seed patch of the surfaces, 3 different surface sets were obtained. Since SMS designs are in general close to the final solutions as demonstrated in Chapter 3, a short optimization run was used to optimize the surfaces for the whole pupil. The method turned out to be very effective in following the prescribed distorted mapping, and the fact that various starting points could be generated changing only one condition proved to be very useful for obtaining milder optical surfaces, which is one of the potential issues of foveated imaging designs.

Second foveation application presented in this chapter included a two-lens virtual reality HMD system. The system presented was obtained via a traditional multi-parametric optimization, but it included novel design approaches. Apart from foveation approach which is not normally seen in commercial VR HMD designs, rotations of the human eye were also considered. The second is a technique commonly encountered in a spectacle lens design, and it proved to be particularly useful for designing VR HMD optics. Using the foveated mapping function in combination with eye-rotation technique, a design closely adapted to the resolution of the human eye was achieved. Since the digital displays available for VR HMD systems exhibit low spatial resolution, the problems related to manufacturing strong aspheric surfaces disappear, since the resolution of the nominal design highly surpasses the resolution of the display and therefore fairly loosed tolerances are acceptable. This design was manufactured via a direct-cut technique, and its spatial resolution was measured experimentally using two standard methods: a USAF resolution test chart and a Slanted-Edge MTF measurement. As expected, the manufactured design shows MTF values above 50% (evaluated at the Nyquist frequency of the display) for the entire usually gazed region (0-20°).

In Chapter 5 an afocal two freeform mirror design problem was solved analytically using partial differential equations. It was proven that for a single imposed condition that coordinate systems are coincident this system has a decoupled ray-to-ray mapping equations, one relating x -variables and the other y -variables, with full symmetry between them. Furthermore, it was demonstrated that for a semi-aplanatic case, where Abbe sine condition is fulfilled in x -direction, choosing a specific contour condition a full freeform two-mirror aplanat can be achieved. This is the example of an aplanatic system without rotational symmetry that uses fewer surfaces than the general aplanat in three dimensions, which requires three surfaces. Alternative mirror system geometries and a refractive case were investigated as well, but the ray-to-ray mapping function in those situations is not necessarily decoupled. Moreover, the

problem presented in this chapter can also be seen as a limit case of SMS 3D design problem with $\alpha = \alpha(x)$ when α tends to zero as discussed in the text.

Chapter 6 introduced a new contribution to the SMS 3D design method. An algorithm for designing 3 plane-symmetric freeform surfaces using SMS 3D method was developed in a similar manner as the existing SMS 3D for two surfaces. A specific design example was presented. The resulting system fulfilled stigmatic imaging requirements for the design fields, as expected. Furthermore, we demonstrated that for the limit case, when three design wavefronts are coincident, the example shown becomes a semi-aplanatic design, where the aplanatic condition is fulfilled in x-direction up to the second order. Additionally, we showed that semi-aplanatic systems have a specific characteristic regarding the RMS spot size distribution curve: in one cross-section this curve has a non-zero slope, and in the other cross-section, depending on the order of the aplanatism, it has either a parabolic shape for the classical aplanatism of the 1st order, or a cubic shape for the 2nd order aplanatism (which is the case presented in this chapter).

As previously stated SMS designs for imaging systems present a starting point for multi-parameter optimization. For rotationally symmetric SMS designs, it is easy to choose the surface representation for further optimization, since those representations have been extensively studied in the past. However, in the case of freeform systems, additional variables such as aperture shape need to be considered when deciding the suitable surface representation. Even though specific studies showed that Forbes Q-polynomials without rotational symmetry show superior behaviour compared to other representations in the case of circular apertures, there are no studies that deeply investigate non-circular aperture shapes.

Aperture shapes that are of a particular importance for high aspect-ratio systems and head mounted display systems include, above all, rectangular shapes. A much needed study of surface representations for optical systems with rectangular apertures was presented in Chapter 7. Two freeform mirror design examples were considered and four orthogonal surface shapes were compared. Even though there is no single freeform surface representation expected to be suited to all optical systems, the results of this study show superior behaviour of the polynomials with the rectangular orthogonal base. The suspected reason is that the orthogonal polynomials can lose their good properties when the area of orthogonality does not coincide with the aperture of the surface they represent. Between the two polynomials with rectangular basis considered, the new presented Q-Legendre polynomials with slope orthogonality showed a slightly better behaviour than the sag orthogonal Legendre polynomial representation. It is somewhat analog to the case of circular aperture where it was proven that Forbes Q-polynomials orthogonal in slope show better behaviour compared to Zernike polynomials that are orthogonal in sag.

2. Future work

Regarding the first part of this thesis (Aspheric optics) and in particular the foveated imaging approach (Chapter 4), an interesting study that would explore the possibilities of the SMS method sharply imaging multiple point sources presented in Chapter 2 is desirable. In this way, two different design strategies within the SMS method (SMS method with pupil discretization and SMS method with sharp imaging of multiple object points) could be directly compared.

The comparative study presented in Chapter 3 that shows the superiority of SMS 2D method compared to a traditional spherical system when used as a starting point for optimization, should be extended to freeform case. Even though some preliminary studies of the potential of SMS 3D as a starting point were shown, a more extensive analysis that would include different system geometries is desirable.

An interesting line of investigation being developed in our research group in the past years involves the application of the design approach for VR HMD systems presented in Chapter 4 (including foveation and eye rotations) to the design of freeform VR HMD systems in combination with multichanneling. This technology (known as ThinEyes® technology) successfully exploits multichannel idea for reducing the overall length of the systems while keeping the acceptable resolution. In these designs it is essential to combine the multichannel approach with freeform surfaces. Each channel can be designed separately, dealing with the specific problems depending on the area of field of view it covers, and freeform geometry provides us with additional degrees of freedom needed. In this way we can use the features of freeform surfaces to further improve our system performance, either improving the spatial resolution or increasing the field of view for the purpose of more immersive experience. Combining multichanneling approach with foveation and eye rotation techniques, apart from providing the optics that closely follows the resolution of the eye, it makes the whole system more compact, solving one of the main issues of head-worn-display technology.

Regarding the part of the thesis titled “Freeform optics”, various future lines of investigations are proposed.

In Chapter 5, apart from the extensively studied freeform mirror system geometry with coincident input and output wavefronts, alternative mirror geometries, together with the refractive system, were shortly considered. However, the partial differential equations for these alternative geometries were not solved. In order to know whether cases of decoupled mapping, or even aplanatic systems with these geometries exist, these partial differential equations have to be solved. Investigation of system geometries not included here could also present an interesting line of investigation.

Future lines of investigation related to the SMS 3D method include an extension for designing 3 freeform surfaces where the object and image points are not distributed along a

single line. This work is already being developed in our research group as a part of another PhD thesis. Specifically, a design of three object points where two of them are positioned on x-axis and the third one is not will be shown in the near future. Additionally, an extension of SMS 3D method to a design of more than three surfaces is planned. A generalization of the algorithm would be very helpful as well. This generalization would include choosing input parameters, such as surface vertices, and input and output wavefronts which could be positioned anywhere in a 3D system, and using an iterative process to calculate SMS freeform surfaces.

The results presented in Chapter 7 suggest the superiority of the polynomials orthogonal in a rectangular domain when optical systems with rectangular apertures are considered. However, it is not clear if the results can be generalised, and additional studies for other design cases are needed to assert it. Investigation of cases where plane symmetry is excluded could offer a valuable supplement to the study performed. Moreover, doing the comparison using polynomials represented as a deviation from a base conic as well as a deviation from the best fit sphere would also enrich the analysis.

APPENDIX A

Specification of the machine vision camera and the lens used in experimental resolution measurements

In the following pages the main characteristics of the machine vision camera with a Sony® EXMOR CMOS sensor and the Ricoh® objective lens are shown. They were used in the experimental resolution measurements of a VR HMD lens presented in Chapter 3.

1. The Imaging Source® DFK23UX174 machine vision camera

The specification of the camera along with its spectral response curve and mechanical drawings explaining the size are shown.

DFK 23UX174 Specification

GENERAL BEHAVIOR

Dynamic range	8/12 bit
Video formats @ frame rate (maximum)	1,920×1,200 (2.3 MP) Y800 @ 54 fps 1,920×1,200 (2.3 MP) Y16 @ 25 fps 1,920×1,200 (2.3 MP) RGB32 @ 54 fps

INTERFACE (OPTICAL)

IR cut filter	✓
Sensor type	CMOS Exmor
Sensor specification	Sony IMX174LQ
Shutter	global
Format	1/1.2 inch
Resolution (maximum)	H: 1,920 px, V: 1,200 px
Frame rate (maximum)	54 fps
Pixel size	H: 5.86 µm, V: 5.86 µm
Lens mount	C/CS

INTERFACE (ELECTRICAL)

Interface	USB 3.0
Supply voltage	4.5 VDC to 5.5 VDC
Current consumption	approx 250 mA @ 5 VDC
Auto iris control	✗
Trigger	✓
I/Os	✓

INTERFACE (MECHANICAL)

Dimensions	H: 29 mm, W: 29 mm, L: 43 mm
Mass	65 g

ADJUSTMENTS

Shutter	1/100,000 s to 30 s
Gain	0 dB to 36 dB
White balance	-2 dB to 6 dB

ENVIRONMENTAL

Temperature (operating)	-5 °C to 45 °C
Temperature (storage)	-20 °C to 60 °C
Humidity (operating)	20 % to 80 % (non-condensing)
Humidity (storage)	20 % to 95 % (non-condensing)

Subject to change

Sony IMX174LQ Spectral Response Curve

CMOS Exmor Sensor in DFK 23UX174

DFK 23UX174 Dimensional Diagram

Dimensions: mm
Tolerances: DIN ISO 2768m

2. SONY® IMX174LQ sensor used in the camera

More details regarding the camera CMOS sensor are presented.

SONY

IMX174LLJ/IMX174LQJ

Diagonal 13.4mm (Type 1/1.2) 2.35M-Effective Pixel
Monochrome and Color CMOS Image Sensors

CMOS Image Sensors with Global Shutter Function for Industrial Use

In such cases where a high-speed moving subject is to be shot with standard CMOS Image sensors, focal plane distortion will occur and it might be a problem especially for the fields of industrial use. The new sensors feature a global

shutter function and able to capture a high-speed moving image without focal plane distortion. Also the sensors have various functions including trigger mode or ROI (Region of interest) mode specially required in the fields of industrial use.

- Global shutter function
- High frame rate (10 bit 164.5 frame/s,
12 bit 128.2 frame/s)
- ROI mode (able to set at a maximum of 16 areas)
- Trigger mode (able to control accumulation
time by external pulse)
- A variety of operating modes (V inversion
output and multiple frame set output mode)

Exmor

* Exmor is a trademark of Sony Corporation. The Exmor is a version of Sony's high performance CMOS image sensor with high-speed processing, low noise and low power dissipation by using column-parallel A/D conversion.

Pregius

* Pregius is a trademark of Sony Corporation. The Pregius is global shutter pixel technology for active pixel-type CMOS image sensors that use Sony's low-noise CCD structure, and realize high picture quality.

Global Shutter Function

For industrial use, it is required to capture the exact shape of a high-speed moving subject. The existing CMOS image sensors have a rolling shutter as the electronic shutter function; therefore, focal plane distortion was inevitable in principle. A new pixel with analog memory was developed for the IMX174LLJ/LQJ and eliminated generation of the focal

plane distortion by enabling to scan all pixel signals at once (referred to as "global shutter function" below). (See photograph 1.) The sensors accomplished high-picture quality combining with the column-parallel A/D conversion technology used for the existing Sony's CMOS image sensors. (See photograph 2.)

High Frame Rate

For industrial use, it is required to capture the sequence of high-speed moving subject; therefore, frame rate is equally important as global shutter function. The IMX174LLJ/LQJ optimized the internal circuit and adopted LVDS 8ch output,

and then realized high-speed imaging by achieving high frame rate with a maximum 164.5 frame/s using 10 bit ADC, and with a maximum 128.2 frame/s using 12 bit ADC. (See table 3.)

ROI Mode

To meet market needs, there is a requirement to capture moving subject only in partial areas rather than a whole image view angle. ROI mode makes this possible. The IMX174LLJ/LQJ have the ROI mode and able to set at a

maximum of 16 areas inside the effective pixels and output cropped partial pictures out of the image view angle. Also, since ROI mode scans less pixel count, even higher frame rate is achievable. (See photograph 3.)

Trigger Mode

The IMX174LLJ/LQJ are equipped with trigger mode, and the external pulse can control accumulation time. The sensors also have a pulse output function to indicate respective condi-

tions during shutter operation and can be coordinated with peripheral circuits.

A Variety of Operating Modes

The IMX174LLJ/LQJ have internal register settings which can switch vertical scan direction (normal/inverted) of the sensors and support 2 frame set output mode utilizing high-speed frame rate. In this 2 frame set output mode, separate exposure time can be set for 2 consecutive frames, and each set of 2 frames

can be handled as 1 set to automatically output consecutive images. This mode can generate a picture with a wide dynamic range as a result of the combination of multiple frames. Also the sensors support all-pixel scan (WUXGA), including UXGA, and full HD scan mode as the image data output format.

<photograph 1>
Moving Object Image

<Photograph 2>
Sample Images

Condition: 2000 lx F5.6 (UXGA image)
ADC 12 bit mode, 60 frames/s,
internal gain 0 dB

<Photograph 3>
ROI Mode Cropping Images

Condition: 2000 lx F5.6 (UXGA image)
ADC 12 bit mode, 60 frames/s, internal
gain 0 dB

<Table 1> Device Structure

Item	IMX174LLJ/IMX174LQJ
Image size	Diagonal 13.4 mm (Type 1/1.2) (WUXGA mode) Diagonal 11.9 mm (Type 1/1.35) (UXGA mode) Diagonal 13.0 mm (Type 1/1.23) (Full HD mode)
Number of effective pixels	1836 (H) × 1216 (V) Approx. 2.35M pixels
Unit cell size	5.86 μm (H) × 5.86 μm (V)
Optical blacks	Horizontal: First: 0 pixels, rear: 0 pixels
	Vertical: First: 10 pixels, rear: 0 pixels
Input drive frequency	37.125 MHz / 74.25 MHz
Package	116-pin LGA
Power supply voltage V _{DD} (Typ.)	3.3 V / 1.8 V / 1.2 V

<Table 2> Image Sensor Characteristics

Item	IMX174LLJ/IMX174LQJ	Remarks
Sensitivity	Typ. 825mV (F8.0 Monochrome) / 1000mV (F5.6 color)	1/30s accumulation
Saturation signal	Min. 850 mV	T _J = 60 °C

<Table 3> Basic Drive Mode

Drive mode	Number of recommended recording pixels	ADC	Frame rate
All-pixel scan (WUXGA)	1920 (H) × 1200 (V) Approx. 2.30M pixels	10 bit	164.5 frames/s
		12 bit	129.2 frames/s
UXGA	1600 (H) × 1200 (V) Approx. 1.95M pixels	10 bit	164.5 frames/s
		12 bit	129.2 frames/s
Full HD	1920 (H) × 1080 (V) Approx. 2.07M pixels	10 bit	120.0 frames/s
		12 bit	120.0 frames/s

3. RICOH® FL-BC7528-9M lens specification

The information of the lens used with the previously presented camera is given herein. A detail specification followed with the mechanical drawings is shown.

SPEC. INFO	FL-BC7528-9M	K65784-5001-813 Dec. 05, 2014
1. Model	FL-BC7528-9M	
2. Application	For 1 Format Camera	
3. Specification		
3-1. Physical (As per the attached drawing)		
(1) Dimensions	$\phi 42.0 \text{ mm} \times 81.0 \text{ mm}$	
(2) Weight	189 g	
(3) Mount	C-Mount (1 – 32 UN – 2A)	
(4) Filter Screw Size	M40.5 mm, P = 0.5 mm	
3-2. Optical		
(1) Focal Length	75 mm $\pm 5\%$	
(2) Max. Aperture Ratio	1: 2.8	
(3) Iris Range	F / 2.8 to F / 32	
(4) Angle of View	Diagonal	12.2°
		Horizontal
	Vertical	7.3°
(5) Picture format	12.8 mm $\times$ 9.6 mm	
(6) Focusing Range	Inf. to 0.25 m	
(7) Back Focal Length	21.3 mm (in air)	
(8) Flange Back Length	17.526 mm $\pm$ 0.05 mm (in air)	
3-3. Mechanical		
(1) Focusing Mechanism	Straight Helicoid	
	Rotation Angle 309°	
	Rotation Torque 0.2 N·cm to 7.0 N·cm	
	Focus Ring can be fixed at any positions by	
	3 fixing screws (A) or 1 thumb screw (refer to external view)	
(2) Iris Mechanism	Non-click, Close-down	
	Rotation Angle 86°	
	Rotation Torque 0.01 N·cm to 3.0 N·cm	
	Iris Ring can be fixed at any positions by 3 fixing screws (B)	
	or 1 thumb screw (refer to external view)	
3-4. Vibration & Shock Test Standard		
(1) Vibration Conditions		
Wave	Sine Wave	
Acceleration	1.96 m/s ² to 59.78 m/s ²	
Cycle Range	10 Hz to 55 Hz	
Cycle Period	3 Minutes	
Duration of Vibration	120 Minutes Respectively in Three	
	Dimensional Directions	
(2) Shock Conditions		
Tester	Shock Testing Machine by Dropping	
Impact Force	588 m/s ²	
Duration of Shock	3.5 ms	
Directions & Number of Times	1 Time Respectively in Three Dimensional	
	Directions	
3-5. Environmental Temperature Range	-10°C to +50°C

SPEC.INFO	FL-BC7528-9M
	K65784-S001-B33
RICOH Industrial Solutions Inc. Dec.05.2014	

PUBLICATIONS

Journal papers

1. Pablo Benitez, **Milena Nikolic**, and Juan C. Miñano, "Analytical solution of an afocal two freeform mirror design problem," *Opt. Express* **25**, 4155-4161 (2017).
<https://doi.org/10.1364/OE.25.004155>
2. **Milena Nikolic**, Pablo Benítez, Bharathwaj Narasimhan, Dejan Grabovickic, Jayao Liu, Juan C. Miñano, "Optical design through optimization for rectangular apertures using freeform orthogonal polynomials: a case study," *Opt. Eng.* **55**(7), 071204 (2016).
<http://dx.doi.org/10.1117/1.OE.55.7.071204>.
3. Bharathwaj Narasimhan, Pablo Benitez, Juan C. Miñano, **Milena Nikolic**, Dejan Grabovickic, "Freeform aplanatic systems," submitted for publication in *Opt. Express*.

Patents:

1. "Display devices with reflectors", Inventors: Pablo Benitez, Juan Carlos Miñano, Dejan Grabovickic, **Milena Nikolic**, WO2016160099 A3, 24 November 2016.

Conference proceedings

1. P. Benitez, J. C: Miñano, P. Zamora, D. Grabovickic, M. Buljan, **M. Nikolic**, B. Narasimhan, J. Gorospe, J. Lopez, E. Sanchez, "Advanced freeform optics enabling ultra-compact VR headsets", accepted for SPIE Digital Optical Technologies Conference, scheduled for 25-29 of June in Munich, Germany.
2. D. Grabovickic, P. Benitez, J. C: Miñano, P. Benitez, M. Buljan, B. Narasimhan, **M. Nikolic**, J. Lopez, E. Sanchez, "Super-resolution optics for Virtual Reality", accepted for SPIE Digital Optical Technologies Conference, scheduled for 25-29 of June in Munich, Germany.
3. **M. Nikolic**, J. C. Miñano, P. Benítez, B. Narasimhan, J. Mendes-Lopes, P. Zamora, M. Buljan, D. Grabovickic, "Design of compact optical systems using multichannel configurations", SPIE, Paper No. 9948-21, Novel Optical System Design and Optimization XIX, SPIE Optics + Photonics (August 2016).
4. J. C. Miñano, P. Benítez, J. Mendes-Lopes, B. Narasimhan, **M. Nikolic**, D. Grabovickic, "Recent advances in the SMS method: Diffraction and 3D aplanatism", SPIE, Paper No.

- 995-12, Nonimaging Optics: Efficient design for illumination and solar concentration XIII, SPIE Optics + Photonics (August 2016).
5. **M. Nikolic**, P. Benitez, J. C. Miñano, D. Grabovickic, J. Liu, B. Narasimhan, M. Buljan, "Optical design through optimization using freeform orthogonal polynomials for rectangular apertures", Proc. SPIE 9626, Optical System Design 2015: Optical Design and Engineering VI, 96260V (23 September 2015).
 6. B. Narasimhan, P. Benitez, D. Grabovickic, J.C. Miñano, **M. Nikolic**, J. Infante, "Design of Freeform aplanatic systems", Proc. SPIE 9629, Optical System Design 2015: Illumination Optics IV, 96290C (23 September 2015).
 7. **M. Nikolic**, P. Benitez, J. C. Miñano, D. Grabovickic, J. Liu, B. Narasimhan, M. Buljan, "Comparative analysis of optimization with freeform orthogonal polynomials for rectangular apertures", SPIE, Paper No. 9579-16 (August 2015)
 8. B. Narasimhan, P. Benitez, D. Grabovickic, J.C. Miñano, **M. Nikolic**, J. Infante, "Freeform aplanatic concentrators" Proc. SPIE 9572, Nonimaging Optics: Efficient Design for Illumination and Solar Concentration XII, 95720M (25 August 2015).
 9. B. Narasimhan, P. Benitez, D. Grabovickic, J.C. Miñano, **M. Nikolic**, J. Infante, "Design of three freeform mirror aplanat", Proc. SPIE 9579, Novel Optical System Design and Optimization XVIII, 95790K (3 September 2015).
 10. P. Benitez, J. C. Miñano, **M. Nikolic**, J. Liu, J. Infante, F. Duerr, "Conditions for perfect focusing multiple point sources with the SMS design method" Proc. SPIE 9191, Nonimaging Optics: Efficient Design for Illumination and Solar Concentration XI, 919102 (17 September 2014).

Oral presentations in international conferences

1. P. Benitez, J. C. Miñano, P. Zamora, D. Grabovickic, M. Buljan, **M. Nikolic**, B. Narasimhan, J. Gorospe, J. Lopez, E. Sanchez, "Advanced freeform Optics enabling ultra-compact VR headsets", accepted for oral presentation in SPIE Digital Optical Technologies Conference, scheduled for 25-29 of June in Munich, Germany.
2. D. Grabovickic, P. Benitez, J. C. Miñano, P. Benitez, M. Buljan, B. Narasimhan, **M. Nikolic**, J. Lopez, E. Sanchez, "Super-resolution optics for Virtual Reality", accepted for oral presentation in SPIE Digital Optical Technologies Conference, scheduled for 25-29 of June in Munich, Germany.
3. P. Benítez, J. C. Miñano, **M. Nikolic (presenter)**, "Multichannel optics for imaging and nonimaging applications", UK Optical Design Meeting (September 2016).
4. **M. Nikolic (presenter)**, J. C. Miñano, P. Benítez, B. Narasimhan, J. Mendes-Lopes, P. Zamora, M. Buljan, D. Grabovickic, "Design of compact optical systems using multichannel

- configurations”, SPIE, Paper No. 9948-21, Novel Optical System Design and Optimization XIX, SPIE Optics + Photonics (August 2016).
5. J. C. Miñano, P. Benítez, J. Mendes-Lopes, B. Narasimhan, **M. Nikolic**, D. Grabovickic, “Recent advances in the SMS method: Diffraction and 3D aplanatism”, SPIE, Paper No. 995-12, Nonimaging Optics: Efficient design for illumination and solar concentration XIII, SPIE Optics + Photonics (August 2016).
 6. J. C. Miñano, P. Benitez, **M. Nikolic**, B. Narasimhan, J. Mendes-Lopes, P. Zamora, M. Buljan, D. Grabovickic, “Compact Optics in Imaging and Nonimaging applications,” ODF ’16 Optical Simposia, Oberkochen, Germany (March 2016).
 7. **M. Nikolic (presenter)**, P. Benitez, J. C. Miñano, D. Grabovickic, J. Liu, B. Narasimhan, M. Buljan, “Optical design through optimization using freeform orthogonal polynomials for rectangular apertures”, Proc. SPIE 9626, Optical System Design 2015: Optical Design and Engineering VI, 96260V (23 September 2015).
 8. B. Narasimhan, P. Benitez, D. Grabovickic, J.C. Miñano, **M. Nikolic**, J. Infante, “Design of Freeform aplanatic systems”, Proc. SPIE 9629, Optical System Design 2015: Illumination Optics IV, 96290C (23 September 2015).
 9. P. Benitez, J. C. Miñano, J. Liu, B. Narasimhan, **M. Nikolic**, J. Ye, “Freeforms: how many degrees of freedom do they provide”, SPIE Optical System Design 2015, Jena, Germany (23 September 2015), **Plenary talk**
 10. J. C. Miñano, P. Benitez, J. Liu, B. Narasimhan, **M. Nikolic**, J. Ye, “On the degrees of freedom of freeform Optics,” in Imaging and Applied Optics 2015, OSA Technical Digest (online) (Optical Society of America, 2015), paper FM2B.1.
 11. **M. Nikolic**, P. Benitez, J. C. Miñano, D. Grabovickic, J. Liu, B. Narasimhan, M. Buljan, “Comparative analysis of optimization with freeform orthogonal polynomials for rectangular apertures”, SPIE, Paper No. 9579-16 (August 2015)
 12. B. Narasimhan, P. Benitez, D. Grabovickic, J.C. Miñano, **M. Nikolic**, J. Infante, “Freeform aplanatic concentrators” Proc. SPIE 9572, Nonimaging Optics: Efficient Design for Illumination and Solar Concentration XII, 95720M (25 August 2015).
 13. B. Narasimhan, P. Benitez, D. Grabovickic, J.C. Miñano, **M. Nikolic**, J. Infante, “Design of three freeform mirror aplanat”, Proc. SPIE 9579, Novel Optical System Design and Optimization XVIII, 95790K (3 September 2015).
 14. J. C. Miñano, J. Liu, P. Benitez, **M. Nikolic**, B. Narasimhan, “Anastigmatic Imaging with unconstrained object to image mapping, “ in Classical Optics 2014, OSA Technical Digest (online) (Optical Society of America, 2014), paper ITh3A.1.
 15. P. Benitez, J. C. Miñano, **M. Nikolic (presenter)**, J. Liu, J. Infante, F. Duerr, “Conditions for perfect focusing multiple point sources with the SMS design method” Proc. SPIE 9191,

Nonimaging Optics: Efficient Design for Illumination and Solar Concentration XI, 919102 (17 September 2014). **Invited Talk.**

16. P. Benitez, J. C. Miñano, **M. Nikolic**, L. Jiayao, J. M. Infante, F. Duerr, "Sharp imaging of Multiple object Points: How and Why," in *Renewable Energy and the Environment*, OSA Technical Digest (online) (Optical Society of America, 2013), paper FW3A.2.

Poster presentations

1. **M. Nikolic**, B. Narasimhan, P. Benitez, J. C. Miñano, "Analytical solution of an afocal two freeform mirror design problem," ES Optical design meeting, Madrid, Spain (23rd of March 2017).
2. B. Narasimhan, **M. Nikolic**, P. Benitez, J. C. Miñano, "Freeform aplanatic systems," ES Optical design meeting, Madrid, Spain (23rd of March 2017).
3. **M. Nikolic**, "Anastigmatic ophthalmic lens design", UK Optical design meeting, Harwell Oxford, UK (15th of September 2016).