Optical Design for Far-field Light Shaping

A thesis submitted for the degree of

doctor rerum naturalium (Dr. rer. nat.)

Author:

Liangxin Yang

Institute of Applied Physics

Friedrich-Schiller-Universität Jena

Submitted to:

the Council of Faculty of Physics and Astronomy,
Friedrich-Schiller-Universität Jena

Supervisor: Prof. Dr. Frank Wyrowski, Friedrich-Schiller-Universität Jena
Reviewer: Prof. Dr. Alois Herkommer, Universität Stuttgart
Reviewer: Prof. Dr. Jari Turunen, University of Eastern Finland
Date of the Disputation: 22 September 2021
Dissertation, Friedrich-Schiller-Universität Jena, 2021

I would like to dedicate this thesis to my loving parents ...

Abstract

The design of an optical element for spatial energy redistribution is a fundamental problem of light shaping. Different solutions for far-field light shaping are reviewed. For those design algorithms based on geometric optics, typically, a mapping between the irradiance distribution of the input source and the target signal is assumed, such as the "ray mapping method" for freeform surface design. However, the validity of the mapping assumption is rarely discussed. In this thesis, the validity is analyzed from a physical-optics point of view. The light-shaping system is modeled under the framework of field tracing. It is revealed that the mapping assumption is true only when all the operators in the system modeling are pointwise operators. Examples of freeform surface design by ray mapping method are presented. Field tracing techniques provide a tool to investigate the validity of the design.

With the physical-optics modeling techniques, a design strategy is provided that starts from designing the functional embodiment of the light-shaping element and continues with the structure embodiment design based on the result from the previous step. In the functional design, an inverse of the modeling techniques is applied, and the light-shaping problem is reduced to a Fourier pair synthesis process. For synthesizing the Fourier pair, a mapping-type algorithm is introduced to overcome the stagnation issue of IFTA in the homeomorphic case. The mapping function is solved in Parseval's equation with the mathematical model of the L^2 Monge–Kantorovich problem applied. By the obtained mapping between the fields of a Fourier pair, an output wavefront phase can be achieved by the stationary phase method in a single integration step, rather than an iterative way like the IFTA. In addition, compared to the phase retrieval method based on the mapping between irradiance distributions in the spatial domain, the solution of the mapping between the field of the Fourier pair is integrable; therefore, a smooth output wavefront phase can be achieved. If the homeomorphic situation does not exist in the system, the mapping-type algorithm cannot hold its accuracy. However, IFTA can then be applied because the stagnation issue does not happen now. Moreover, the result from the mapping-type method is a well-suited initial guess for the IFTA. After retrieving the output wavefront phase, the functional embodiment is nothing other than a wavefront phase response (WPR) function.

The structural design of light-shaping elements is developed with the obtained WPR function, or more directly with the output wavefront phase. The design of both a holographic optical element (HOE) and a freeform lens for light shaping is demonstrated. The algorithms are based on their physical models. The local linear grating approximation (LLGA) is addressed for the HOE, and the local plane interface approximation (LPIA) is applied for the freeform lens.

The element function of the HOE design is the same as the WPR function. The period of the local gratings is then derived from its element function. A Gaussian-to-Top-hat shaping task is taken as an example for demonstrating the algorithm. A hybrid component by adding a curved surface to the HOE is suggested to reduce the grating effects from the HOE.

For the design of a freeform lens, the algorithm starts with an initial surface profile. The retrieval of the output wavefront phase and the construction of the freeform surface alternatively proceed until a proper freeform surface is found. The algorithm has no restriction about the input wavefront and the shape of the predefined surface. Moreover, the freeform surface can be designed flexibly as either the front or back surface of the lens. Numerical examples are provided for illustrating the effectiveness of the algorithm.

Zusammenfassung

Das Design von optischen Elementen für die Umformung der spatialen Energieverteilung des Lichtes ist eine der fundamentalen Aufgabenstellungen im Bereich der „Lichtformung“. Verschiedene Lösungsagorithmen zur Formung des Fernfeldes werden untersucht. Typischerweise werden diese auf geometrische Optik basierende Designalgorithmen bijektive Zuordnungen zwischen den Irradianzen des Eingangfeldes und des Ausgangsfeldes angenommen, z. B. die „Ray-Mapping“-Methode für das Design von Freiformoberflächen. Jedoch wird die Gültigkeit dieser Annahmen selten diskutiert. In dieser Arbeit wird die Gültigkeit dieser Zuordnung unter Zuhilfenahme der physikalischen Optik analysiert. Das System zur Formung des Lichtes wird im Rahmen des „Field Tracing“ modelliert und ergab das Ergebnis, dass die Annahme der bijektiven Zuordnung nur gültig ist, wenn alle Operatoren innerhalb des Systems punktweise definiert sind. Beispieldesigns von Freiform Oberflächen, die durch die „Ray-Mapping“-Methode modelliert wurden, werden demonstriert und mit Hilfe von „Field Tracing“ auf ihre Gültigkeit hin überprüft.

Durch die Modellierungstechnicken der physikalischen Optik kann eine Designstrategie gegeben werden, die die funktionale Darstellung eines Elements der Lichtformung als Grundlage nimmt und anschließend das Design der Struktur basierend auf dem Ergebnis der vorangegangen Rechnungen weiterführt. Im funktionalen Design wird ein inverser Modellierungsansatz verwendet und das Problem der Lichtformung zu einem „Fourier Pair Synthesis“ - Prozess reduziert. Zur Erzeugung des Fourier-Paares wird der iterative Fouriertransformationsalgorihtmus (IFTA) erweitert, indem ein Algorithmus basierend auf bijektiven Zuordnungen eingeführt wird, um Stagnationsprobleme für die Algorithmen bei Homeomorphismen zu lösen. Die Funktion der Zuordnung wird mathematisch als Lösung der Parseval-Gleichung durch Zuhilfenahme des L^2 Monge–Kantorovich Problems gerechnet. Durch die Zuordnung zwischen den Feldern des Fourier-Paares kann die Wellenfront des Ausgangsfeldes berechnet werden durch die Methode der statischen Phasen, welche eine einfache Integration darstellt im Vergleich zum iterativen Vorgehen des IFTAs. Außerdem, im Vergleich zum Phasenrekonstruktionsmethode basierend auf der bijektiven Zuordnung zwischen den Irradianzverteilungen in der räumlichen Domäne, ist die Zuordnung zwischen den Feldern des Fourier-Paars integrierbar, wodurch eine glatte Phase der Wellenfront im

Ausgangsfeld erzielt werden kann. Weiterhin wird gezeigt, dass die bijektiven Zuordnungen nicht akkurat sind, wenn kein Homeomorphismus im System existiert.

Andererseits kann in diesen Fällen der IFTA verwendet werden, da Stagnationsprobleme dann nicht auftauchen. Zudem sind die Ergebnisse der Methode der bijektiven Zuordnung ein guter Startpunkt für den IFTA. Nach der Rekonstruktion der Phase ist die funktionale Darstellung der Komponente nichts anderes als die Responsefunktion der Wellenfrontphase (WPR).

Das Design der Struktur der Lichtformungs-Komponente wird entwickelt mithilfe der erhaltenen WPR-Funktion, bzw direkt durch die Wellenfrontphase des Ausgangsfeldes. Das Design sowohl von einem holographischen Element (HOE) und von einer Freiform-Linse wird demonstriert. Aufgabe der Systeme ist die Umwandlung einer Gausverteilung in eine „Top-Hat“-Verteilung. Die Algorithmen basieren auf den physikalischen Modellen der „Local Linear Grating Approximation“ (LLGA) für das HOE und der „Local Plane Interface Approximation“ (LPIA) für die Freiform-Linse.

Im Falle des HOE wird die elementare Funktion der Komponente durch die WPR beschrieben. Die Periode der lokalen Gitter wird dann durch diese Funktion berechnet. Als Beispiel für den Algorithmus wird die Umwandlung einer Gausverteilung in einen „Top-Hat“ gezeigt. Eine hybride Komponente, bei der gekrümmte Oberflächen zum HOE hinzugefügt werden, wird vorgeschlagen, um das Problem der Amplitudenmodulation im HOE zu lösen.

Für das Design der Freiform-Linse startet der Algorithmus mit einem initialen Oberflächenprofil und bestimmt anschließend jeweils abwechselnd die rekonstruierte Wellenfrontphase und Freiformoberfläche, bis eine akzeptable Lösung gefunden wird. Es gibt keine Restriktion in Bezug auf die Eingangswellengront und die Form der vordefinierten Oberfläche. Zudem kann die Freiform-Oberfläche flexible als entweder Vor- oder Rückseite der Linse designt werden. Numerische Simulationen werden gezeigt, um die Effizienz des Algorithmus zu demonstrieren.

Table of contents

Chapter 1

Introduction

The applications of light shaping have influenced people's daily life in a way we cannot imagine decades ago. The development of innovative technology enables the fabrication of precision machinery for optical devices used in industry and in consumer products. Light shaping is used in information technology and telecommunications, health care and medicine, optical sensing, lighting and energy control. Harnessing the physical properties of light is the essential issue in all these applications, and different requirements lead to the key challenges for the designers of optical technology.

Optical design can be roughly cataloged into two branches, imaging optics and non-imaging optics[1], based on its applications. In imaging optics, people try to tackle the problem of how to map the point sources to the image points as exactly as possible. In non-imaging optics, the focus is on the transmission of energy from the source to the target. The work of this thesis mainly focuses on the topic of non-imaging, or light shaping, where the energy distribution of the light is controlled.

1.1 Light-shaping problem

The design of light-shaping optics can be formulated as follows: given an input (a light source) and a target (a prescribed energy distribution), one or multiple optical elements are to be designed so that the light from the source is redistributed to produce the prescribed energy. In general, the target energy distribution can be defined on a vertical plane, on a tilted plane or even on a curved surface.

Light-shaping design is not trivial, and no general workflow has been developed for all kind of optical systems. Traditional design methods for optical elements, such as lens design, is usually a forward method. Modeling techniques are applied to simulate the optical system, for example, the Monte–Carlo ray-tracing method. Merit functions are set for the

parameters of the optical element. Optimization algorithms are then applied to search for the element profile that minimize the merit functions. However, using the forward method for the light-shaping design is not an effective strategy. The parameterization of optical elements for light shaping, like diffractive optical elements and freeform surfaces, is usually far from trivial. Precise optimization may result in a slow procedure because millions of rays and a large number of merit functions are needed. Because of the huge quantity of the parameters, the optimization could lead to a local minimum of the merit function or even result in unphysical solutions.

Therefore, an inverse design strategy is typically applied in light shaping, where the algorithm starts from the target. The target physical quantity, such as irradiance or radiant intensity, is either derived as a function of electromagnetic fields or ray bundles, which are traced back to the position where the optical element is to be designed. With the information from the source and the target, the element can be determined, sometimes iteratively. The inverse design strategy is used in this thesis, although the parametric optimization is not discarded but brought at the end of the design procedure. The result from the inverse design normally is a good initial guess for the parametric optimization step.

1.2 Possible solutions for light shaping

Depending on the application requirement, different kinds of optical elements can be used for light shaping, such as holography optical element (HOE) , freeform surface, diffusers, metasurface or hybrid components with both diffractive and refractive elements. In the following, typical design methods for HOE and freeform surface for light shaping are reviewed.

Holography optical element In the book "Optical Holography: Materials, Theory and Applications" [2], Prof. Blanche wrote that,

> *Holograms are very well known for the awe inspiring 3D images they can recreate. But they can also be used to generate arbitrary wavefronts. Examples of such wavefronts are focalization exactly like a lens, or reflection exactly like a mirror. The difference of the hologram from the original element (lens or mirror) is that, in both cases, diffraction is involved, not reflection or refraction . That type of hologram, called holographic optical element, is found in optical setups where for reason of space, weight, size, complexity, or when it is not possible to use classical optical elements. Some examples include combiner in*

head-up display, dispersion grating in spectrometers, or spot array generators for cameras and laser pointers.

HOE is one of the commonly used solutions for light shaping indeed due to its ability to generate arbitrary wavefronts. In the beginning, HOEs can only be acquired optically and recorded on photographic film as hardcopies, for which the light from the source and the target is interfered. However, due to the development of electronics and computer technology, an HOE can be formulated numerically from a virtual object model (2D or 3D) with the help of physical-optics simulations for light propagation, known as the computer-generated hologram (CGH). The modeling and design of the HOEs in this thesis mainly concentrates on the CGH.

Although the CGH can eliminate the bulky setups required for optically recorded HOEs and is convenient in terms of processing or storage, the design of CGH usually faces a huge challenge in terms of the amount of calculation required.

One way to generate the CGH is to imitate the optical recording process such that the target pattern/object is represented in various electric fields, usually as independent point sources at different positions, and each electric field is propagated backwards to the element plane to form a hologram. The elementary holograms from different point sources are then overlapped to generate the entire CGH [3, 4]. Various approaches are proposed for the calculation of the propagation of each point source for CGH computation based on the point-light-source model. The propagation can be simulated using the inverse function of the Kirchhoff diffraction integral [5] to compute CGHs in their most rigorous form. However, in general cases, an exact solution for Kirchhoff diffraction integral equation cannot be found. Even in the cases that the solution can be achieved, rather expensive numerical efforts are demanded. Therefore, to overcome the computational bottle neck, approximated methods are introduced to get around the problem. Simpler expressions like the Fresnel integral and the Fraunhofer integral can be used as substitutes [6]. The coherent ray trace method [7, 8] is also a simple and widely used technique that yields excellent fidelity in the reconstructed images of three-dimensional objects with simple geometries. The look-up table (LUT) approaches are also proposed to reduce the computational time [9–13]. The LUT method utilizes pre-calculated data to replace complicated calculations. Each complex amplitude distribution of the point light source is pre-calculated and stored. Then, the CGH can be easily obtained by calling the pre-calculated data. However, this method requires a large amount of memory and has high requirements for the computer hardware [12].

If the target pattern/object is considered as an entire electric field, the construction of the CGH is mainly a phase-retrieval problem because the character of the HOE normally is an element with phase-only modulations under control, for example, liquid crystal on

silicon (LCoS) or spatial light modulator (SLM). In general, the complete information of the target complex field (amplitude and phase) is not captured from measurement because typically, the sensor for measurement only provides the energy of light. An effective way to obtain a required output phase to reconstruct the target pattern/object is to use an iterative computation such as the Gerchberg–Saxton (GS) algorithm [14]. By iteratively applying the Fourier transform (FT) and the inverse Fourier transform (IFT) and setting constraints at the space-domain and its corresponding Fourier domain, the final output phase can be obtained when the error between the target field and the iteratively reconstructed field converges. Based on the idea of the GS algorithm, different phase retrieval algorithms are developed for the design of CGH for light shaping [15–26]. Algorithms are developed for speckles reduction, efficiency enhancement, or to achieve the diffraction pattern on the plane in the Fresnel field zone or even in 3D space. Geometric-based algorithms are also proposed to solve the required output phase without an iterative procedure [27–34]. The algorithms assume a one-to-one map of the irradiance between the source and the target, and the stationary phase method is addressed to construct the phase from the mapping function [34].

Freeform surface Another common solution for light shaping is elements with freeform surfaces. For general illumination, the majority of the system is beyond rotational symmetric. Therefore, the application of freeform optics introduces more freedom for the design, which may also change bulky optics systems into compact ones. The recent progress of high-precision fabrication technology also expedites the application of freeform surface in modern optical systems [35].

The design of freeform surfaces for light-shaping applications can be divided into two categories according to the different types of light source [36]. One is for the light source with zero étendue (single spherical wave or plane wave), and the other one is for the extended light source.

The literature describes, three main types of method for the design of freeform surfaces regarding the zero étendue light source [36]: the Monge–Ampère equation method [37–45], the “ray mapping method” [46–57], and the supporting quadric method (SQM) [58–62].

In the geometric-optics assumption, a one-to-one mapping is assumed for the source and target energy distribution. In most cases, the design problem can be formulated as a nonlinear partial differential equation (PDE) of the elliptic type, the elliptic Monge–Ampère differential equation [37, 63]. The design method is then based on numerically finding the solution of the equation [39, 64], where the equation is discretized with the finite difference scheme for the derivative terms, and Newton’s method is applied to solve the discretized

nonlinear equations. Depending on the complexity of the problem, the number of equations in the system can reach tens of thousands, which leads to a heavy computational effort.

In contrast to the Monge–Ampère equation method, the ray mapping method is relatively more practical [65]. Similarly, with the geometric-optics assumption, the method is separated into two main steps. The key step of this method is to find a mapping between the source and the target with energy conservation. The calculation for the mapping is usually solved as a variational problem within the framework of the L^2 Monge–Kantorovich problem, also known as the "optimal mass transport" (OMT) problem [66]. Numerical algorithms are proposed to solve the problem [67–70], resulting in a curl-free mapping between two coordinates from the source and the target. The second step of the ray mapping method is to find the corresponding target surface that realizes this mapping. Usually, the surface is computed using Snell's law, either in a point-by-point scheme [51, 71], or by performing a least-squares optimization with the calculated normal vector field [48, 50, 52]. The critical problem for the ray mapping method is that, although the mapping from the OMT solution is curl-free, the normals of the surface derived from the mapping are not necessarily integrable due to the non-linearity introduced by Snell's law [52, 72]. To obtain the target surface, an optimization process is performed either directly applied on the surface height value [52], or to include the integrability condition of the surface into the ray-mapping calculation step [57, 73].

The SQM, as the name indicates, attempts to discretize the design problem to a local imaging problem where a quadric surface is searched to solve the local problem. The target energy distribution is pixelated, and one sampling point is used to indicate each pixel. Depending on the discretized problem, different quadric surfaces, such as paraboloids [58], ellipsoids [61], or hyperboloids [59, 62], are used as a surface patch, and the final freeform surface is composed by these surface patches. The number of sample points should be set to a huge number to avoid the discontinuity between the patches.

For the design problem that the light source is beyond the zero étendue approximation, the above-mentioned methods become limited. The simultaneous multiple surfaces (SMS) method is the most successful design method for tackling the extended source illumination problem [74–78]. The basic idea of the SMS method is to characterize the source into two [75] or three [78] wavefronts, and correspondingly, specify the same number of wavefronts to achieve the target pattern. After the wavefront pair is set, freeform surfaces are constructed locally to couple the wavefronts.

Because the HOE for light shaping modulates the incident by means of diffraction, the element is sensitive to the wavelength of the incident. Besides, additional diffraction orders will introduce stray light to the target. At the same time, the freeform surface, as a refractive

component, is relatively insensitive to the wavelength: only the dispersion of the material plays the role. The smoothness of the freeform surface is also an advantage to avoid stray light. However, the structure of the freeform surface is bulky compared to the HOE, especially for non-paraxial light-shaping cases. Therefore, for certain applications, both solutions are determined due to their special characters, and variant components are also introduced to balance their drawbacks [79, 80].

1.3 From functional design to structural design

From the reference listed above, the algorithms for different types of solutions, for example, the diffractive ones or the refractive ones, are quite different from each other. Typical design algorithms for the diffractive elements, like the GS algorithm, are based on a Fourier relation of electric fields. Nevertheless, typical design algorithms for the freeform surface, like the Monge–Ampère equation method or the ray mapping method, are based on a conservation relation of the irradiance distribution. We ask the question that, can the shaping problem be tackled in a systematic way, so that both the diffractive and refractive elements can be designed in a same process.

From a physical-optics point of view, instead of finding a solution directly on the construction of optical element, the optical system is analyzed and a question about the functionality of the required optical elements is asked first. To this end, the system should be designed by a systematic investigation into how to obtain system information using a forward-propagated input field and an inverse propagation of the desired output field.

This approach is done in a first design step in the functional embodiment of the system that only the function of the required components and their position in the system is considered. In this step, the structure of the optical elements and its physical effect are not considered. Therefore, the optical element is not restricted to either as a refractive one or a diffractive one. In the second step, the structural design searches for an optical element that can realize the functionality, along with the probable fabrication constraints. This two-step approach is usually favored in diffractive optics, especially for paraxial beam shaping [31]. Typically, once the element function is obtained, the thin element approximation (TEA) method [5] is applied to design the element structure. In this thesis, the approach is generalized so that it is not restricted to the paraxial situation.

We learn from the design method for diffractive elements, that a Fourier pair synthesis is mainly required for the far-field light-shaping system; we also learn from the geometric-optics-based design methods for refractive elements, that a homeomorphic (one-to-one map) assumption is usually applied in the algorithm. How can these two different approaches

relate to each other? Can the approaches be unified in a more general way for the functional and structural design? Following the questions, the major contributions of this thesis are as follows:

a) The far-field light-shaping system is analyzed with the field tracing techniques. It is found that if all the modeling operators in the system are pointwise operators, a homeomorphism is established for the electric fields through the system, from the input plane to the target plane. This assumption is typically used in those geometric-optics-based algorithms for designing a freeform surface. The field tracing technique is used as a tool to investigate the validity of the design with the homeomorphic assumption. The homeomorphism of the Fourier transform for the electric field behind the optical element is noted as the critical point of the validity.

b) Using the inverse of the field-tracing modeling technique, the far-field light-shaping problem mainly required a Fourier pair synthesis, by which the required output phase is retrieved. For synthesizing the Fourier pair, the iterative Fourier transform algorithm (IFTA) is extended, such that a mapping-type algorithm is introduced to overcome the stagnation issue for IFTA in the homeomorphic case. The mapping function is solved in Parseval's equation with the mathematical model, that is, the L^2 Monge–Kantorovich problem.

c) By the mapping-type Fourier pair synthesis strategy, the smooth output wavefront phase is achieved in a single integration step using the obtained mapping function, rather than an iterative way. Moreover, the existence and character of the mapping function between the electric field in two domains guarantee the integrability of the output wavefront phase, while its integrability is usually a problem for algorithms based on the mapping of irradiance distribution in the spatial domain.

d) The algorithm from Prins [70] is adapted to solve a mapping from a Jacobian equation. However, the algorithm is modified with an additional integral step so that the algorithm converges to an integrable solution faster. Moreover, a dummy uniform function is introduced for tackling the mapping between two functions with arbitrary support shapes. The mapping between two functions is then obtained by the combination of two separate maps. The integrability of the combined mapping is also discussed mathematically and optimized.

e) With the obtained output wavefront phase, an HOE can be designed for the shaping task. However, due to its amplitude modulation, the designed HOE may introduce errors to the target irradiance distribution. A hybrid component is suggested with an

additional surface that is designed for reducing the high-frequency feature of the HOE. The surface is selected such that an approximated wavefront phase to the required one is introduced in front of the HOE, whereby the local grating periods of the HOE are relaxed.

f) The obtained output wavefront phase is also applied for the design of a freeform lens. An iterative algorithm is proposed for the surface design, where the algorithm alternatively performs the functional design and the structural design until a proper surface is found. The algorithm is developed in a general way so that it can be applied for any type of wavefront of the input field. The freeform lens is designed with a predefined surface that can be set either as the front or the back surface of the lens, with an arbitrary shape.

1.4 Synopsis of the thesis

The synopsis of the thesis is provided as followed. In **Chapter 2**, the basic principle of the field tracing techniques are introduced. The homeomorphic situation in the light-shaping system is discussed in the framework of field tracing. After the theory, the validity of the ray mapping algorithm that includes the homeomorphic assumption for the design of the freeform surface is investigated via physical-optics analysis, with examples demonstrated.

In **Chapter 3**, the light-shaping problem is analyzed and the design of functional embodiment is presented. The IFTA is reviewed, and its stagnation problem in the homeomorphic situation is mathematically illustrated. In such a situation, the mapping-type Fourier pair synthesis is proposed to retrieve the required output wavefront phase. The existence of a solution with its integrable characteristic is addressed and discussed through numerical examples. The mapping is solved by applying the L^2 Monge–Kantorovich mathematical model to Parseval's equation. The algorithm for solving the mapping is presented in detail.

Based on the result from functional design, in **Chapter 4**, the design of the structure of the HOE and the freeform surface is developed. The physical modeling techniques of both elements are presented. The design algorithms are provided to construct the structure for the HOE and the freeform surface. Different light-shaping tasks are demonstrated to illustrate the design. The extra physical effect of different structures is discussed and compensated with the proposed solution.

Finally, **Chapter 5** summarizes the work in this thesis and gives a brief outlook for the possible research topics and extended applications of the work.

Chapter 2

Light shaping from a physical-optics point of view

The modeling of the optical system is important because the analysis and the design should be based on reliable modeling techniques. In this chapter, the fundamental principle of a physical-optics modeling technique, field tracing, is introduced for system modeling. The basic model of a light-shaping system is discussed based on the field tracing technique. Different algorithms can be selected to perform the Fourier transforms in field tracing. One is the homeomorphic Fourier transform (HFT) algorithm, which can be accurately applied under the stationary phase condition and establishes a homeomorphism between the electric fields in the free space propagation step. The homeomorphic situation is usually assumed in the geometric-optics-based design algorithms. However, the validity of the designed result obtained based on the assumption must be verified via physical-optics modeling. The geometric-optics-based design algorithms and examples of their application are investigated under the framework of field tracing.

2.1 Physical-optics modeling techniques

In physical optics, light is modeled as an electromagnetic field that is mathematically governed by Maxwell's equations. The physical nature of the electromagnetic field is expressed in different domains: the time domain ($t-$domain), the temporal frequency domain ($\omega-$domain), the spatial domain ($\boldsymbol{\rho}-$domain) and the spatial frequency domain ($\boldsymbol{\kappa}-$domain), where $\boldsymbol{\rho} \coloneqq (x, y)$ is the 2D transversal spatial coordinate and $\boldsymbol{\kappa} \coloneqq (k_x, k_y)$ is the 2D

transversal spatial frequency coordinate. For example, Maxwell's equations, considered in the $(\boldsymbol{\rho}, \omega)$ domain where the electromagnetic fields are defined on 2D planes, are written as:

$$\boldsymbol{\nabla} \times \boldsymbol{E}(\boldsymbol{\rho}, \omega) = \mathrm{i}\omega \boldsymbol{B}(\boldsymbol{\rho}, \omega), \tag{2.1a}$$

$$\boldsymbol{\nabla} \times \boldsymbol{H}(\boldsymbol{\rho}, \omega) = -\mathrm{i}\omega \boldsymbol{D}(\boldsymbol{\rho}, \omega), \tag{2.1b}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{D}(\boldsymbol{\rho}, \omega) = 0, \tag{2.1c}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B}(\boldsymbol{\rho}, \omega) = 0, \tag{2.1d}$$

without considering the free charges and currents.

In general, Maxwell's equations are solved with different boundary conditions to model an optical system. Some advanced Maxwell solvers, like the finite-difference time-domain (FDTD), the finite element method (FEM), and Fourier modal method (FMM), can provide rigorous solutions for the equations in different domains. However, these solvers usually suffer from expensive numerical efforts, particularly when the feature sizes of the optical element become larger or simulation along the system with long distance. Therefore, tailoring solvers for different subdomains of the system in a wise way can provide necessary physical information for the system on the one hand, and on the other hand, perform the simulation in a fast manner.

Field tracing is a technique that combines various modeling methods for different subdomains of an optical system [81, 82]. Certainly, each selected modeling method must rely on electromagnetic field representation. The resulting field of each subdomain is then interconnected, sequentially or non-sequentially, resulting in a total solution for the whole optical system. In general, the modeling methods are often applied for the field represented in different domains. With sufficient accuracy, the domain choice depends on where the modeling method is more efficient; that is also one of the basic ideas of fast physical optics.

A field tracing diagram shown in Fig. 2.1 indicates an example of system modeling with the field tracing techniques. The diagram shows a sequential modeling example, whereas the field tracing techniques, in general, can be performed sequentially or non-sequentially. The path in the figure illustrates how the modeling techniques are applied through the system. The four parallel lines indicate the four different domains where the electromagnetic field are defined; for example, the second line is the spatial and temporal frequency domain, $(\boldsymbol{\rho}, \omega)$ domain. The over-tilde "˜" symbol of the operators indicates it is defined in the $\boldsymbol{\kappa}-$domain, which distinguishes it from the operators in the $\boldsymbol{\rho}-$domain.

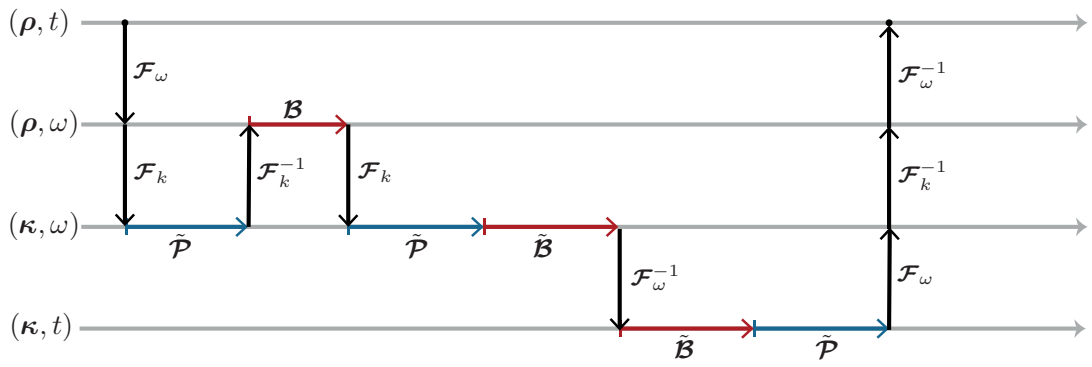

Fig. 2.1 General field tracing diagram for illustrating the field tracing techniques.

$\mathcal{B}$ operator In the field tracing diagram, $\mathcal{B}$ is an operator of a certain optical component. It realizes the functionality of the component and interacts with the fields passing through it. Normally, the operator is set in the domain where the functionality of the component is defined such that it minimizes its calculation effort. For example, the $\mathcal{B}$ operator for an aperture in the system is usually defined in the $\boldsymbol{\rho}-$domain, because it is only a product calculation with the coming field; for a grating, the $\mathcal{B}$ operator is more efficient in the $\boldsymbol{\kappa}-$domain because for each diffraction order, $\mathcal{B}$ can be a pointwise operator for the field in the $\boldsymbol{\kappa}-$domain.

The $\mathcal{B}$ operator is in a matrix form so that the vectorial effect of the optical component on the field is taken into account. Considering an electric field, with its field vector as $\boldsymbol{E}^{\text{in}} = (E_x^{\text{in}}, E_y^{\text{in}}, E_z^{\text{in}})$, passing through an optical component, the corresponding output field is $\boldsymbol{E}^{\text{out}} = (E_x^{\text{out}}, E_y^{\text{out}}, E_z^{\text{out}})$. The relation between the input and output field can be described with the $\mathcal{B}$ operator explicitly as:

$$\begin{bmatrix} E_x^{\text{out}} \\ E_y^{\text{out}} \\ E_z^{\text{out}} \end{bmatrix} = \begin{bmatrix} \mathcal{B}_{xx} & \mathcal{B}_{xy} & \mathcal{B}_{xz} \\ \mathcal{B}_{yx} & \mathcal{B}_{yy} & \mathcal{B}_{yz} \\ \mathcal{B}_{zx} & \mathcal{B}_{zy} & \mathcal{B}_{zz} \end{bmatrix} \begin{bmatrix} E_x^{\text{in}} \\ E_y^{\text{in}} \\ E_z^{\text{in}} \end{bmatrix} \tag{2.2}$$

In general, a component should include $\mathcal{B}$ operators for different channels, to represent its response of an input field from the front or back of the component in a transmission or reflection manner.

$\tilde{\mathcal{P}}$ operator Between different regions in the optical system is the free space, which is defined in [81] as homogeneous and isotropic media. The $\tilde{\mathcal{P}}$ operator in Fig. 2.1 indicates

the operator for the free space propagation of the fields. In field tracing techniques, the $\tilde{\boldsymbol{\mathcal{P}}}$ operator is usually applied in the $\boldsymbol{\kappa}-$domain because there the operator can be simply performed as a product calculation. For the field in the $\boldsymbol{\rho}-$domain, by the Fourier transform, the field in the $\boldsymbol{\kappa}-$domain is actually the spectrum of its plane wave components. With the theory of spectrum of plane wave (SPW) method for the free space propagation, each plane wave in the $\boldsymbol{\kappa}-$domain transforms as it propagates from the input plane to the target plane. Mathematically, the process is written as:

$$\begin{aligned}\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}) &= \tilde{\boldsymbol{\mathcal{P}}}\tilde{\boldsymbol{E}}^{\mathrm{in}}(\boldsymbol{\kappa}) \\ &= \exp(\mathrm{i}\check{k}_z(\boldsymbol{\kappa})\Delta z)\cdot \tilde{\boldsymbol{E}}^{\mathrm{in}}(\boldsymbol{\kappa})\end{aligned} \tag{2.3}$$

where $\check{k}_z(\boldsymbol{\kappa}) = \sqrt{{k_0}^2\check{n}^2 - \boldsymbol{\kappa}^2}$ is the $z-$component of the wave vector in the corresponding medium, and Δz the propagation distance. Therefore, performing the free space in the $\boldsymbol{\kappa}-$domain is a product calculation and can be done pointwise; that saves much numerical effort compared to propagation operator with integrals calculations in $\boldsymbol{\rho}-$domain.

Fourier transform The Fourier transforms are the operators that switch the field in different domains when it is required. $\boldsymbol{\mathcal{F}}_\omega$ represents the Fourier transform between the $t-$domain and the $\omega-$domain, and $\boldsymbol{\mathcal{F}}_k$ represents the one between the $\boldsymbol{\rho}-$ and the $\boldsymbol{\kappa}-$domain. The inverse of the transforms are $\boldsymbol{\mathcal{F}}_\omega^{-1}$ and $\boldsymbol{\mathcal{F}}_k^{-1}$, respectively.

Normally, the fast Fourier transform (FFT) method is applied for the Fourier transform process. The development of FFT reduces the numerical effort of computing the discrete Fourier transform (DFT) from $\mathcal{O}\left(N^2\right)$ to $\mathcal{O}(N\log N)$, where N is the sampling points of the function. However, sampling a complex field can still yield a large sampling number of N, making the FFT impractical. Several rigorous solutions have been proposed to overcome the problem: by handling a quadratic phase term, the Fourier transform efficiency can be improved; examples include the chirped-z transform [83] or the semi-analytical Fourier transform [84].

Moreover, when the optical system appears in non-paraxial situations, or more precisely, when the stationary phase condition of a field is satisfied [85], an approximated algorithm in terms of the homeomorphic Fourier transform (HFT) can also be applied for performing the Fourier transform [86, 87]. The HFT is a fast, mathematically pointwise method governed by a mapping between the coordinates in one domain and its Fourier domain.

Therefore, the system modeling tailors the interaction between the field and the optical components into different $\boldsymbol{\mathcal{B}}$ operators in different domains. The resulting fields are then switched to the applied domain for the next operator by the Fourier transform. Different

algorithms enrich the Fourier transform class that flexibly models optical systems. Choosing the right Fourier transform method under a certain circumstance is essential to both accelerate the system modeling and to help to analyze the property of the field, which is useful information for the optical design.

2.2 Homeomorphic situation in light shaping

The design of an optical element for spatial energy redistribution is a fundamental problem of light shaping. One intuitive idea to redistribute the energy of light is to find a mapping between the energy of the source and the target. Therefore, local energy conservation, or in literature, also called the intensity law of geometrical optics [88], is usually assumed in the design algorithms, either for the design of diffractive optics [27, 28, 30, 32, 89] or refractive optics [36, 37, 48, 49, 52]. All the design methods that include the mapping assumption are the geometric-optics-based methods. However, whether the mapping assumption from geometric optics is accurate is still open to question. From a physical-optics point of view, the assumption is valid only if all the operators for the system modeling are pointwise operators because a pointwise operator gives a one-to-one map between the input and output field. In this section, the light-shaping system is analyzed via the field tracing techniques, and the critical point of the mapping relation is revealed.

A basic light-shaping system is illustrated in Fig. 2.2, together with a field tracing diagram that denotes the modeling sequence used in the light path. An optical element is used in the system to shape the input field from the source and achieve the desired energy distribution in the target plane, located in the far-field zone. In fact, for the general case, the target energy can be on any kind of surfaces, such as a curved surface or tilted plane. We use a planar surface perpendicular to the optical axis here as an example to simplify the later discussion.

In Fig. 2.2, $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ is the electric field defined on the plane in front of the optical element. $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ are fields on the plane behind the optical element, in the $\boldsymbol{\rho}-$ and $\boldsymbol{\kappa}-$domains, respectively. $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ and $\tilde{\boldsymbol{E}}^{\mathrm{tar}}(\boldsymbol{\kappa}^{\mathrm{tar}})$ are the fields on the target plane. The field tracing algorithm for the system can be written as:

$$\begin{aligned}\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) &= \boldsymbol{\mathcal{F}}^{-1}\left\{\tilde{\boldsymbol{\mathcal{P}}}\boldsymbol{\mathcal{F}}\left\{\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})\right\}\right\} \\ &= \boldsymbol{\mathcal{F}}^{-1}\left\{\tilde{\boldsymbol{\mathcal{P}}}\boldsymbol{\mathcal{F}}\left\{\boldsymbol{\mathcal{B}}\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})\right\}\right\},\end{aligned} \tag{2.4}$$

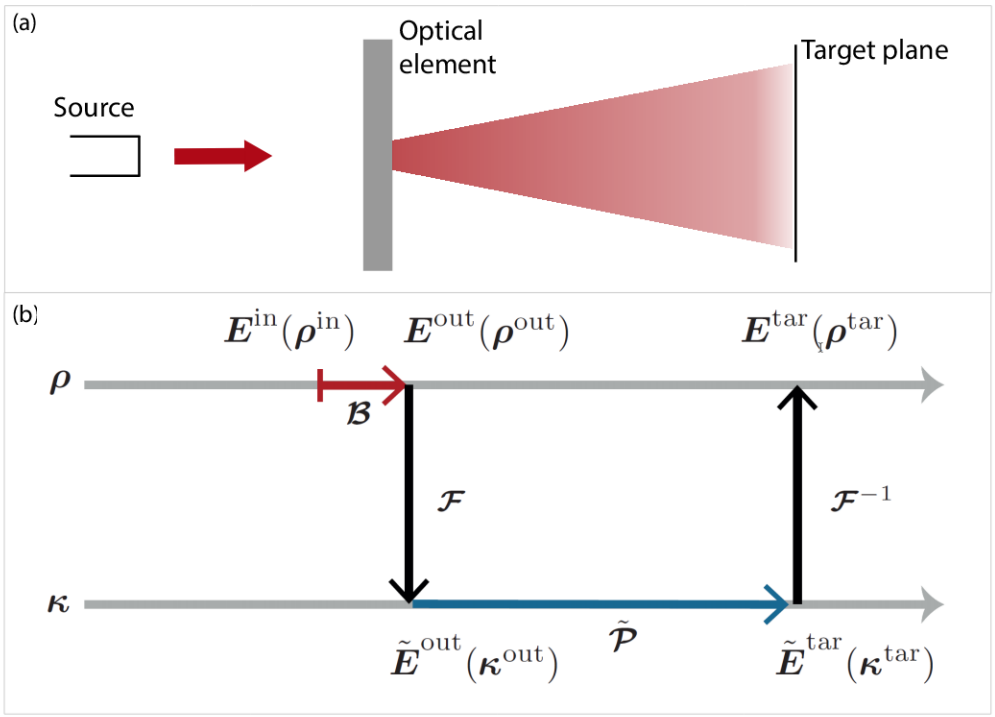

Fig. 2.2 A light-shaping system with a field tracing diagram illustrating the modeling techniques.

Pointwise operators in light-shaping system

For optical elements that the modeling of them is a pointwise operator, their $\mathcal{B}$ operators establish a one-to-one map, also known as a homeomorphism, between the field in front of the optical element and the one behind, $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$, respectively.

For example, the CGH is usually modeled with the local linear grating approximation. Therefore, the $\mathcal{B}$ operators of CGHs are local operators such that the interaction with the input field is considered pointwise for each sampled field value; For paraxial light shaping with DOEs, the DOEs are simply modeled with the thin element approximation (TEA) method [5]. Similarly, the operator of the TEA model $\mathcal{B}^{\mathrm{TEA}}$ is also a phase response function such that the phase values in space are proportional to the local height value of the element; If the optical element is a freeform lens, the local plane interface approximation (LPIA) method [90] is usually applied for modeling the light propagation through both surfaces. Each application of the LPIA operator $\mathcal{B}^{\mathrm{LPIA}}$ contains three sequential steps [90]: 1. a coordinate transformation for the field on the surface from the global coordinate system to the (position-dependent) surface coordinate system; 2. applying the position-dependent Fresnel coefficients to the local field components; 3. a coordinate transformation for the field on the surface from the local coordinate system back to the global one. Therefore, LPIA is also

a pointwise operator for the electric field, which gives homeomorphism behavior between $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ in the $\boldsymbol{\rho}-$domain.

$\tilde{\boldsymbol{\mathcal{P}}}$ in Fig. 2.2 is the free space propagation operator, which is a simple product calculation in the $\boldsymbol{\kappa}$ domain, as shown in Eq. (2.3). Therefore, $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ and $\tilde{\boldsymbol{E}}^{\mathrm{tar}}(\boldsymbol{\kappa}^{\mathrm{tar}})$ also exhibit a mapping relation.

Both the $\boldsymbol{\mathcal{B}}$ and $\tilde{\boldsymbol{\mathcal{P}}}$ operators can be realized by pointwise calculation. Therefore, if the Fourier transforms ($\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{F}}^{-1}$) in Eq. (2.3) can also be performed pointwise, the whole field tracing process is a pointwise algorithm, which builds up a homeomorphism between the input and the target field. Therefore, the Fourier transforms are the critical points that determine whether the homeomorphic condition can be fulfilled.

As mentioned in the last section, different kinds of algorithms can be applied for the Fourier transform in the field tracing. The HFT is the one that is exactly an approximated pointwise Fourier transform. The HFT produces a one-to-one map between the coordinates of the fields expressed in the $\boldsymbol{\rho}-$ domain and in the $\boldsymbol{\kappa}-$ domain [86, 87].

Specifically, similar to the notation in [86], the electric field in the $\boldsymbol{\rho}-$ domain is denoted by

$$E_\ell(\boldsymbol{\rho}) = U_\ell(\boldsymbol{\rho}) \exp(\mathrm{i}\psi(\boldsymbol{\rho})), \tag{2.5}$$

where E_ℓ stands for any element of the set (E_x, E_y, E_z). $U_\ell(\boldsymbol{\rho}) \in \mathbb{C}$ in general. $\psi(\boldsymbol{\rho})$ is the wavefront phase, which is the common smooth part of the phase for all the field components (E_x, E_y, E_z), and $\arg[U_\ell(\boldsymbol{\rho})] = \arg[U_\ell(\boldsymbol{\rho})] - \psi(\boldsymbol{\rho})$ is the residual phase.

With the above in mind, the HFT operation is formulated as follows:

$$\tilde{E}_\ell(\boldsymbol{\kappa}) = \boldsymbol{\mathcal{F}}^{\mathrm{hom}}\{U_\ell(\boldsymbol{\rho})\} = \alpha(\boldsymbol{\kappa}) U_\ell(\boldsymbol{\rho}(\boldsymbol{\kappa})) \exp(\mathrm{i}\tilde{\psi}(\boldsymbol{\kappa})), \tag{2.6}$$

with $\tilde{\psi}(\boldsymbol{\kappa}) = \psi(\boldsymbol{\rho}(\boldsymbol{\kappa})) - \boldsymbol{\kappa} \cdot \boldsymbol{\rho}(\boldsymbol{\kappa})$, and $\alpha(\boldsymbol{\kappa})$, a weight factor that depends on the second derivatives of $\psi(\boldsymbol{\rho})$. The HFT performs the Fourier transform governed by a bijective map between $\boldsymbol{\rho}$ and $\boldsymbol{\kappa}$. Similarly, the inverse HFT is also performed by the same one-to-one map of the coordinates of the fields between the two domains [86, 87]. The bijective map is illustrated in Fig.2.3. The functions of electric field are defined in each domain, with bounded supports $\Omega_{\boldsymbol{\rho}}$ and $\Omega_{\boldsymbol{\kappa}}$ of arbitrary shape in general. With the mapping $\boldsymbol{\rho} \leftrightarrow \boldsymbol{\kappa}$, each sampling of the field in $\Omega_{\boldsymbol{\rho}}$ is mapped to a single value of the field in $\Omega_{\boldsymbol{\kappa}}$, and vice versa.

HFT (or inverse HFT) is purely a mathematical tool that performs the Fourier transform via a pointwise calculation. This behavior of the Fourier transform was discussed in the work of Bryngdahl in the context of geometric transformations in optics [85] and is based on the stationary phase method [91]. The accuracy of the HFT is mathematically determined by the validity of the stationary-phase approximation and the bijective characteristic of the

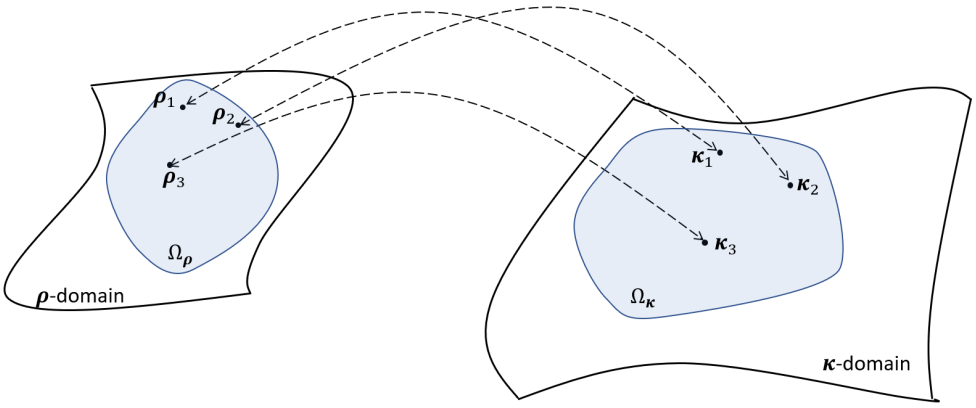

Fig. 2.3 A bijective map between functions on two different domains.

Table 2.1 The pointwise properties of the operators through the light path of the system and their resulting homeomorphic field pairs

Operators	**Description**	**Homeomorphism between fields**
$\boldsymbol{\mathcal{B}}$	pointwise operator (most elements)	$\boldsymbol{E}^{\text{in}}(\boldsymbol{\rho}^{\text{in}})$ and $\boldsymbol{E}^{\text{out}}(\boldsymbol{\rho}^{\text{out}})$
$\boldsymbol{\mathcal{F}}$	pointwise if the HFT is accurate	$\boldsymbol{E}^{\text{out}}(\boldsymbol{\rho}^{\text{out}})$ and $\tilde{\boldsymbol{E}}^{\text{out}}(\boldsymbol{\kappa}^{\text{out}})$
$\boldsymbol{\mathcal{P}}$	pointwise operator	$\tilde{\boldsymbol{E}}^{\text{out}}(\boldsymbol{\kappa}^{\text{out}})$ and $\tilde{\boldsymbol{E}}^{\text{tar}}(\boldsymbol{\kappa}^{\text{tar}})$
$\boldsymbol{\mathcal{F}}^{-1}$	pointwise operator	$\tilde{\boldsymbol{E}}^{\text{tar}}(\boldsymbol{\kappa}^{\text{tar}})$ and $\boldsymbol{E}^{\text{tar}}(\boldsymbol{\rho}^{\text{tar}})$

wavefront phase terms $\psi(\boldsymbol{\rho})$ (or $\tilde{\psi}(\boldsymbol{\kappa})$) [86]. As a rule of thumb, when the system appears in a non-paraxial situation, and the field contains a strong wavefront phase, the HFT can be accurately applied for the Fourier transform of the field.

Because the target plane is at the far-field zone, where the wavefront phase $\tilde{\psi}(\boldsymbol{\kappa})$ in $\boldsymbol{\kappa}-$domain is approximated as a strong spherical phase function per definition, the inverse HFT can be applied for $\boldsymbol{\mathcal{F}}^{-1}$ in high accuracy. Therefore, to build a homeomorphic relation through whole light path, the only concern is the direct Fourier transform $\boldsymbol{\mathcal{F}}$ behind the optical element. For every certain system, the homeomorphism between $\boldsymbol{E}^{\text{out}}(\boldsymbol{\rho}^{\text{out}})$ and $\tilde{\boldsymbol{E}}^{\text{out}}(\boldsymbol{\kappa}^{\text{out}})$ must always be checked by numerical approaches. When their behavior is closed to a homeomorphic relationship, the total homeomorphic relationship between $\boldsymbol{E}^{\text{in}}(\boldsymbol{\rho}^{\text{in}})$ and $\boldsymbol{E}^{\text{tar}}(\boldsymbol{\rho}^{\text{tar}})$ is established. That is, $\boldsymbol{\rho}^{\text{tar}}(\boldsymbol{\rho}^{\text{in}})$ is a bijective mapping function.

Table 2.1 shows a summary of the pointwise properties of the operators through the light-shaping system in Fig. 2.2 and the homeomorphic fields pairs they provide. Therefore, we can conclude that in the literature, the geometric-optics-based design algorithms are sufficiently precise when the HFT is accurate enough for the operator $\boldsymbol{\mathcal{F}}$. If the homeomorphic condition for $\boldsymbol{\mathcal{F}}$ is not satisfied, extra physical effects will be introduced to the system, and the design result based on geometric optics is not valid.

Local energy conservation

Without considering energy loss through propagation, the radiant flux from the source to the target should be conserved. In the system shown in Fig. 2.2, the energy conservation law with respect to different positions is written as

$$\iint E_e^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})\,\mathrm{d}\boldsymbol{\rho}^{\mathrm{in}} = \iint E_e^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})\,\mathrm{d}\boldsymbol{\rho}^{\mathrm{out}} = \iint E_e^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})\,\mathrm{d}\boldsymbol{\rho}^{\mathrm{tar}}, \tag{2.7}$$

where E_e^{in}, E_e^{out} are the irradiance distributions on the plane in front of and behind the optical element respectively, and E_e^{tar} the distribution on the target plane.

Irradiance is defined as the local radiant flux received by a surface and directly equal to the $z-$component of the Poynting vector, where the z-axis is parallel to the normal of the illuminated surface, which is assumed macroscopically planar. Therefore,

$$E_e(\boldsymbol{\rho}) = \langle s_z(\boldsymbol{\rho})\rangle\,, \tag{2.8}$$

where $\langle s_z\rangle$ is the $z-$component of the time-averaged Poynting vector, which in general is defined as

$$\langle \boldsymbol{s}(\boldsymbol{\rho})\rangle = \frac{1}{2}\,\mathbb{R}\mathrm{e}\left\{\boldsymbol{E}(\boldsymbol{\rho})\times\boldsymbol{H}^{*}(\boldsymbol{\rho})\right\}. \tag{2.9}$$

The magnetic field $\boldsymbol{H}(\boldsymbol{\rho})$ of the light can be derived from its electric field $\boldsymbol{E}(\boldsymbol{\rho})$ by Maxwell's equations.

Therefore, from Eqs. 2.8 and 2.9, the homeomorphic assumption between the electric fields on different planes indicates the irradiance on the corresponding plane also exists in a homeomorphic relationship. With the one-to-one map of the irradiance, Eq. 2.7 is derived to its derivative form, which is usually known as the local energy conservation law,

$$\det\left[J\left(\boldsymbol{\rho}^{\mathrm{tar}}\left(\boldsymbol{\rho}^{\mathrm{in}}\right)\right)\right] = \frac{E_e^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})}{E_e^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{in}}))}, \tag{2.10a}$$

$$\det\left[J\left(\boldsymbol{\rho}^{\mathrm{tar}}\left(\boldsymbol{\rho}^{\mathrm{out}}\right)\right)\right] = \frac{E_e^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})}{E_e^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{out}}))}, \tag{2.10b}$$

where $J\left(\boldsymbol{\rho}^{\mathrm{tar}}\left(\boldsymbol{\rho}^{\mathrm{in}}\right)\right)$ and $J\left(\boldsymbol{\rho}^{\mathrm{tar}}\left(\boldsymbol{\rho}^{\mathrm{out}}\right)\right)$ are the Jacobian matrices of the mapping function $\boldsymbol{\rho}^{\mathrm{tar}}\left(\boldsymbol{\rho}^{\mathrm{in}}\right)$ and $\boldsymbol{\rho}^{\mathrm{tar}}\left(\boldsymbol{\rho}^{\mathrm{out}}\right)$, respectively. Their determinants are denoted by $\det$.

Eq. 2.10a is the assumption for the algorithms to design a freeform surface [36, 37, 48, 49, 52]; Eq. 2.10b is the one applied for those geometric-optics algorithms that search for a required output phase [28, 30, 32].

2.3 Validity of the homeomorphic assumption in design

For those algorithms that consider the light-shaping problem under the geometric-optics assumption, the validity of the design is still open to question because the homeomorphic behavior for the field behind the element should be confirmed for each certain case. A ray-tracing simulation is usually performed in the literature to analyse the design result. When the result from ray tracing coincides exactly with the target, it is proof of the design's validity. However, using a simulation technique that is also under the geometric-optics assumption is not enough for a full analysis of the accuracy of the design because any eventual physical-optics effects would go unnoticed, both in the design and in the subsequent simulation. Therefore, to answer whether the geometric-optics assumption is accurate enough for the design algorithm, the system with the designed result must be investigated via physical-optics simulation.

In this section, two simple light-shaping systems are shown in Fig. 2.4 for discussion. In this example, the optical element requested is a freeform lens with a square aperture. The freeform lens is sought to shape the input field to a given irradiance distribution with two different sizes. In both cases, the freeform lens is configured with a fixed, predefined, planar surface and the freeform surface to be designed. N-BK7 glass is used for the lens, with a refractive index of 1.52 at a wavelength of $532\ \mathrm{nm}$. The size of the target irradiance pattern for the first case is $900 \times 900\ \mathrm{mm}$ (shown in Fig. 2.4 (a)). The size of the pattern for the second case is smaller at $150 \times 150\ \mathrm{mm}$ (shown in Fig. 2.4 (b)). The source is taken to be a plane-wave field, with wavelength $532\ \mathrm{nm}$, and field size $1 \times 1\mathrm{mm}$. For a more straightforward discussion, the target plane is set perpendicular to the optical axis and located $1\ \mathrm{m}$ away from the freeform lens. We selected the parameters for our examples to demonstrate the theoretical considerations. The conclusions for this example remain valid when the size of the freeform surfaces, distances and target sizes are scaled accordingly, which shows the conclusions are independent of fabrication constraints.

Design of freeform lens

The freeform surface is designed using the algorithm presented in the author's previous work [92], which can also be understood as a ray mapping method developed under the geometric-optics assumption. The algorithm assumes a homeomorphism, that is, a bijective map between the coordinate of the irradiance function on the input plane and the target, with the input plane defined right in front of the lens. Therefore, the local energy conservation law by Eq. 2.10a is used in the design.

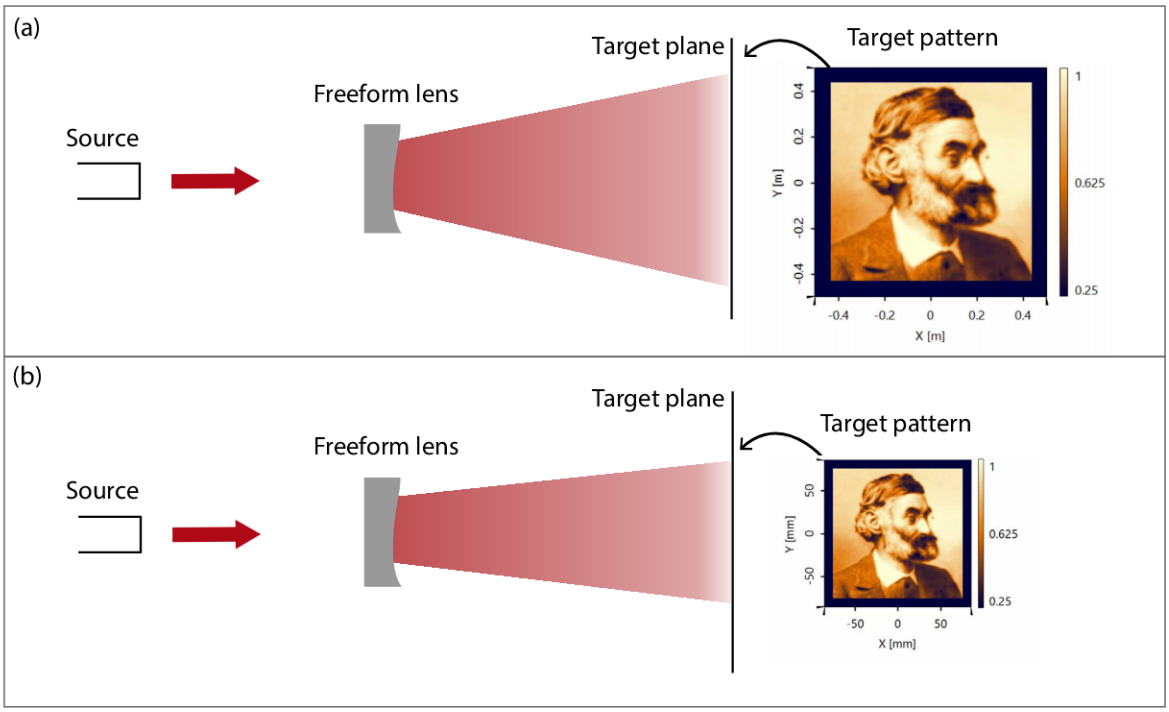

Fig. 2.4 Light-shaping tasks with the freeform surface to be designed.

The search for a smooth, bijective mapping $\boldsymbol{\rho}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{in}})$ in Eq. (2.10a) can be mathematically derived as the L^2 Monge–Kantorovich problem or the optimal mass transport (OMT) problem. Numerical methods have been proposed to solve the problem [67–70]. Here, we consider the irradiance defined on 2D supports in $\mathbb{R}^2$ and implement the algorithm from Prins[70] with some modifications because said algorithm is flexible enough to solve the mapping between two supports with different boundaries. The algorithm for solving the OMT problem will be introduced in detail in Section 3.4. In both light-shaping tasks, 301×301 sampling points are used to calculate the mapping. The same number of sampling points is also used later for the surface design.

Once the mapping is solved, an iterative algorithm is performed to design the surface structure [92]. An initial guess for the surface is necessary to start the algorithm. Here, the surface is simply initialized as a planar surface because the algorithm is not sensitive to the initialization. The required normal vector of the surface is found by Snell's law in each iteration, with the mapping, the input wavefront and the current surface profile. Then, a B-spline method is used to integrate the surface height function from its gradient [93], which can be concluded directly from the normal vector. The difference of the surface height value between two iterations is calculated, and the L^2 norm of the relative difference is used as a

merit function. The algorithm is applied iteratively until the surface converges. Usually, the surface converges quickly, within five iterations.

The height profiles of the designed freeform surface for both cases are shown in Fig. 2.5. The freeform surface is represented by B-spline functions. The results show that the profiles are similar in both cases, but have different gradients.

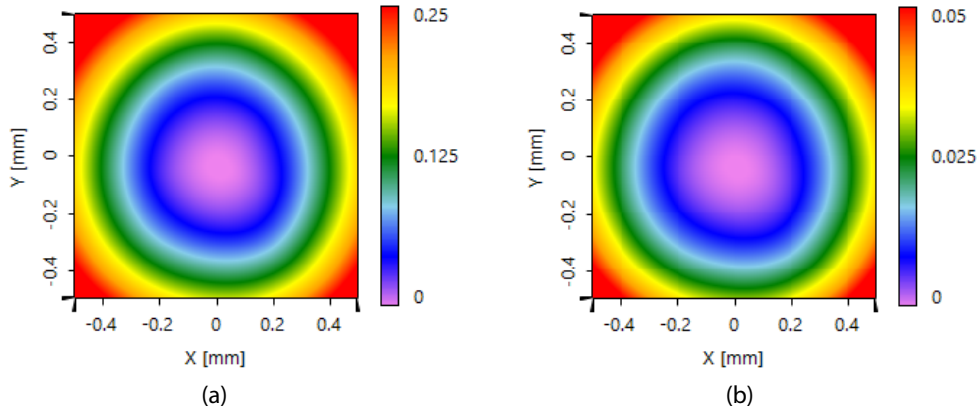

Fig. 2.5 The 2D height profile (Unit: mm) of the designed freeform surface in (a) the non-paraxial case, and in (b) the paraxial case.

Investigation of the design result

The design is investigated with different simulation techniques. A ray tracing simulation is performed to provide a preliminary view of the designed surface then continues with a field tracing process to reveal the physical effects in the system.

The ray-tracing simulation for the freeform surface of case 1 (Fig. 2.4 (a)) is shown in Fig. 2.6 (a) – source plane and (b) – target plane. The ray positions are connected to illustrate the deformation of the input mesh by the freeform surface. It also shows that the freeform surface slightly deforms the square shape of the input. This is a well-known issue because the surface gradient derived from the mapping is not necessary integrable [72]. The B-spline technique has found an optimal approximated surface with the obtained gradient data. The deformation of the surface could be avoided by a further optimization step. For this example, it is of no concern because investigating the validity of the mapping assumption itself is the focal point, which is to be studied next.

To verify that the homeomorphic situation and the assumption in the design algorithm are valid for both examples, it is important to analyze whether $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is in the homeomorphic zone, or, in other words, if the HFT can be applied accurately to perform the Fourier transforms. Therefore, the system with the designed freeform surface is first simulated with

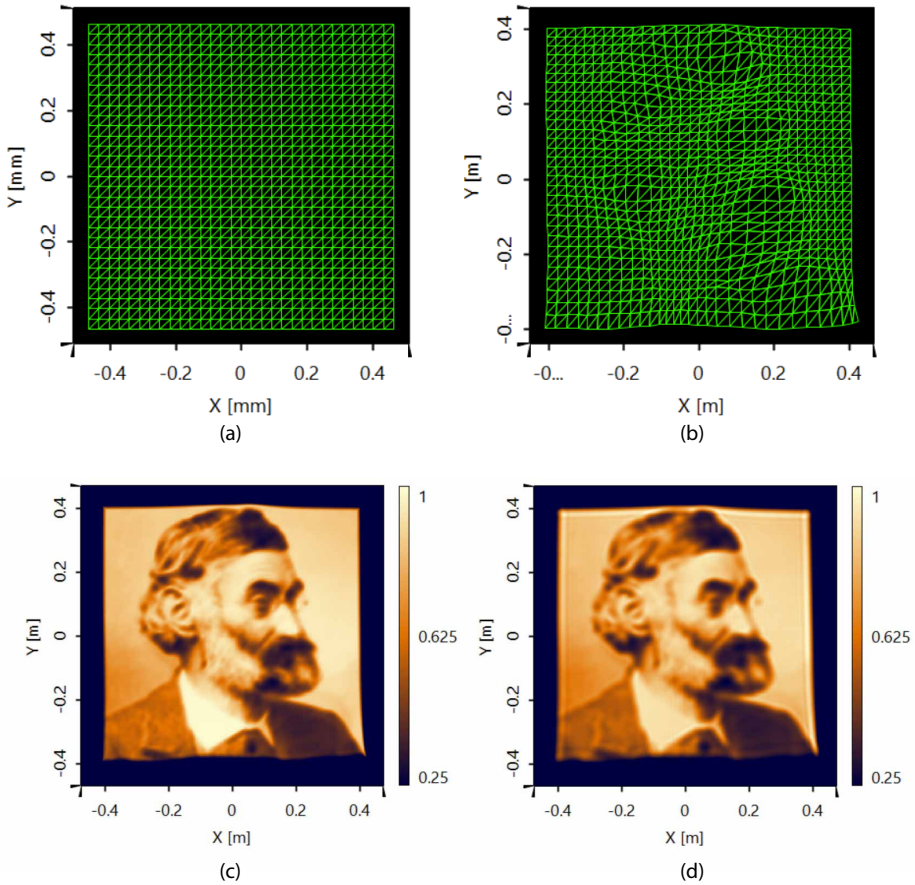

Fig. 2.6 (a) Source rays sampled on the mesh nodes. (b) Ray-tracing result on the detector plane, with a mesh established for the rays using the same indexing from the source mesh. (c)(d) Irradiance detector result with field tracing techniques. Simulation with (c) the HFT and, (d) the FFT.

HFT. Then, we investigate their accuracy by comparing the simulation result against the one obtained with the rigorous FFT.

$\mathcal{F}$ in Fig. 2.2 is selected as the HFT to begin investigating the system of the first example ((Fig. 2.4 (a)). Therefore, a pointwise calculation between $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ is performed in the whole-system simulation. The irradiance at the target plane is measured by a detector, and shown in Fig. 2.6 (c). The result in Fig. 2.6 (c) coincides with the target pattern in Fig. 2.4. However, as discussed, that is not sufficient proof of the validity of the designed surface. Next, the Fourier transform for $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is switched to the FFT to calculate $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$. The detected irradiance at the target plane is shown in Fig. 2.6 (d). The resulting irradiance distribution also mostly coincides with the target distribution, with the exception of only slight errors which occur at the edges of the pattern, small enough to be neglected. Therefore, the validity of the designed freeform surface for the task is

proved. Both techniques give coincident results that show the HFT and the inverse HFT are good approximate algorithms for the direct and inverse Fourier transform in this case. Consequently, the full sequence from $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ exhibits a homeomorphic behavior, so that the geometric-optics assumption $\boldsymbol{\rho}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{in}})$ is quite a good approximation for the design in this example.

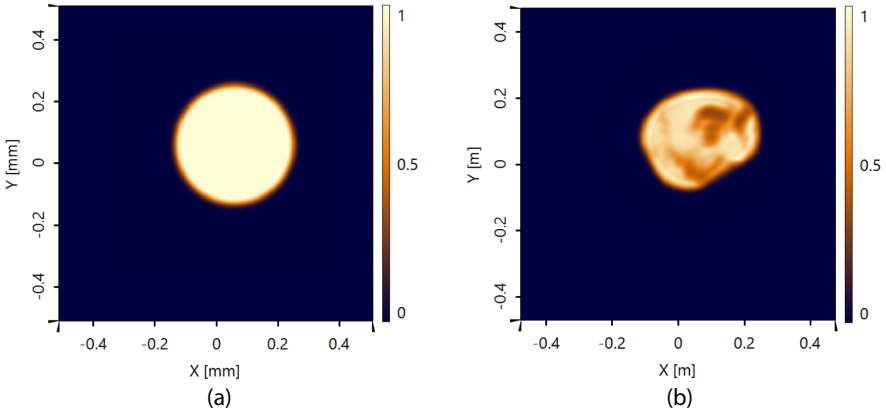

Fig. 2.7 (a) Irradiance (normalized) of a truncated plane wave. (b) The corresponding irradiance (normalized) on the detector, obtained with the truncated plane wave.

The homeomorphism also can be illustrated between a truncation of the source and its corresponding simulation result obtained on the detector. Fig. 2.7 (a) shows the irradiance of the input plane wave when an off-axis circular aperture is introduced. The corresponding irradiance with the truncated plane wave is shown in Fig. 2.7 (b). The figures indicate that the homeomorphism indeed happens in this example. Although the pattern in Fig. 2.7 (b) is deformed from the shape of the aperture of the input, the mapping between the source and the target still provides a clear pattern showing that the neighboring sample of the source does not contribute to a same area at the target plane that may cause blurring and interfering.

For case 2 (Fig. 2.4 (b)), simulations with the homeomorphic assumption and the rigorous calculation are performed as in the previous example. The resulting irradiance distribution on the target plane is shown in Fig. 2.8 (a) for the simulation with the HFT and in Fig. 2.8 (b) for the simulation with the FFT. The result in Fig. 2.8 (a) again coincides with the target pattern as expected. However, the other result shows obvious diffraction effects, with low resolution in the irradiance distribution, which produce a large deviation from the target of the design task. The blurring of the result in Fig. 5 (b) is in nice agreement with the conclusions of Zwick et al [94], and provides a 2D demonstration. Fig. 2.8 (a) and Fig. 2.8 (b) together show that the simulation with the HFT in this case is far from a good approximation, and there is no homeomorphism between $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$. Therefore, the freeform surface designed via the geometric-optics assumption cannot fulfill the requirements of the light-shaping task.

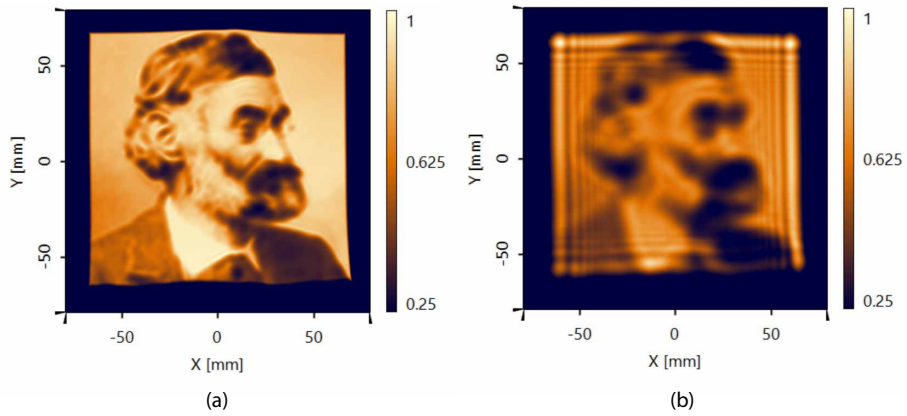

Fig. 2.8 Detector result: irradiance. Simulation with (a) the HFT (b) the FFT.

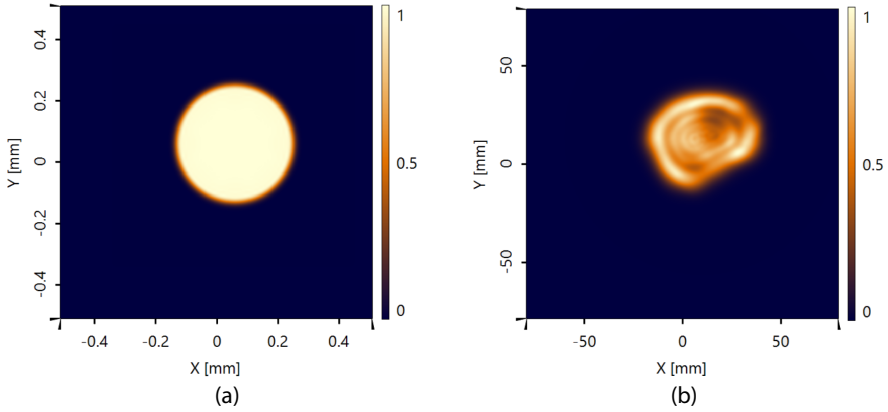

Fig. 2.9 (a) Irradiance (normalized) of a truncated plane wave. (b) The corresponding irradiance (normalized) on the detector, obtained with the truncated plane wave.

Similarly, a truncated source with its corresponding result is shown in Fig. 2.9. With the same aperture introduced for the source, the simulation result on the detector appears blurred with strong interfere fringes. The figures again indicate that the homeomorphic condition is not satisfied in this paraxial case.

With the two examples in hand, we would like to point out that for those design algorithms that include the homeomorphic assumption for the field (or irradiance distribution) between the input and target planes, such as ray mapping design algorithms, the accuracy of the assumption should be determined by physical-optics simulation with the designed result. If the analysis is done by the geometric-optics-based ray tracing technique, the simulation result will always coincide with the target pattern for the design because both design and analysis follow the same assumption. Additional physical effects that might appear in reality, such as diffraction, will go unnoticed. As a rule of thumb, the more non-paraxial (both input and output side) the system, the more accurate the geometric-optics assumption. The simulation

results of the two examples also reveal that the accuracy of the geometric-optics assumption essentially depends on how accurate the homeomorphic behavior is, or in other words, how accurately the HFT and inverse HFT can be applied in the field tracing.

The simulations were done with the software VirtualLab Fusion [95]. One physical-optics simulation on a PC (2.39 GHz Intel(R) Core(TM) Quad CPU), in the current implementation, takes a few seconds. In field tracing, the sampling of the field for the propagation typically dominates the calculation time and this sampling is not related to the selected resolution in the detector.

2.4 Summary

In this chapter, the field tracing techniques are illustrated and applied to analyze typical far-field light-shaping systems. The field tracing techniques reveal from a physical-optics modeling point of view that if all the operators in the system are pointwise operators, the homeomorphism of the fields through the system is established. The most critical point for the homeomorphic situation is laid on the Fourier transform operator $\mathcal{F}$ for the field $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ behind the optical element. If $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ satisfied the stationary phase condition, the pointwise HFT can be accurately applied here; therefore, the homeomorphism exists in the system.

For those geometric-optics-based design algorithms that use the homeomorphic assumption between the input and target field, the design validity cannot be examined by a geometric-optics approach because both the design and modeling are under the same assumption. Field tracing techniques provide a tool to investigate the result of the design. An example with two light-shaping tasks demonstrates that in the non-paraxial case where $\mathcal{F}$ for $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ can be accurately performed by the HFT, the designed freeform surface with the homeomorphic assumption realizes the shaping task well. In contrast, in the paraxial case where the HFT is not accurate for the $\mathcal{F}$ operator, the designed freeform surface is not the solution for the task.

Chapter 3

Design functional embodiment of the light-shaping element

The functionality of the required optical element for a certain far-field light-shaping system can be determined with the modeling techniques introduced in the last chapter. The inverse approach for optical design starts from the source and the target simultaneously. The field tracing technique helps to propagate each field from the source and the target to the optical element's position, with or without other predefined optical components in the system. Therefore, the functionality of the optical element, that is, its corresponding $\mathcal{B}$ operator, can be concluded from its input and output fields. However, in practice, due to the lack of information about the field or additional constraints for the system, the formulation of the required element function can be far from straightforward. In this chapter, the search for the functional embodiment of the required optical element is discussed based on the field tracing techniques, and algorithms are proposed to solve the problem.

The element's function for the far-field light shaping is mainly a phase-only response function. The optical element shapes the phase of the input field into a required output phase, which dominates the output propagation and attains the target energy distribution in the far-field zone. Apparently, the optical element, in general, will introduce amplitude modulation to the input field as well; however, it is not considered in the functional design step, but considered as an added effect that will be optimized later in the structural design step. Therefore, the functional design step neglects any constraints from a real structure and concentrates on the physics of the field itself. Thus, for the objective of designing a phase-only response function, knowing the output phase on request to achieve the target energy is the key point.

One intuitive thought for the energy redistribution is to find a 2D mapping between the input and the target irradiance distribution. Those geometric-optics-based algorithms either

for designing the freeform surface [48, 51] or for the output phase retrieval [32, 96] are based on the achieved mapping. By knowing how the irradiance redistributes in the space domain, the 2D mapping can be applied to compute the gradient of the output phase with the stationary phase method. The output phase is achieved quickly; however, the validity of the mapping assumption should be checked for certain optical systems, as was discussed in Chapter 2. In addition, the obtained gradient field is not necessarily integrable cannot be used to construct the output phase directly.

Another phase retrieval method usually used in the literature is based on a Fourier pair synthesis. The field on the target plane is calculated from the target energy quantities, for example, irradiance or radiant intensity. The calculated target field is used as the constraint for the design. Another constraint is introduced from the given input field. With the far-field assumption, both the input field and the target field appears a Fourier relationship. The Fourier pair synthesis searches a required phase for the input field so that the Fourier relationship between the fields is established and both constraints are satisfied.

In fact, the constraint from the target field can be applied at any position in the field tracing sequence for the free space propagation. We trace the target field back to the output plane of the required optical element, and apply the constraint for the field in the $\boldsymbol{\kappa}-$domain. The process will be demonstrated in Section 3.1. The Fourier pair synthesis is performed in a more direct way between the field on the output plane in the $\boldsymbol{\rho}-$domain and its corresponding field in the $\boldsymbol{\kappa}-$domain. This is also a more general way for light-shaping problems with its target plane located either in the far-field zone or NOT in the far-field zone. Because once the target field is known, the far-field assumption is not needed.

The iterative Fourier transform algorithm (IFTA) is a well-developed approach that iteratively synthesizes the Fourier pair, as its name implies. However, it may lead to stagnation problems when the fields of the Fourier pair are in a homeomorphic situation. The problem usually occurs when the output propagation is in a non-paraxial situation, where the output field contains a strong wavefront phase. A mathematical proof is demonstrated later in Section 3.1. A mapping-type Fourier pair synthesis is introduced to overcome the above-mentioned problem. Instead of an iterative process, the approach results in a required output phase in a single step, without iterations. Moreover, the output phase obtained by the approach is a smooth function, from which a smooth element function or even a smooth element structure can be designed, thus reducing of speckles [97, 98] or offering higher conversion efficiency [20] may result. The algorithm is discussed in Section 3.2.

Comparing the mapping-type Fourier pair synthesis algorithm with the other mapping-type algorithm in which the homeomorphism is assumed between the input and target irradiance distributions in the space domain, the calculating mapping function of the fields in

the Fourier pair can be directly applied to the integration of the request output phase. The existence and integrability of the mapping function are addressed in Section 3.3. Therefore, a smooth output phase is obtained by direct integration without any approximations.

3.1 Design with Fourier pair synthesis

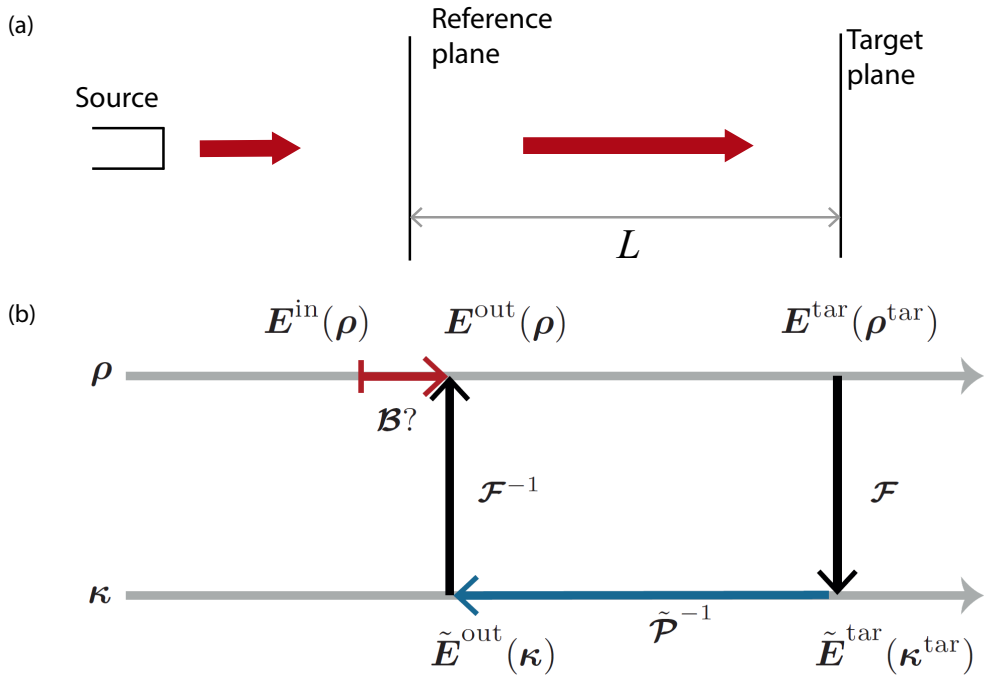

Fig. 3.1 Analysis of the design task under the framework of field tracing techniques.

The light-shaping problem is analyzed under the framework of field tracing. The light-shaping system, together with the modeling techniques, is shown in Fig. 3.1 for the following discussion. The requested optical element is defined on its reference plane, where its functionality is to be searched. The field tracing diagram (Fig. 3.1(b)) indicates the modeling methods used for the design. We consider the problem with the given information from the design task listed as follows:

- the input field $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho})$
- the signal irradiance $E_e^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$
- the position of the reference plane $\mathrm{P}^{\mathrm{ref}}$ where the requested optical element is defined
- the position of the target plane $\mathrm{P}^{\mathrm{tar}}$ where the signal is defined

Both the reference plane and the target plane are considered perpendicular to the optical axis for this discussion. However, that alignment can be generalized to tilted planes or curved surfaces without changing the design logic illustrated in the following content. The distance between the planes is indicated as L.

The input field is written explicitly as:

$$E_\ell^{\mathrm{in}}(\boldsymbol{\rho}) = \left|E_\ell^{\mathrm{in}}(\boldsymbol{\rho})\right| \exp[\mathrm{i}\gamma_\ell^{\mathrm{in}}(\boldsymbol{\rho})], \tag{3.1}$$

Here, the notation E_ℓ is used to represent any component of the field $\boldsymbol{E}$, with $\ell = x, y, z$. $\left|E_\ell^{\mathrm{in}}(\boldsymbol{\rho})\right|$ is the amplitude of any component of the field $E_\ell^{\mathrm{in}}(\boldsymbol{\rho})$. $\gamma_\ell^{\mathrm{in}}$ is the phase of each component. Similarly, the output field is written as:

$$E_\ell^{\mathrm{out}}(\boldsymbol{\rho}) = \left|E_\ell^{\mathrm{out}}(\boldsymbol{\rho})\right| \exp[\mathrm{i}\gamma_\ell^{\mathrm{out}}(\boldsymbol{\rho})]. \tag{3.2}$$

Here, both the input and output field are defined on the reference plane of the optical element with its transversal coordinate as $\boldsymbol{\rho}$.

In general, the input field is obtained by forward tracing the field from the source, with or without predefined optical elements in the system, while the output field is obtained by backward tracing the field from the target.

The field on the target plane is written as:

$$E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) = \left|E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})\right| \exp[\mathrm{i}\psi^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})]. \tag{3.3}$$

Because the target plane is at the far-field zone, the phase of the target field is an approximated spherical phase that is common for all its three components per definition. Here, ψ is used to denote the smooth phase, with

$$\psi^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) = k_0 n \sqrt{\left|\boldsymbol{\rho}^{\mathrm{tar}}\right|^2 + L^2}, \tag{3.4}$$

where L is the propagating distance from the output plane to the target plane.

The amplitude of the target field can be determined using the given irradiance distribution. Under the far-field assumption, the relation of the irradiance and the field can be formulated as follows:

$$\begin{aligned} E_e^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) &= \left\langle \boldsymbol{s}(\boldsymbol{\rho}^{\mathrm{tar}}) \right\rangle \cdot \hat{N}(\boldsymbol{\rho}^{\mathrm{tar}}) \\ &= \frac{n}{2\mu_0 c} \hat{\boldsymbol{s}}(\boldsymbol{\rho}^{\mathrm{tar}}) \hat{N}(\boldsymbol{\rho}^{\mathrm{tar}}) \left\| \boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) \right\|^2, \end{aligned} \tag{3.5}$$

where $\langle \boldsymbol{s} \rangle$ is the time-averaged Poynting vector, and $\hat{N}$ is the normal vector of the target plane. $\hat{N} = (0, 0, 1)$ if the target plane is perpendicular to the optical axis. n is the refractive index, μ_0 the vacuum permeability, and c the speed of light in vacuum. $\hat{\boldsymbol{s}}$ is the unit vector of the Poynting vector, which can be calculated by $\hat{\boldsymbol{s}} = (x/r, y/r, L/r)$, where (x, y) is the coordinate of the irradiance function on the target plane, and $r = \sqrt{x^2 + y^2 + L^2}$. $\|\boldsymbol{E}\|$ denotes the L^2 norm of the field, with

$$\begin{aligned} \|\boldsymbol{E}\|^2 &= E_x^2 + E_y^2 + E_z^2 \\ &= E_x^2 + E_y^2 + (\frac{s_x E_x + s_y E_y}{s_z})^2. \end{aligned} \tag{3.6}$$

The amplitude of $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ can be calculated from the irradiance if a polarization state is assumed. In fact, the polarization state of the target field is a free parameter for the design, for example, global polarized. The global polarized field is indicated as:

$$\begin{bmatrix} E_x^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) \\ E_y^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) \end{bmatrix} = \begin{bmatrix} \mathcal{J}_x \\ \mathcal{J}_y \end{bmatrix} E^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}), \tag{3.7}$$

where $\begin{bmatrix} \mathcal{J}_x \\ \mathcal{J}_y \end{bmatrix}$ is the Jones vector, and $E^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ is the common field distribution. Therefore, with Eq. (3.6) and Eq. (3.7), we can derive

$$\left\| \boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) \right\|^2 = P \left| E^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) \right|^2, \tag{3.8}$$

where P is a polarization term that is computed by:

$$P = \mathcal{J}_x^2 + \mathcal{J}_y^2 + (\frac{s_x \mathcal{J}_x + s_y \mathcal{J}_y}{s_z})^2, \tag{3.9}$$

Thus, $E^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ is derived from the irradiance $E_e^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ with the equation

$$\left| E^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) \right|^2 = \frac{2\mu_0 c}{n} \frac{E_e^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})}{\hat{\boldsymbol{s}}(\boldsymbol{\rho}^{\mathrm{tar}}) \hat{N}(\boldsymbol{\rho}^{\mathrm{tar}})} \frac{1}{P}, \tag{3.10}$$

So far, the target field is calculated as

$$\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) = \begin{bmatrix} \mathcal{J}_x \\ \mathcal{J}_y \\ \dfrac{s_x \mathcal{J}_x + s_y \mathcal{J}_y}{s_z} \end{bmatrix} \left| E^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}}) \right| \exp[\mathrm{i}\psi^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})]. \tag{3.11}$$

With the obtained target field, in general, one can conclude the output field behind the optical element, via an inverse application of the modeling methods in the field tracing diagram. Therefore, $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ is derived as

$$E_\ell^{\mathrm{out}}(\boldsymbol{\rho}) = \boldsymbol{\mathcal{F}}^{-1}\left\{\tilde{\boldsymbol{\mathcal{P}}}^{-1}\boldsymbol{\mathcal{F}}\left\{E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})\right\}\right\}. \tag{3.12}$$

In general, the amplitude of $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ attained by the backward propagation is not necessarily equivalent to the one of $E_\ell^{\mathrm{in}}(\boldsymbol{\rho})$. Therefore, if one assumes the functionality of the optical element is a phase-only modulation, $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ can not be directly used for the calculation of the element function because the amplitude difference from $E_\ell^{\mathrm{in}}(\boldsymbol{\rho})$ introduces an amplitude modulation to the element function.

Therefore, in the design process, an amplitude constraint is applied for $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ that ensures both the amplitudes of the input and output field are the same. The requirement of the $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ is naturally another constraint. With the far-field assumption, $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ and $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ appears a Fourier relation. Therefore, in the literature, a Fourier pair synthesis is usually applied for obtaining the phase of $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ so that the Fourier pair is established and both the constraints are met.

However, one can see from Fig. 3.1(b) that, the constraint from the target field can be considered on $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$, $\tilde{E}_\ell^{\mathrm{tar}}(\boldsymbol{\kappa}^{\mathrm{tar}})$ or $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ by backward-propagation modeling. The most direct Fourier pair synthesis is apply the constraint from the target on $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$. Therefore, the field on the target plane is traced until $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$, which is the corresponding spectrum of $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ in the $\boldsymbol{\kappa}-$domain, as illustrated in the field tracing diagram in Fig. 3.1(b),

$$\begin{aligned}\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa}) &= \tilde{\boldsymbol{\mathcal{P}}}^{-1}\boldsymbol{\mathcal{F}}\left\{E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})\right\} \\ &= \exp[-\mathrm{i}k_z(\boldsymbol{\kappa})L]\boldsymbol{\mathcal{F}}\left\{E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})\right\}.\end{aligned} \tag{3.13}$$

Note that the constraint for $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ is simply an amplitude constraint, while the phase of $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ is usually considered as the *phase freedom*. That is because, under the far-field assumption, the phase introduced from the free space propagation operator ($\exp[\mathrm{i}k_zL]$) is much larger than the phase of $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ and contributes to the majority of the spherical phase of $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$.

The calculation from $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ to $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ in Eq. (3.13) is a pointwise calculation, which establishes a homeomorphism between $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ and $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$. In the far-field zone, the phase of $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ is a spherical phase, and the HFT can be applied for the Fourier transform in Eq. (3.13) accurately per definition [86]. Therefore, all the operators in Eq. (3.13) are pointwise operators. Fig. 3.2 shows different examples illustrating the homeomorphism

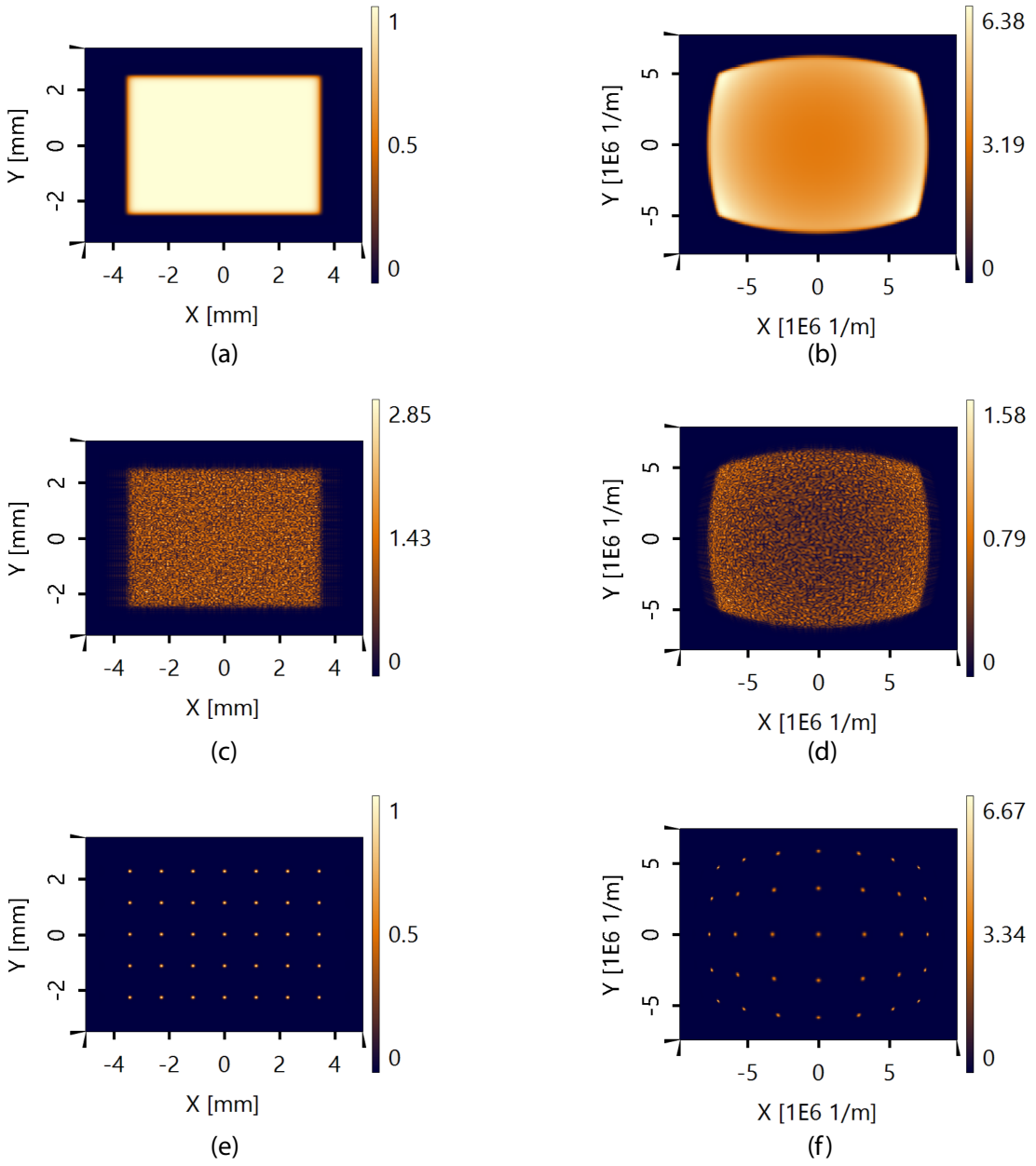

Fig. 3.2 Different examples illustrating the homeomorphism between $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ and $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$. The left column shows the amplitude of different $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$, while their corresponding amplitudes of $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ are shown in the right column. (a)(b), case of a uniform distribution; (c)(d), case of speckle pattern; (e)(f), dot pattern.

between $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ and $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$. With the known amplitude of $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$, as shown in Fig. 3.2 (a), (c) and (e), the amplitude of $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ is calculated by Eq. (3.13), as shown in Fig. 3.2 (b), (d) and (f). Although the target fields exhibit different amplitude distributions in Fig. 3.2 (a), (c) and (e), Fig. 3.2 (b), (d) and (f) all show a corresponding one-to-one map.

Under both amplitude constraints for the fields of a Fourier pair, $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ and $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$, determining the phase of $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ by the Fourier pair synthesis is the key point for the design of the element function. Several algorithms are proposed in the literature for solving the Fourier pair synthesis problem. Details of the algorithms will be discussed in the following sections.

As a result, after the Fourier pair synthesis problem is solved, with both the phases in front of the optical element and the one behind, one can conclude the phase response function $\Delta\gamma_\ell$ that must be realized by the element is nothing other than the subtraction between the input and output phase,

$$\Delta\gamma_\ell(\boldsymbol{\rho}) = \gamma_\ell^{\mathrm{out}}(\boldsymbol{\rho}) - \gamma_\ell^{\mathrm{in}}(\boldsymbol{\rho}). \tag{3.14}$$

3.1.1 Iterative Fourier transform algorithm (IFTA)

IFTA is the algorithm for the Fourier pair synthesis, a term introduced to catalog a group of methods that bounce between the domains of the Fourier pair forwards and backwards [99]. The origin of this error-reducing algorithm can be dated from the early 1970s [100]. Gerchberg and Saxton adapted it for phase retrieval problems [14], and based on their work, the IFTA was developed and has been widely used in many applications for 50 years. Therefore, this kind of iterative algorithm is also referred to as the Gerchberg–Saxton (GS) algorithm. Various developments of the IFTA have been proposed in the last decades [15–20, 22, 26, 101, 102]. For example, Fienup extensively improved the convergence speed for the error reduction [15]. Wyrowski and Bryngdahl [16] have chosen a better initial phase to avoid the stagnation problem and proposed the soft-quantization concept for multi-level phase design[17]. Brenner introduced an unwrap phase step in each iteration to obtain a smooth phase function. Additional constraints were introduced to enhance the diffraction efficiency[22] or further speed up the convergence [26].

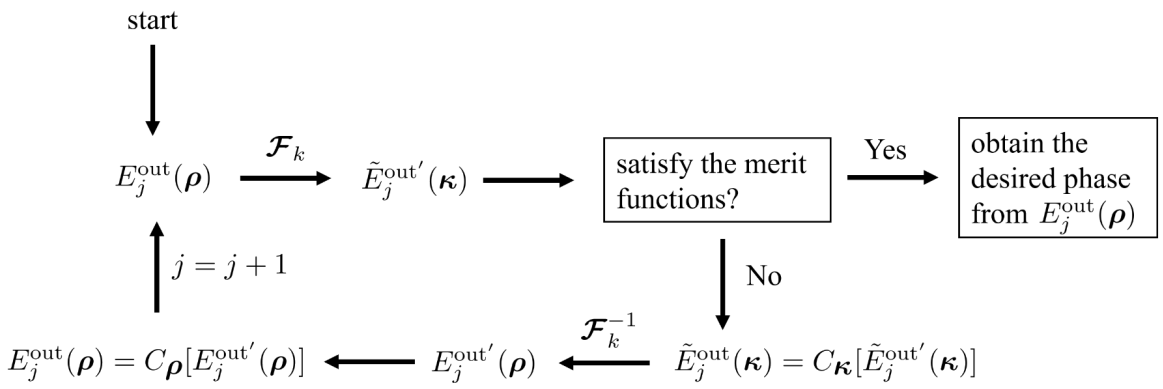

Fig. 3.3 The basic IFTA is displayed in a flowchart-like layout.

The basic algorithm of IFTA is illustrated in Fig. 3.3. In the Fourier pair synthesis problem, we reach the point that both the amplitudes of the fields $E_\ell^{\text{out}}(\boldsymbol{\rho})$ and $\tilde{E}_\ell^{\text{out}}(\boldsymbol{\kappa})$ are obtained. The IFTA attempts to search for a proper phase for $E_\ell^{\text{out}}(\boldsymbol{\rho})$ so that the Fourier relation can be established. The iterative process is along with the amplitude constraints $C_{\boldsymbol{\rho}}$ in the $\boldsymbol{\rho}$ domain and $C_{\boldsymbol{\kappa}}$ in the $\boldsymbol{\kappa}$ domain respectively. The amplitude constraint $C_{\boldsymbol{\rho}}$ replaces the amplitude of $E_\ell^{\text{out}'}(\boldsymbol{\rho})$ that is calculated from the inverse Fourier transform with the amplitude of $E_\ell^{\text{out}}(\boldsymbol{\rho})$ in each iteration. The amplitude constraint $C_{\boldsymbol{\kappa}}$ replaces the calculated amplitude of the field in the $\boldsymbol{\kappa}$ domain with the amplitude of $\tilde{E}_\ell^{\text{out}}(\boldsymbol{\kappa})$. The algorithm retrieves the required output phase for $E_\ell^{\text{out}}(\boldsymbol{\rho})$ that fulfills these constraints, or at least does so approximately.

The numerical implementation of the basic IFTA is shown in Algorithm 1. Here the subscript ℓ of the field component indicator is skipped in the notation. The fields of both domains are sampled, and written as:

$$E^{\text{out}}(\boldsymbol{\rho}_{\text{i}}),\ \boldsymbol{\rho}_{\text{i}} \in \Omega_{\boldsymbol{\rho}},\ \Omega_{\boldsymbol{\rho}} \in \mathbb{R}^2$$
$$\tilde{E}^{\text{out}}(\boldsymbol{\kappa}_{\text{i}}),\ \boldsymbol{\kappa}_{\text{i}} \in \Omega_{\boldsymbol{\kappa}},\ \Omega_{\boldsymbol{\kappa}} \in \mathbb{R}^2$$

where $\boldsymbol{\rho}_{\text{i}}$ (or $\boldsymbol{\kappa}_{\text{i}}$) are the sampling points of $E^{\text{out}}(\boldsymbol{\rho}_{\text{i}})$ (or $\tilde{E}^{\text{out}}(\boldsymbol{\kappa}_{\text{i}})$). $\Omega_{\boldsymbol{\rho}}$ and $\Omega_{\boldsymbol{\kappa}}$ are the bounded support in $\mathbb{R}^2$, where both fields in the two domains are defined, respectively.

Algorithm 1: The basic iterative Fourier transform algorithm.

Input: $|E^{\text{out}}(\boldsymbol{\rho}_{\text{i}})|$, $\left|\tilde{E}^{\text{out}}(\boldsymbol{\kappa}_{\text{i}})\right|$, an initial phase in $\boldsymbol{\rho}-$domain $\gamma_0^{\text{out}}(\boldsymbol{\rho}_{\text{i}})$, the maximum iteration $j^{\max}$ and a tolerance value TOL

Output: an optimized phase in $\boldsymbol{\rho}-$domain $\gamma^{\text{out}}(\boldsymbol{\rho}_{\text{i}})$

Initialization: let $E^{\text{out}}(\boldsymbol{\rho}_{\text{i}}) = |E^{\text{out}}(\boldsymbol{\rho}_{\text{i}})|\, \gamma_0^{\text{out}}(\boldsymbol{\rho}_{\text{i}})$
while $j <= j^{\max}$ **do**
    Fourier transform: $\tilde{E}_j^{\text{out}'}(\boldsymbol{\kappa}_{\text{i}}) = \boldsymbol{\mathcal{F}}\{E_j^{\text{out}}(\boldsymbol{\rho}_{\text{i}})\}$
    if $\varepsilon > TOL$ **then**
        Set $\boldsymbol{\kappa}-$domain amplitude constraint: $\tilde{E}_j^{\text{out}}(\boldsymbol{\kappa}_{\text{i}}) = C_{\boldsymbol{\kappa}}[\tilde{E}_j^{\text{out}'}(\boldsymbol{\kappa}_{\text{i}})]$
        Perform inverse Fourier transform: $E_j^{\text{out}'}(\boldsymbol{\rho}_{\text{i}}) = \boldsymbol{\mathcal{F}}^{-1}\{\tilde{E}^{\text{out}}(\boldsymbol{\kappa}_{\text{i}})\}$
        Set $\boldsymbol{\rho}-$domain amplitude constraint: $E_j^{\text{out}}(\boldsymbol{\rho}_{\text{i}}) = C_{\boldsymbol{\rho}}[E_j^{\text{out}'}(\boldsymbol{\rho}_{\text{i}})]$
        $j \leftarrow j + 1$
    else
        break
Return $\gamma_j^{\text{out}}(\boldsymbol{\rho}_{\text{i}})$

In the algorithm, ε is a macroscopic mean error calculated by

$$\varepsilon = \frac{\sum^{N} \left(\left| \tilde{E}_j^{\text{out}'}(\boldsymbol{\kappa}) \right| - \left| \tilde{E}_j^{\text{out}}(\boldsymbol{\kappa}) \right| \right)^2}{N}. \tag{3.15}$$

where N is the number of sampling points of $\tilde{E}_j^{\text{out}}(\boldsymbol{\kappa})$. ε is typically used as a merit function to determine the convergence. The convergence of the algorithm was given in the original paper from Gerchberg and Saxton [14] with a good conceptual argument.

The original IFTA indicated in Algorithm 1 is known to give poor performance when the phase is in a quantized form [99]. It tends to converge to a nearest local optimum rather than finding the desired global optimum. Some additional design freedom is applied in IFTA to overcome this problem.

Beside of the *phase freedom* mentioned in the last subsection, an amplitude freedom is usually used in the algorithm. The *amplitude freedom* is applied on $\tilde{E}_\ell^{\text{out}}(\boldsymbol{\kappa})$, such that its amplitude constraint of $C_{\boldsymbol{\kappa}}$ is set only in a certain region where $\tilde{E}_\ell^{\text{out}}(\boldsymbol{\kappa})$ is defined, while the values outside its defined region can be arbitrary complex values.

Moreover, because the fidelity of the generated signal is more important than the quantity of input light, a scale factor may be utilized to adjust the power repartition between the region of $C_{\boldsymbol{\kappa}}$ and the region outside. This freedom is usually referred to as the *scale freedom*.

Therefore, with all the three freedoms, the desired output field in the $\boldsymbol{\kappa}-$domain belongs to the set

$$\begin{aligned} \tilde{E}_\ell^{\text{out}}(\boldsymbol{\kappa}) = \{ \tilde{E}(\boldsymbol{\kappa}) \, | \tilde{E}(\boldsymbol{\kappa}) = \alpha^{SF} \left| \tilde{E}_\ell^{\text{out}}(\boldsymbol{\kappa}) \right| \exp[\mathrm{i}\gamma^{\text{PF}}(\boldsymbol{\kappa})], \; \forall \boldsymbol{\kappa} \in \Omega_{\boldsymbol{\kappa}}; \text{ and} \\ \tilde{E}(\boldsymbol{\kappa}) = \tilde{E}^{AF}(\boldsymbol{\kappa}), \; \forall \boldsymbol{\kappa} \notin \Omega_{\boldsymbol{\kappa}} \}. \end{aligned} \tag{3.16}$$

where $\Omega_{\boldsymbol{\kappa}}$ is the region where the amplitude constraint $C_{\boldsymbol{\kappa}}$ is defined. γ^{PF} is the phase freedom and $\tilde{E}^{AF}(\boldsymbol{\kappa})$ is the amplitude freedom for the region outside $\Omega_{\boldsymbol{\kappa}}$. α^{SF} is the scale freedom factor, which is formulated as,

$$\alpha^{SF} = \frac{\iint_{\Omega_{\boldsymbol{\kappa}}} \left(\tilde{E}_j^{\text{out}'}(\boldsymbol{\kappa}) \tilde{E}_j^{\text{out}}(\boldsymbol{\kappa}) \right) \, \mathrm{d}\boldsymbol{\kappa}}{\iint_{\Omega_{\boldsymbol{\kappa}}} \left| \tilde{E}_j^{\text{out}}(\boldsymbol{\kappa}) \right|^2 \, \mathrm{d}\boldsymbol{\kappa}}. \tag{3.17}$$

Merit functions

The signal-to-noise ratio (SNR) measures the correspondence between a desired function and another function that possesses errors. Here, the SNR is defined as the ratio between the

integral intensity of the output field to the integral intensity of the noise, where the noise is defined as the subtraction between the calculated output field and the target signal field, both in the $\boldsymbol{\kappa}-$domain. In a paraxial situation, the formula is simply expressed in the following with a unit of dB.

$$\mathrm{SNR}\left(\tilde{E}_j^{\mathrm{out}}(\boldsymbol{\kappa}), \tilde{E}_j^{\mathrm{out}'}(\boldsymbol{\kappa})\right) = \frac{\iint_{\Omega_{\boldsymbol{\kappa}}} \left|\tilde{E}_j^{\mathrm{out}'}(\boldsymbol{\kappa})\right|^2 \mathrm{d}\boldsymbol{\kappa}}{\iint_{\Omega_{\boldsymbol{\kappa}}} \left(\left|\tilde{E}_j^{\mathrm{out}'}(\boldsymbol{\kappa})\right| - \alpha^{SF}\left|\tilde{E}_j^{\mathrm{out}}(\boldsymbol{\kappa})\right|\right)^2 \mathrm{d}\boldsymbol{\kappa}}. \tag{3.18}$$

Because the amplitude freedom is introduced in the algorithm, another important merit function is the window efficiency, which is considered as the amount of energy of the input field diffracted in the signal region $\Omega_{\boldsymbol{\kappa}}$ in the target plane [103]. Any noise in the signal region is ignored. The window efficiency can be simply formulated as

$$\eta_{\mathrm{win}} = \frac{\iint_{\Omega_{\boldsymbol{\kappa}}} \left|\tilde{E}_j^{\mathrm{out}'}(\boldsymbol{\kappa})\right|^2 \mathrm{d}\boldsymbol{\kappa}}{\iint_{\Omega_{\boldsymbol{\rho}}} \left|\tilde{E}_j^{\mathrm{in}}(\boldsymbol{\rho})\right|^2 \mathrm{d}\boldsymbol{\rho}}. \tag{3.19}$$

In a non-paraxial situation, all the integral intensity in the above merit functions shall include all three components of the corresponding electric field.

3.1.2 Stagnation in IFTA

In general, the IFTA uses the fast Fourier transform (FFT) method for performing the Fourier transform, which implies each field value of $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ is contributed by all the field values from $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$. However, in certain cases, in which a homeomorphism exists between the fields $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$ and $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$, or in other words, each field value of $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$ is contributed by a single corresponding field value of $E_\ell^{\mathrm{out}}(\boldsymbol{\rho})$, the Fourier transform can be accurately performed by the homeomorphic Fourier transform (HFT) [86], whereas IFTA will fail to converge.

Let us consider the case of a non-paraxial system where the homeomorphic assumption is valid. The output field in $\boldsymbol{\rho}-$domain in Eq. (3.2) is rewritten as:

$$\begin{aligned} E_\ell^{\mathrm{out}}(\boldsymbol{\rho}) &= \left|E_\ell^{\mathrm{out}}(\boldsymbol{\rho})\right| \exp[\mathrm{i}\gamma_\ell^{\mathrm{out}}(\boldsymbol{\rho})] \\ &= \left|E_\ell^{\mathrm{out}}(\boldsymbol{\rho})\right| \exp[\mathrm{i}\phi_\ell^{\mathrm{out}}(\boldsymbol{\rho})] \exp[\mathrm{i}\psi^{\mathrm{out}}(\boldsymbol{\rho})], \end{aligned} \tag{3.20}$$

where ψ^{out} is a smooth wavefront phase that is common all three field components, and $\phi_\ell^{\mathrm{out}} = \gamma_\ell^{\mathrm{out}} - \psi^{\mathrm{out}}$ is the residual phase for each field component after the smooth phase is extracted. Similarly, the output field in the $\boldsymbol{\kappa}-$domain is written as:

$$\begin{aligned}\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa}) &= \left|\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})\right| \exp[\mathrm{i}\tilde{\gamma}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})] \\ &= \left|\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})\right| \exp[\mathrm{i}\tilde{\phi}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})] \exp[\mathrm{i}\tilde{\psi}^{\mathrm{out}}(\boldsymbol{\kappa})],\end{aligned} \tag{3.21}$$

where $\tilde{\psi}^{\mathrm{out}}$ is also a common smooth wavefront phase for each field component in $\boldsymbol{\kappa}-$domain.

According to the theory of the HFT [86], in a homeomorphic situation, the smooth phases of ψ^{out} and $\tilde{\psi}^{\mathrm{out}}$ dominate the Fourier transform so that the Fourier transform results in a bijective mapping from the $\boldsymbol{\rho}-$domain to the $\boldsymbol{\kappa}-$domain. The relationship between the wavefront phase in the two domains is shown in the following equations.

$$\tilde{\psi}^{\mathrm{out}}(\boldsymbol{\kappa}) = \psi^{\mathrm{out}}[\boldsymbol{\rho}(\boldsymbol{\kappa})] - \boldsymbol{\kappa} \cdot \boldsymbol{\rho}(\boldsymbol{\kappa}), \tag{3.22}$$

where $\boldsymbol{\rho}(\boldsymbol{\kappa})$ is a bijective mapping: $\boldsymbol{\rho} \leftrightarrow \boldsymbol{\kappa}$. The inverse Fourier transform also leads to a similar relation,

$$\psi^{\mathrm{out}}(\boldsymbol{\rho}) = \tilde{\psi}^{\mathrm{out}}[\boldsymbol{\kappa}(\boldsymbol{\rho})] + \boldsymbol{\rho} \cdot \boldsymbol{\kappa}(\boldsymbol{\rho}). \tag{3.23}$$

From Eqs. (3.22) and (3.23), one can see that, every iteration of IFTA will result in the same phase. Therefore, for the system with an output field that exhibits a homeomorphic situation, IFTA reaches a stagnant state.

3.2 Fourier pair synthesis: the homeomorphic case

3.2.1 Mapping-type Fourier pair synthesis

The last section discusses a stagnation problem for IFTA that happens in the homeomorphic case. However, if the Fourier pair exhibits a homeomorphic behavior, the stationary phase formula

$$\nabla\psi^{\mathrm{out}}(\boldsymbol{\rho}) = \boldsymbol{\kappa}(\boldsymbol{\rho}) \tag{3.24}$$

in the HFT method can be applied to calculate the phase for the field in the $\boldsymbol{\rho}$ domain by integrating the mapping function $\boldsymbol{\kappa}(\boldsymbol{\rho})$ directly. Therefore, in a homeomorphic situation, instead of an iterative approach that may have stagnation problems, the mapping relation can be applied to synthesize the Fourier pair in a single step.

Finding an appropriate mapping is a critical point. However, with the amplitudes in both domains known, the mapping can be solved by the well-known Parseval's equation,

$$\iint \left\|\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})\right\|^2 \mathrm{d}\boldsymbol{\rho} = \iint \left\|\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})\right\|^2 \mathrm{d}\boldsymbol{\kappa}. \tag{3.25}$$

where $\|\boldsymbol{E}\|$ denotes the L^2 norm of the field, $\|\boldsymbol{E}\| := \sqrt{E_x^2 + E_y^2 + E_z^2}$.

If we assume $\boldsymbol{\kappa}(\boldsymbol{\rho})$ is a one-to-one map, Eq. (3.25) can be derived to a differential equation:

$$\det[J(\boldsymbol{\kappa}(\boldsymbol{\rho}))] = \frac{\left\|\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})\right\|^2}{\left\|\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}(\boldsymbol{\rho}))\right\|^2}, \tag{3.26}$$

where $\det[J(\boldsymbol{\kappa}(\boldsymbol{\rho}))]$ is the determinant of the Jacobian matrix $J(\boldsymbol{\kappa}(\boldsymbol{\rho}))$.

Solving $\boldsymbol{\kappa}(\boldsymbol{\rho})$ from Eq. (3.26) is similar to a mathematical model, the L^2 Monge–Kantorovich problem, or the "optimal mass transport" (OMT) problem. Several numerical algorithms have been proposed to solve the OMT problem [68, 70] and result in a curl-free mapping. The detail of the algorithm will be explained in Section 3.4. Therefore, by applying the OMT model to Eq. (3.26), $\boldsymbol{\kappa}(\boldsymbol{\rho})$ can be solved. In addition, its curl-free characterization guarantees the smoothness of the $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ function.

Similarly, the resulting wavefront phase response (WPR) function for the element is calculated by

$$\Delta\psi(\boldsymbol{\rho}) = \psi^{\mathrm{out}}(\boldsymbol{\rho}) - \psi^{\mathrm{in}}(\boldsymbol{\rho}), \tag{3.27}$$

where $\psi^{\mathrm{in}}(\boldsymbol{\rho})$ is the smooth phase part of the input field.

So far, another mathematical model, a mapping-type Fourier pair synthesis, is introduced for light shaping in the homeomorphic situation [104]. The approach starts with a mapping relation between the fields in both domains and obtains a smooth common output wavefront phase for all three field components. As a result, the WPR function for the optical element can be obtained. The mapping-type approach not only tackles the stagnation problem, but also provides a smooth function without dislocations. In Eq. (3.26), one can see that the mapping-type Fourier pair synthesis method takes the full vectorial field into account.

3.2.2 Character of the mapping solution

In the mapping-type Fourier pair synthesis method, the Fourier pair is assumed to have a homeomorphic relationship where the Fourier transform can be accurately performed with the HFT method. The reconstructed wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is based on a bijective map between the $\boldsymbol{\rho}-$ and $\boldsymbol{\kappa}-$domain that is solved in Parseval's equation. Therefore, the question

arises if the output wavefront phase derived from Parseval's equation is proper for the HFT method. To this end, a mathematical proof is given in the following paragraphs.

The derivation of the output wavefront phase starts from assuming a smooth one-to-one map of $\boldsymbol{\kappa}(\boldsymbol{\rho})$ to Parseval's equation, such that Eq. (3.25) can be derived into its differential form, as shown in Eq. (3.26). The Jacobian matrix $J(\boldsymbol{\kappa}(\boldsymbol{\rho}))$ is written explicitly as $J(\boldsymbol{\kappa}(\boldsymbol{\rho})) = \begin{pmatrix} \partial k_x/\partial x & \partial k_x/\partial y \\ \partial k_y/\partial x & \partial k_y/\partial y \end{pmatrix}$. Its determinant is computed by $\det[J(\boldsymbol{\kappa}(\boldsymbol{\rho}))] = \dfrac{\partial k_x}{\partial x}\dfrac{\partial k_y}{\partial y} - \dfrac{\partial k_x}{\partial y}\dfrac{\partial k_y}{\partial x}$.

Therefore, Eq. (3.26) is derived as:

$$\begin{aligned} \left\| \tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}) \right\| &= \sqrt{\frac{1}{\det[J(\boldsymbol{\kappa}(\boldsymbol{\rho}))]}} \left\| \boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}) \right\| \\ &= \sqrt{\frac{1}{\dfrac{\partial k_x}{\partial x}\dfrac{\partial k_y}{\partial y} - \dfrac{\partial k_x}{\partial y}\dfrac{\partial k_y}{\partial x}}} \left\| \boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}) \right\|. \end{aligned} \tag{3.28}$$

The reconstructed wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is obtained according to the stationary formula in Eq. (3.24). Therefore, with Eqs. (3.28) and (3.24), the reconstructed wavefront phase yields a relation for the amplitudes in both domains,

$$\left\| \tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}) \right\| = \sqrt{\frac{1}{\psi_{xx}^{\mathrm{out}}(\boldsymbol{\rho})\psi_{yy}^{\mathrm{out}}(\boldsymbol{\rho}) - \psi_{xy}^{\mathrm{out}^2}(\boldsymbol{\rho})}} \left\| \boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}) \right\|. \tag{3.29}$$

where ψ_{xx}^{out}, ψ_{yy}^{out} and ψ_{xy}^{out} are the second derivative of ψ^{out}, e.g. $\psi_{xx} = \partial^2\psi/\partial x^2$.

This conclusion is the same as the one from the HFT. The HFT is performed analytically according to the formula in the paper of [86]:

$$\begin{aligned} \tilde{E}_\ell(\boldsymbol{\kappa}) &= \boldsymbol{\mathcal{F}}^{\mathrm{hom}}\left\{E_\ell(\boldsymbol{\rho})\right\} \\ &= \alpha(\boldsymbol{\rho}(\boldsymbol{\kappa}))\,|E_\ell(\boldsymbol{\rho}(\boldsymbol{\kappa}))|\exp[\mathrm{i}\phi_\ell(\boldsymbol{\rho}(\boldsymbol{\kappa}))]\exp[\mathrm{i}\psi(\boldsymbol{\rho}(\boldsymbol{\kappa})) - \mathrm{i}\boldsymbol{\kappa}\cdot\boldsymbol{\rho}(\boldsymbol{\kappa})], \end{aligned} \tag{3.30}$$

where

$$\alpha(\boldsymbol{\rho}) = \sqrt{\frac{1}{\psi_{xx}(\boldsymbol{\rho})\psi_{yy}(\boldsymbol{\rho}) - \psi_{xy}^2(\boldsymbol{\rho})}}, \tag{3.31}$$

is a weight factor for the amplitude. Although Eq. (3.30) is written in field component form, the weight factor $\alpha(\boldsymbol{\rho})$ is a common term for all three field components, which gives the same relationship for the amplitudes of $\boldsymbol{\rho}-$ and $\boldsymbol{\kappa}-$domain as in Eq. (3.29). Therefore, the reconstructed wavefront phase that is derived based on Parseval's equation gives the same

conclusion as HFT. Therefore, the resulting $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is the requested output wavefront phase for the Fourier pair synthesis under the homeomorphic assumption.

3.2.3 Numerical experiment

In this subsection, the algorithm with the homeomorphic assumption for the Fourier pair synthesis is demonstrated with an example. A simple light-shaping task is shown in Fig. 3.1. It attempts to design a smooth WPR function, modulating the input light and achieving a required energy distribution at the far-field zone. The source is a Gaussian wave, with a wavelength of $532\ \mathrm{nm}$ and a half divergence angle of $1°$. The target pattern is the portrait of Mr Ernst Abbe, which is given as an irradiance distribution with a size of $1 \times 1\ \mathrm{m}$ (shown in Fig. 3.4(a)). For a straightforward discussion, the target plane is set perpendicular to the optical axis and located $1\ \mathrm{m}$ away from the component.

The amplitudes of both the field $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})$ are prepared first for the Fourier pair synthesis. With the amplitude constraint in the $\boldsymbol{\rho}$ domain, the amplitude of $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})$ is equivalent to the amplitude of $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho})$ obtained by propagating the source field to the component plane. At the same time, by assuming a polarization state, for instance linearly polarized along x, the target field $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}')$ can be calculated from the given irradiance distribution by using Eqs. (3.10-3.11). $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}')$ exhibits a spherical phase function because it is located in the far-field zone. The propagation step is modeled inversely to obtain the field $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})$ by using Eq. (3.13). The mapping-type method works with the squared norm of both the field $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})$, which are shown in Fig. 3.4 (b) and (c), respectively.

To calculate $\boldsymbol{\kappa}(\boldsymbol{\rho})$ from Eq. (3.26) with the data in Fig. 3.4 (b) and (c), we modify the numerical algorithm proposed by Prins [70] to solve the equation. A dummy square-shape uniform density function $D(\boldsymbol{\xi})$ is introduced, so that the integration over its defined area is equal to both sides of Eq. (3.25). Instead of calculating $\boldsymbol{\kappa}(\boldsymbol{\rho})$ between $\left\|\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})\right\|^2$ and $\left\|\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})\right\|^2$ directly, we separately solve two mappings, $\boldsymbol{\kappa}(\boldsymbol{\xi})$ and $\boldsymbol{\rho}(\boldsymbol{\xi})$, between the dummy function and each squared norm of the amplitude. By doing so, the mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ is the combination of $\boldsymbol{\kappa}(\boldsymbol{\xi})$ and $\boldsymbol{\rho}(\boldsymbol{\xi})$, while the homeomorphism $\boldsymbol{\rho} \leftrightarrow \boldsymbol{\kappa}$ is still maintained. The detail of the algorithm will be introduced in Section 3.4. The same approach has been proposed in the spatial domain light shaping algorithm [105]. The resulting homeomorphism is shown in Fig. 3.4 (d) and (e), where the coordinates $\boldsymbol{\rho}$ and $\boldsymbol{\kappa}$ are sampled on the mesh nodes.

The mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ can be concluded from the two meshes by interpolating the sampling points of the meshes, with both data $k_x(\boldsymbol{\rho})$ and $k_y(\boldsymbol{\rho})$ shown in Fig. 3.5 (a) and (b), respectively. The wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is calculated with Eq. (3.24) by integrating the data of

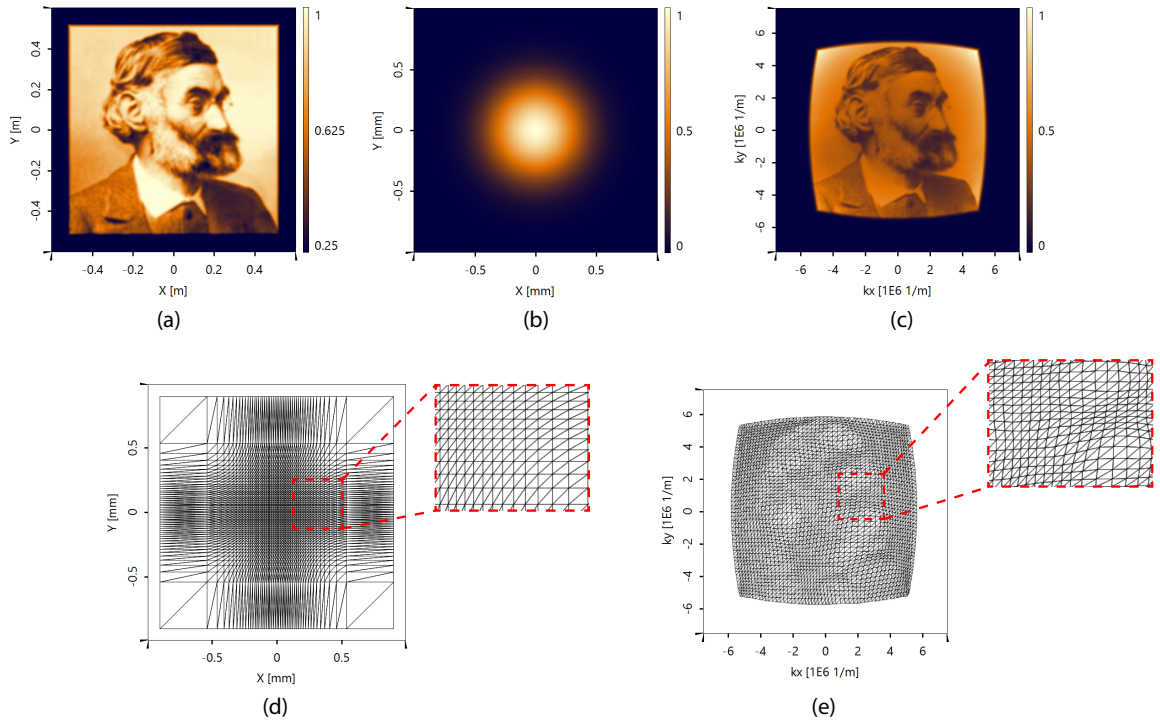

Fig. 3.4 An example for the discussion of the design algorithm. (a) Target irradiance distribution (normalized). The squared norm of the amplitude for the Fourier pair: (b) $\left\|\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})\right\|^2$ in $\boldsymbol{\rho}$ domain , and (c) $\left\|\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})\right\|^2$ in $\boldsymbol{\kappa}$ domain. (d)(e) The designed homeomorphism between the squared norms of the amplitude in both domains, with the mesh (d) for (b) and its mapping one (e) for (c).

the mapping. The B-spline technique [93] is applied here for the numerical integration, and as a result, the integrated wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is represented by the B-spline functions, as shown in Fig. 3.5 (c).

So far, the required wavefront phase for $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})$ is obtained by the Fourier pair synthesis with the homeomorphic assumption. The WPR function for the component is simply calculated following Eq. (3.27) by subtracting the wavefront phase of the input field from that of the output. The resulting $\Delta\psi(\boldsymbol{\rho})$ is shown in Fig. 3.5 (d), which is mainly a spherical phase function with additional modulation in Fig. 3.5 (e). Therefore, the functionality of the required component can be interpreted as shaping the phase of the input field, giving a spherical phase to dominate the field propagation to achieve the majority of the target area, and introducing an aberrant phase to redistribute the energy after propagation.

Physical-optics simulation Simulation with the calculated WPR function is done in the software VirtualLab Fusion [95] to investigate the designed result. In the field tracing

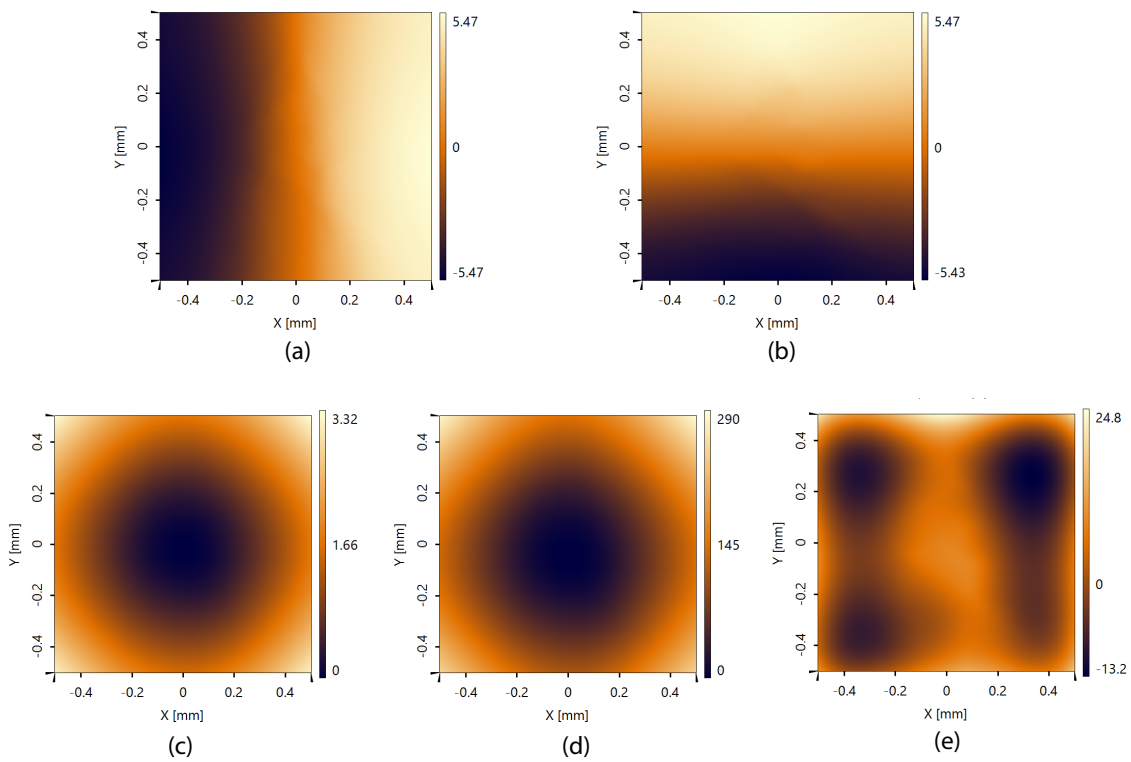

Fig. 3.5 The calculated mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ with (a) $k_x(\boldsymbol{\rho})$ and (b) $k_y(\boldsymbol{\rho})$ (unit: $10^6\mathrm{rad/m}$). (c) The required wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ resulting from the mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ (unit: $10^3\mathrm{rad}$). (d) The WPR function $\Delta\psi(\boldsymbol{\rho})$ for the optical element (unit: rad). (e) The residual phase after a spherical phase function fitting from (d) (unit: rad).

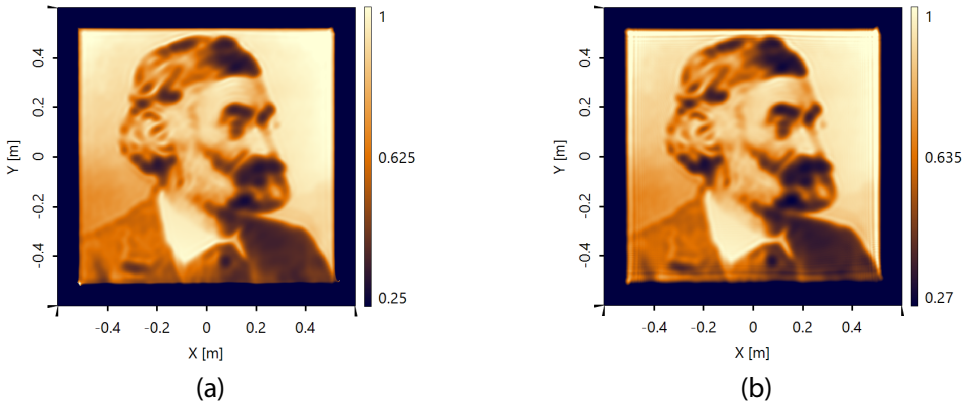

Fig. 3.6 Detector result: irradiance. Simulation with (a) HFT, (b) FFT.

engine, VirtualLab Fusion simulates the system with the techniques shown in the field tracing diagram in Fig. 3.1. A mathematical criterion in the background decides what kind of Fourier transform algorithm can be applied for the $\mathcal{F}$. In this case, the software determines that $\mathcal{F}$ is accurately calculated by the HFT. Therefore, the field behind the functional component in

this example indeed exhibits a homeomorphic behavior. An irradiance detector is set in the system, and the result is shown in Fig. 3.6 (a), which is coincident with the target pattern in Fig. 3.4 (a).

In addition, for more careful testing, we enforced the $\mathcal{F}$ to be performed by the FFT method. The result is shown in Fig. 3.6 (b). The result in Fig. 3.6 (a) is almost equivalent to the one in Fig. 3.6 (b), with only slight deviations at the edge that can be neglected, again revealing that the HFT is a valid choice in this example. Consequently, the homeomorphism assumption between the field $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})$ is valid, and applying the homeomorphic assumption in the Fourier pair synthesis yields an accurate design.

The smooth WPR function in Fig. 3.5 (d) can be used further for the element structure design, for example, a CGH or a freeform surface. The design and comparison of different element structures will be present in the Chapter 4.

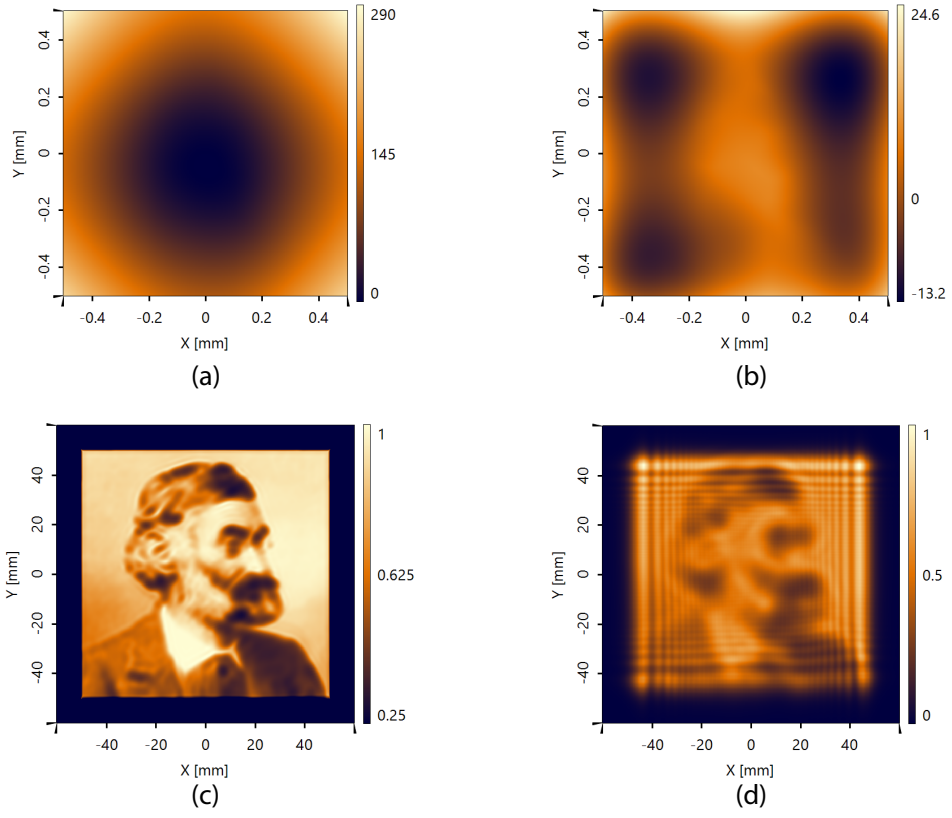

Fig. 3.7 (a) The WPR function $\Delta\psi(\boldsymbol{\rho})$ for the optical element (unit: rad). (b) The residual phase after a spherical function fitting of (a) (unit: rad). (c)(d) Detector result: irradiance. Simulation with (c) HFT, (d) FFT.

3.2.4 Strategy of Fourier pair synthesis

Another example where the homeomorphism does not exist is demonstrated for further investigation of the design algorithm. Here, the task is similar to the one shown above, with all the parameters in the task kept the same, except the size of the target pattern, which is reduced to 100×100 mm, so that the output is in a more paraxial situation. Following the same logic, the homeomorphic assumption is again made for the Fourier pair of the field behind the optical element. The WPR function is designed by applying the same mapping-type algorithm, as shown in Fig. 3.7 (a). Again, the algorithm results in a smooth element function. Fig. 3.7 (b) shows the residual phase in addition to a fitted spherical part. Compared with the previous example, one can see that the designed WPR function is similar except for the difference in their gradients.

Using the designed $\Delta\psi(\boldsymbol{\rho})$, the simulation is again performed via physical-optics modeling methods. With different Fourier transform algorithms, HFT v.s. FFT, the detector this time gives quite different results of irradiance distribution, as shown in Fig. 3.7 (c) and (d). For the simulation with HFT, because the same assumption is made in both design and modeling, the result coincides with the target pattern. However, by the analysis with rigorous FFT, Fig. 3.7 (d) shows obvious diffraction effects and lower resolution compared to the target pattern, which indicates that the designed function is not valid for the task. The homeomorphism assumption for the Fourier pair is far from a good approximation in this case.

In summary, we bring the homeomorphic assumption into the content of Fourier pair synthesis. The selection of the Fourier transform methods will lead to different strategies for the design. With both the examples above, it is demonstrated that the critical point for the mapping-type algorithm is whether the field behind the optical element is in its homeomorphic zone, or from a modeling point of view, whether the HFT can be accurately applied for the following operator. If so, typically in non-paraxial situations where the output field has a strong wavefront phase, the mapping-type design algorithm can provide a smooth WPR function. Otherwise, the designed result is not a valid function for the light-shaping task. Its validity should be investigated by physical-optics modeling. However, when the homeomorphic behavior is broken, IFTA can be applied properly. Moreover, the IFTA can benefit from the result of the homeomorphic algorithm, which is shown in what follows.

Because the homeomorphic assumption is not fulfilled in the second task, the IFTA is appropriate for the design. We start the IFTA with an initial guess given by the result of the homeomorphic algorithm. The IFTA converges after about 200 iterations. Fig. 3.8 shows the WPR function, with (a) the initial one and (b) the optimized one after IFTA, with both in a 2π modulo mode. The figures only show the residual part after subtracting the same

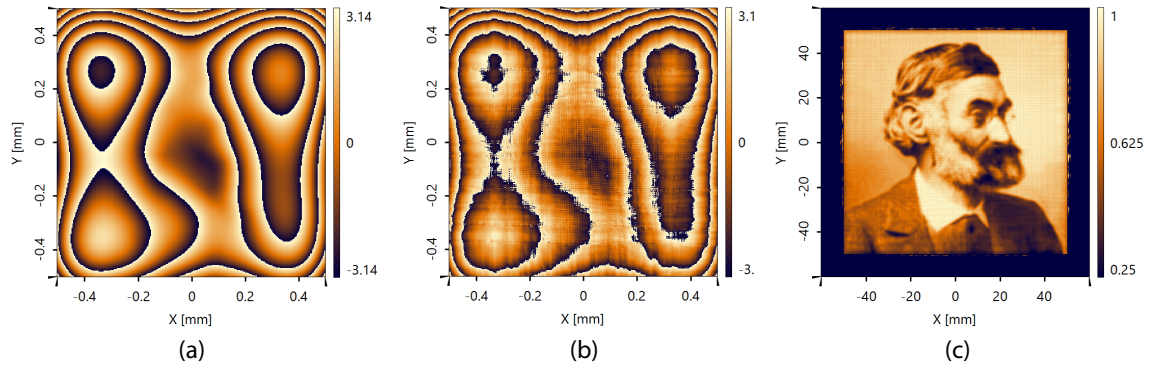

Fig. 3.8 The residual phase part after a spherical phase fitting of the WPR function: (a) the initial one and (b) the optimized one (in 2π modulo, unit: rad). (c) The detector result: irradiance, simulated with the optimized WPR function.

spherical phase. Fig 3.8 (c) shows the simulation result with the optimized function, which has improved resolution and no diffraction fringes compared to the irradiance distribution in Fig. 3.7 (d).

Aagedal et al. [97] showed that a well-introduced initial phase function might resolve the speckle problem in the IFTA. The speckle-free pattern, as shown in Fig 3.8 (c), indeed indicates that the smooth WPR function resulting from the homeomorphic algorithm is a good option to start. In such a paraxial situation, TEA can be applied to the final function after IFTA optimization for the structural design.

3.3 Integrability condition in mapping-type algorithms

In the above section, a mapping relation is assumed in the Fourier pair for attaining a smooth output wavefront phase. In literature, another mapping relation between the irradiance distribution is usually assumed in those geometric-optics-based algorithms also for achieving the required output wavefront phase.

The derivation of the output wavefront phase from the mapping is again under the stationary phase approximation [85], that the gradient of the output wavefront phase is connected with the mapping. For example, Feng et al. [106, 107] coupled the phase gradient and the mapping in a complex non-linear partial differential equation to solve the required output wavefront phase. However, if the mapping is concluded only from the irradiance relation without considering the output wavefront phase information, like in [28, 32, 33], the gradient data of the wavefront phase derived from the mapping is not necessarily integrable. A mathematical proof is provided in this section showing that the obtained gradient data of

the wavefront phase is integrable only in paraxial cases. A similar integrable problem is also well-known in the ray mapping method for the freeform surface design [36, 54, 72].

The mapping-type Fourier pair synthesis method indeed can be applied [104] to obtain a proper output wavefront phase without solving a complex differential equation. Instead of finding a mapping between the irradiance in the space domain, the mapping is assumed between the electric field in the space domain and the corresponding field in the spatial-frequency domain. In this scheme, the integrable gradient data of the output wavefront phase is obtained from the mapping, with no restriction in the paraxial situation. Therefore, a smooth output wavefront phase can be retrieved by direct integration.

In the following, both algorithms with different mapping strategies in the framework of field tracing techniques are presented and analyzed. The integrability condition in the algorithms is discussed mathematically. Numerical examples are shown to demonstrate and compare both approaches.

3.3.1 Mapping-type algorithm in the spatial domain

If a lossless system is assumed, the energy flux conserves in different plane though the system. Eq. (2.7) is recalled here for the following discussion,

$$\iint E_e^{\mathrm{in}}(\boldsymbol{\rho})\,\mathrm{d}\boldsymbol{\rho} = \iint E_e^{\mathrm{out}}(\boldsymbol{\rho})\,\mathrm{d}\boldsymbol{\rho} = \iint E_e^{\mathrm{tar}}(\boldsymbol{\rho}')\,\mathrm{d}\boldsymbol{\rho}', \tag{3.32}$$

where E_e^{in}, E_e^{out} are the irradiance distributions on the plane in front of and behind the optical element respectively, and E_e^{tar} is the distribution on the target plane.

In the literature, the homeomorphism in the design algorithms is calculated by the energy conservation law in the spatial domain. For instance, when the mapping is applied between the second and the last expressions in Eq. (3.32), the equation is derived into its differential form, which is also called the local energy conservation law,

$$\det[J(\boldsymbol{\rho}'(\boldsymbol{\rho}))] = \frac{E_e^{\mathrm{out}}(\boldsymbol{\rho})}{E_e^{\mathrm{tar}}(\boldsymbol{\rho}'(\boldsymbol{\rho}))}, \tag{3.33}$$

where $\det[J(\boldsymbol{\rho}'(\boldsymbol{\rho}))]$ is the determinant of the Jacobian matrix $J(\boldsymbol{\rho}'(\boldsymbol{\rho}))$.

Considering functional embodiment of the optical element, its structure is simply neglected but will be introduced as constrains in the later structural design step. Therefore, we define $E_e^{\mathrm{in}}(\boldsymbol{\rho})$ and $E_e^{\mathrm{out}}(\boldsymbol{\rho})$ on the same element plane by assuming the element function as a phase-only function, $E_e^{\mathrm{out}}(\boldsymbol{\rho}) = E_e^{\mathrm{in}}(\boldsymbol{\rho})$, where $E_e^{\mathrm{in}}(\boldsymbol{\rho})$ is obtained from the source. $E_e^{\mathrm{tar}}(\boldsymbol{\rho}')$

is the target irradiance given from the task. In general, $\boldsymbol{\rho}'(\boldsymbol{\rho})$ can be solved from Eq. (3.33) by the mathematical model, the L^2 Monge-Kantorovich problem.

By the method of stationary phase according to Bryngdahl [85], the gradient of the output wavefront phase is written as:

$$\nabla\psi^{\mathrm{out}}(\boldsymbol{\rho}) = k_0 n \frac{\boldsymbol{\rho}'(\boldsymbol{\rho}) - \boldsymbol{\rho}}{\sqrt{\|\boldsymbol{\rho}'(\boldsymbol{\rho}) - \boldsymbol{\rho}\|^2 + L^2}} \tag{3.34}$$

where k_0 is the wave number, n is the refractive index and L is the propagation distance between the optical element and the target plane.

The existence and the curl-free characterization of the solution for the L^2 Monge–Kantorovich problem are addressed in the theorem by Brenier [108]. However, due to the nonlinear relation in Eq. (3.34), $\nabla\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is not necessary a conservative vector field, even though $\boldsymbol{\rho}'(\boldsymbol{\rho})$ is. The integrability of $\nabla\psi^{\mathrm{out}}(\boldsymbol{\rho})$ only can be reserved in the paraxial approximation such that $\|\boldsymbol{\rho}'(\boldsymbol{\rho}) - \boldsymbol{\rho}\| \ll L$. In general, with the gradient data obtained from the mapping in the spatial domain, the required output wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ cannot be reconstructed by direct integration.

At the same time, the homeomorphism is assumed between the field of $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})$ for the algorithm with mapping type Fourier pair synthesis. The mapping is established between the coordinate of two domains, $\boldsymbol{\rho}$ and $\boldsymbol{\kappa}$, which is solved from Parsevel's equation (Eq. (3.25)), using the same L^2 Monge–Kantorovich model. The gradient of the output wavefront phase is equivalent to the mapping function, as shown in Eq. (3.24). Therefore, the curl-free characterization of the mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ guarantees that $\nabla\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is integrable in any cases. $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ can be obtained by direct integration, regardless of paraxial or non-paraxial, on-axis or off-axis situations.

In summary, both schemes with different homeomorphic assumptions for searching a required smooth output wavefront phase are presented. The mathematical derivation shows that the gradient information of the output wavefront phase calculated with the mapping from the irradiance in the spatial domain cannot satisfy the integrability condition. However, the proposed mapping-type Fourier pair synthesis directly results in integrable data for the output wavefront phase reconstruction.

3.3.2 Comparison of the integrability

A simple off-axis example with a high divergence angle, where the approach assuming mapping of irradiance in the spatial domain fails, is demonstrated for the later discussion, as shown in Fig. 3.9. The input is a plane wave propagating along the optical axis, with a

square aperture of 1×1 mm size. The optical element is required to shape the plane wave into a homogeneous rectangle pattern with a size of 1.2×0.8 m, located on a vertical plane 1 m away from the shaper, and the target pattern is 2 m offset distance in the y-direction.

The irradiance of both the input and target is a constant value. Therefore, it is straightforward to conclude that the homeomorphism of the irradiance is between two regular uniform grids, with the shape of a square and a rectangle, respectively. We sampled both grids with 301×201 points in the x- and y-dimensions. The output wavefront phase gradient data are calculated using Eq.(3.34), as shown in Fig. 3.9 (b) and (c).

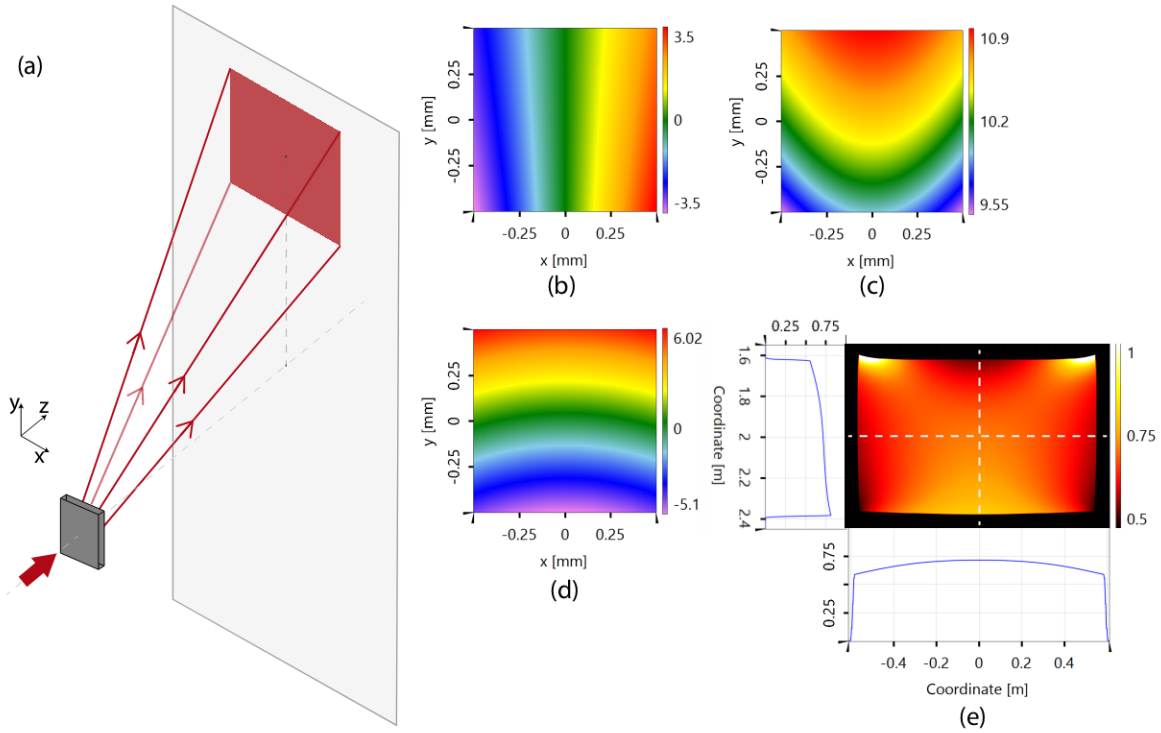

Fig. 3.9 (a) An off-axis light-shaping task, shaping a plane wave into a uniform off-axis rectangle pattern. (b)(c) The calculated output phase gradient data, with x- and y-components, respectively (Unit: $10^6 \mathrm{rad/m}$). (d) The calculated output wavefront phase by direct numerical integration method with the gradient data (Unit: $10^3 \mathrm{rad}$). (e) Irradiance (normalized): simulation result by field tracing, with the designed WPS function used.

The integrability of the resulting gradient data is determined by $\psi_{xy}^{\mathrm{out}} = \psi_{yx}^{\mathrm{out}}$[109, p. 458], where $\psi_{xy} \coloneqq \frac{\partial^2 \psi}{\partial x \partial y}$. Therefore, we use the relative root-mean-squared deviation (RMSD) of ψ_{xy}^{out} and ψ_{yx}^{out} to determine the integrability of the data, which is evaluated by

$$\delta = \sqrt{\frac{\sum_i \sum_j [\psi_{xy}^{\mathrm{out}}(\boldsymbol{\rho}_{ij}) - \psi_{yx}^{\mathrm{out}}(\boldsymbol{\rho}_{ij})]^2}{\sum_i \sum_j [\psi_{xy}^{\mathrm{out}}(\boldsymbol{\rho}_{ij})]^2}}. \tag{3.35}$$

The data from Fig. 3.9 (b) and (c) give $\delta = 40\%$.

To reconstruct the output wavefront phase function based on the data in Fig. 3.9 (b) and (c), the B-spline technique [93] is applied here for the numerical integration. This technique gives an optimal approximate smooth wavefront phase represented by B-spline functions, as shown in Fig. 3.9 (d).

To test if the resulting output phase is the one requested for the light shaping, which can be appropriately used later for the optical element design, the WPS function $\Delta\psi(\boldsymbol{\rho})$ is simply calculated by Eq. (3.27). $\psi^{\text{in}}(\boldsymbol{\rho})$ is the wavefront phase of the input field, which is a constant function in this case. A functional component that stores the WPS function of $\Delta\psi(\boldsymbol{\rho})$ is defined in the software VirtualLab Fusion [95]. Therefore, the obtained output wavefront phase from the design can be recalled in the simulation. The simulation is performed with the techniques shown in the field tracing diagram in Fig. 3.1. A detector set on the target plane gives the irradiance distribution shown in Fig. 3.9 (e). The shape and the inner irradiance distribution indicate that it deviates quite strongly from the targeted uniform rectangle pattern. Therefore, we can conclude that the approach that starts with mapping irradiance distribution in the spatial domain cannot provide an accurate output wavefront phase.

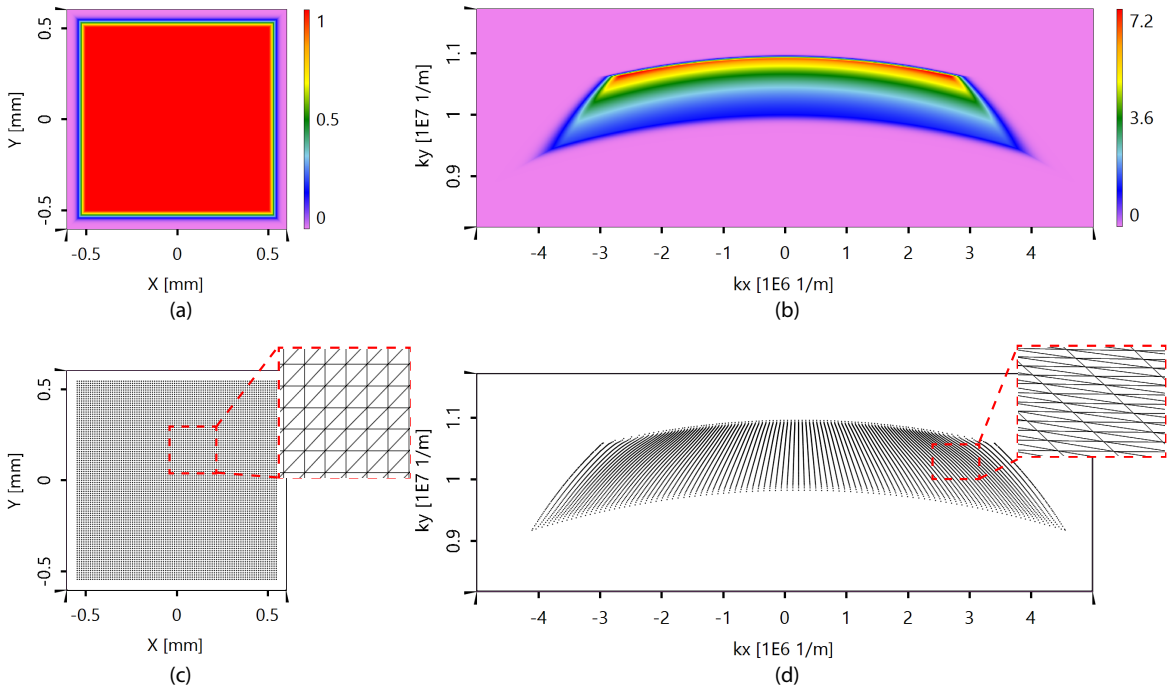

Fig. 3.10 The squared norm of the amplitude for the Fourier pair: (a) $\left\|\boldsymbol{E}^{\text{out}}(\boldsymbol{\rho})\right\|^2$ in $\boldsymbol{\rho}$ domain , and (b) $\left\|\tilde{\boldsymbol{E}}^{\text{out}}(\boldsymbol{\kappa})\right\|^2$ in $\boldsymbol{\kappa}$ domain. (c)(d) The designed homeomorphism between the squared norms of the amplitude in both domains, with the mesh (c) for (a) and its mapping (d) for (b).

For the other algorithm that establishes a mapping between the Fourier pair, the amplitude of the fields $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa})$, in both domains, is first prepared from the given source and target irradiance of the task. Fig. 3.10 (a) and (b) shows both obtained amplitudes of the Fourier pair . We implement the algorithm proposed by Prins [70] to calculate the mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ in Eq. 3.26. The homeomorphism $\boldsymbol{\rho} \leftrightarrow \boldsymbol{\kappa}$ is illustrated by two meshes shown in Fig. 3.10 (c) and (d), with the same number of sampling points 301×201 as in the previous approach.

Because the output wavefront phase gradient is equivalent to the mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ according to Eq. (3.24), it is obtained by applying an interpolation technique to the data of $\boldsymbol{\kappa}(\boldsymbol{\rho})$. Fig. 3.11 (a) and (b) show the gradient of $\psi^{\mathrm{out}}(\boldsymbol{\rho})$. The relative RMSD between ψ^{out}_{xy} and ψ^{out}_{yx} is $\delta = 0.4\%$, where the small value of indicates the gradient we obtained is integrable.

By direct integration, $\psi^{\mathrm{out}}(\boldsymbol{\rho})$ is shown in Fig. 3.11 (c). Similarly, we calculate the WPS function $\Delta\psi(\boldsymbol{\rho})$ for further investigation. Simulation with the resulting $\Delta\psi(\boldsymbol{\rho})$ gives the pattern on the irradiance detector shown in Fig. 3.11 (d). Here, the detected irradiance clearly indicates a uniform rectangle pattern that proves the required output wavefront phase is obtained accurately. The smooth output wavefront phase reconstructed by the mapping-type Fourier pair synthesis is also a promising result that can be used for the further structural design of the optical element.

The above numerical example is a simple one to illustrate the integrability issue in the output wavefront phase retrieval with the mapping from the irradiance relation. In contrast, the wavefront phase gradient obtained from the Fourier pair relation does not have an integrability problem. The latter approach for light shaping is general, with no restriction from the wavefront phase and shapes of the input field. Another example demonstrates a light-shaping problem for a circular shape spherical wave input.

As shown in Fig. 3.12 (a), the task is to shape a spherical wave to a rectangle pattern in the far-field zone. The full divergence angle of the spherical wave is $30°$, with a wavelength of $532\mathrm{nm}$ and a circular aperture. The target pattern is located on a vertical plane $1\ \mathrm{m}$ away from the optical element, with a size of $1.2 \times 0.8\ \mathrm{m}$. By the mapping-type Fourier synthesis method, the calculated gradient of the output wavefront phase is shown in Fig. 3.12 (b) and (c). The relative RMSD between the obtained ψ^{out}_{xy} and ψ^{out}_{yx} reaches $\delta = 0.4\%$, which indicates the gradient data of the output wavefront phase for this example are again integrable. The integrated smooth output wavefront phase is shown in Fig. 3.12 (d). Fig. 3.12 (e) shows the simulation result with the WPR function. The homogeneous irradiance distribution indicates that by the mapping-type Fourier pair synthesis method, the obtained output wavefront phase is accurate for deriving a WPR function for the task.

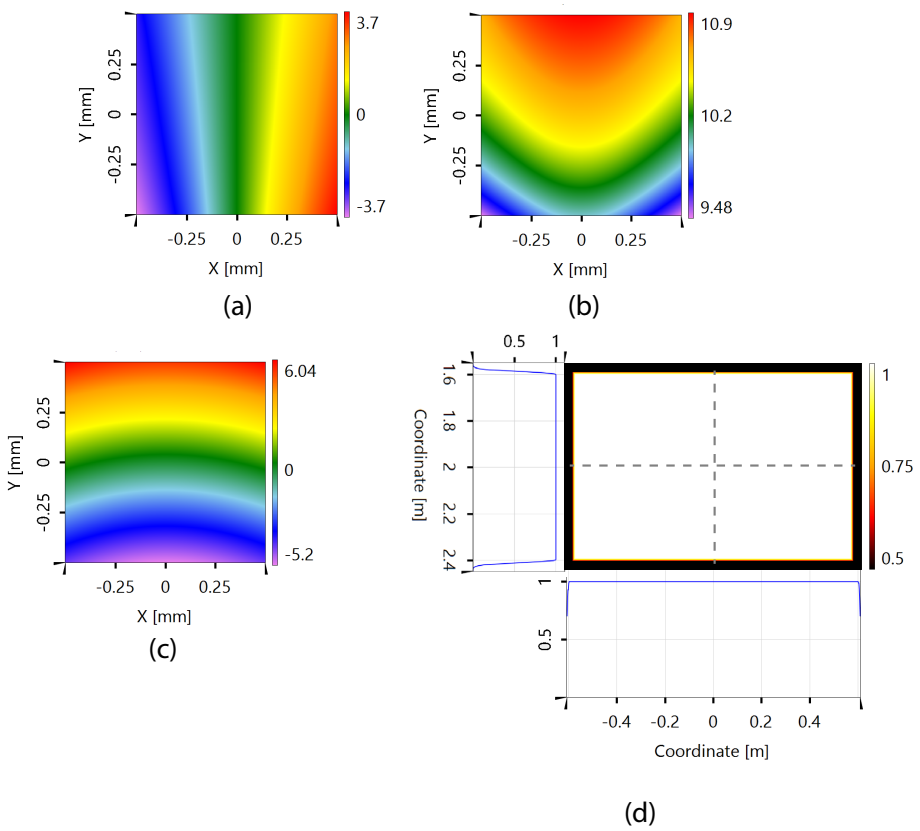

Fig. 3.11 The gradient data of the output wavefront phase calculated with the mapping type Fourier pair synthesis method, with (a) the x component of the gradient and (b) the y component of it (Unit: $10^6 \mathrm{rad/m}$). (c) The output wavefront phase integrated directly from its gradient data (Unit: $10^3 \mathrm{rad}$). (d) The simulated irradiance (normalized) using the designed WPS function.

So far, the algorithm for designing the functional embodiment of the light-shaping element is illustrated in detail. The functional embodiment helps in the later structural design and provides insight into whether the homeomorphic situation exists for the light-shaping system. The above numerical examples show that for those systems that exhibit homeomorphism, design algorithms with a mapping assumption obtain accurate results.

3.4 Homeomorphism design for the Fourier pair

Solving a one-to-one map between two functions connected with a Monge–Ampère equation is usually required in the light-shaping algorithms. For example, in the mapping-type Fourier synthesis method, the mapping has solved the equation of the fields between the $\boldsymbol{\rho}-$ and $\boldsymbol{\kappa}-$domain. The mathematical model of the L^2 Monge–Kantorovich problem is widely

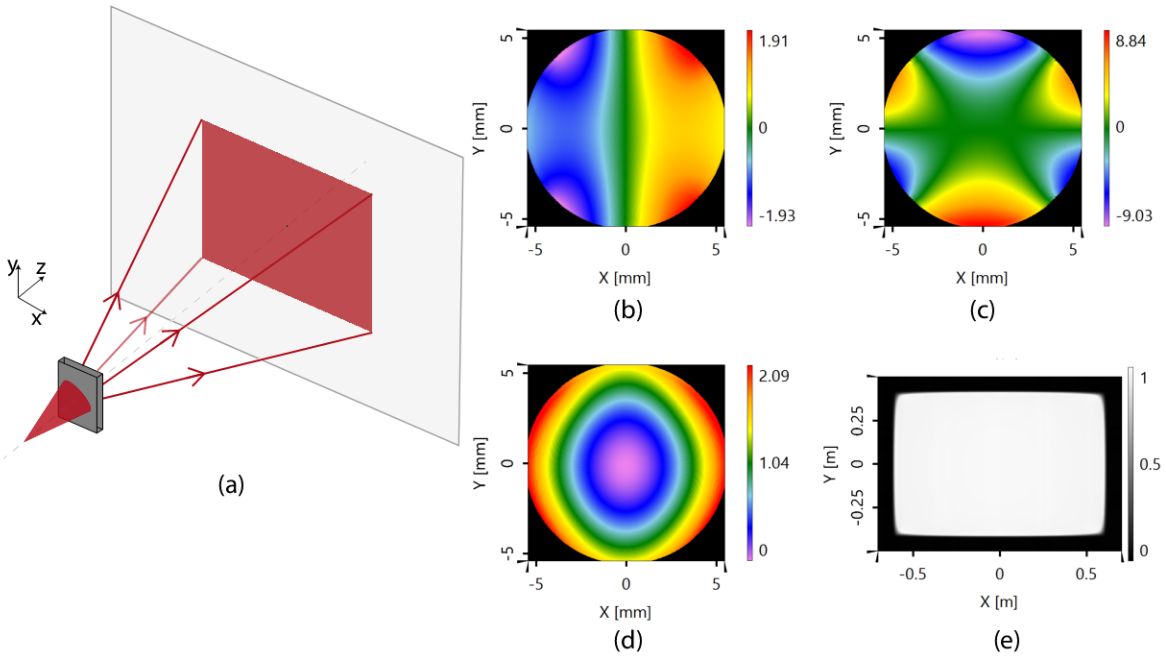

Fig. 3.12 (a) An non-paraxial light-shaping task, shaping a spherical wave with a circular aperture into a uniform rectangle pattern. (b)(c) The calculated output wavefront phase gradient data, with x- and y-component respectively (Unit: $10^6 \mathrm{rad/m}$ and $10^5 \mathrm{rad/m}$ respectively). (d) The calculated output wavefront phase by direct numerical integration method with the gradient data (Unit: $10^4 \mathrm{rad}$). (e) Irradiance distribution (normalized): simulation result with the WPS function.

used to solve the mapping. The mathematical model was first introduced for transporting material from one location to another at a minimal cost. Therefore, it is often referred to as the optimal mass transport (OMT) problem. It was extensively studied by Monge in 1781 and by Kantorovich in 1942, so their theories are known as the Monge–Kantorovich theory of optimal mass transport [110]. Different numerical approaches have been proposed for solving the problem in recent decades. For instance, Haker et al. [67] proposed an elastic registration method that is solved via an optimal warping based on deriving a partial differential equation to minimize the L^2 Kantorovich–Wasserstein distance under a mass preservation constraint. Sulman et al. [68] presented an approach based on solving a logarithmic parabolic Monge–Ampère equation. Prisn et al. [70] developed a least-squares method with a proper boundary condition to solve the normal Monge–Ampère equation from the problem directly. The kernel of our method is implied by the approach proposed by Prins et al. However, an iterative process is introduced to improve the convergence to an integrable solution. In this section, the algorithm is presented in a purely mathematical manner, and numerical experiments investigate its validity.

3.4.1 L^2 Monge–Kantorovich problem

The original Monge–Kantorovich optimal transport problem is reviewed in this section. Although the original problem can be defined in arbitrary dimensions, it is introduced and discussed in a 2D situation because it is usually applied to 2D problems in light shaping.

Let the support Ω_1 and Ω_2 denote a subset of $\mathbb{R}^2$ and the scalar density functions be defined as $D_1 : \Omega_1 \to \mathbb{R}$ and $D_2 : \Omega_2 \to \mathbb{R}$ on the supports of Ω_1 and Ω_2, respectively. In the original mathematical model, these density functions indicate the density of a material, where the integral of the functions on a given domain is the mass. In the application in optics, the density functions are usually interpreted as the unit energy quantities, that is, irradiance. It is assumed that the integrals of the functions over their 2D domains are equivalent:

$$\iint_{\Omega_1} D_1(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \iint_{\Omega_2} D_2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}, \tag{3.36}$$

where their coordinates are defined in 2D, with $\boldsymbol{x} = (x, y)$ and $\boldsymbol{u} = (u, v)$.

The transport problem attempts to find a mapping, $\boldsymbol{M}(\boldsymbol{x}) : \Omega_1 \to \Omega_2$, that each sample of $\boldsymbol{x} \in \Omega_1$ is moved to $\boldsymbol{u} = \boldsymbol{M}(\boldsymbol{x}) \in \Omega_2$. If the mapping $\boldsymbol{M}(\boldsymbol{x})$ is a smooth one-to-one map, Eq. (3.36) can be derived into its Jacobian equation.

$$\det\left(\underline{\boldsymbol{J}_{\boldsymbol{M}}}\right) = \frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}(\boldsymbol{x}))}, \tag{3.37}$$

where $\det\left(\underline{\boldsymbol{J}_{\boldsymbol{M}}}\right)$ is the determinant of the Jacobian matrix that describes the distortion of the local area introduced by the mapping $\boldsymbol{M}(\boldsymbol{x})$. The bold underlined notation "$\underline{\boldsymbol{A}}$" indicates the variable $\boldsymbol{A}$ is a matrix.

The solution of Eq. (3.37) is not unique. However, under an additional closure condition, a unique solution can be found. A cost functional is induced as the closure,

$$C(\boldsymbol{M}) = \iint_{\Omega_1} \|\boldsymbol{x} - \boldsymbol{M}(\boldsymbol{x})\|^2 D_1(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{3.38}$$

Therefore, the one-to-one map that satisfied Eq. (3.37) and minimized the cost functional of Eq. (3.38) is then called the *optimum solution* of the L^2 Monge–Kantorovich problem.

3.4.2 Algorithm solving the L^2 Monge–Kantorovich problem

The algorithm proposed by Prins et al. [70] is modified in this work to solve the L^2 Monge–Kantorovich problem. The essential modification are mainly in two aspects:

1. An integration step is introduced in each iteration of the algorithm. The obtained potential function of the mapping is then differentiated and used as the input for the next iteration. In this way, the algorithm converges to an integrable mapping faster.

2. A dummy square-shape uniform density function is introduced to map both the given density functions, a more flexible way to tackle problems with an arbitrary boundary of the functions' support.

The algorithm in [70] is a least-squares method. The method numerically solves the Jacobian equation by minimizing one functional over the set of symmetric matrices to find the Jacobian matrix that satisfies Eq. (3.37), and minimizing another functional for the boundary condition. The minimization is alternatively performed over the Jacobian matrix, the boundary condition and the mapping. The minimization over the Jacobian matrix and the boundary condition is done with a fixed mapping, and the one over the mapping is done with a fixed Jacobian matrix and boundary condition.

The algorithm starts with an initial map $\boldsymbol{M}^0$. The whole iterative algorithm adapts the mapping so that its Jacobian matrix will satisfy Eq. (3.37).

To correct the mapping, we require the Jacobian matrix of $\boldsymbol{M}^0$, which is given by

$$\underline{\boldsymbol{J}_{\boldsymbol{M}^0}} = \begin{bmatrix} \dfrac{\partial u^0}{\partial x} & \dfrac{\partial u^0}{\partial y} \\ \dfrac{\partial v^0}{\partial x} & \dfrac{\partial v^0}{\partial y} \end{bmatrix}, \tag{3.39}$$

equals to a positive semidefinite matrix $\underline{\boldsymbol{P}}(\boldsymbol{x})$, where $\underline{\boldsymbol{P}}(\boldsymbol{x})$ comes from the set that satisfies

$$\det(\underline{\boldsymbol{P}}(\boldsymbol{x})) = \frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}^0(\boldsymbol{x}))}. \tag{3.40}$$

Note that in [70], the matrix $\underline{\boldsymbol{P}}(\boldsymbol{x})$ is set as a symmetric matrix $\underline{\boldsymbol{P}} = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}$, with its components along one diagonal are equivalent. The matrix form is set to converge the algorithm to an integrable mapping, because the Jacobian of an integrable mapping is a symmetric matrix in this form. However, because we introduce a integration step in each iteration of the algorithm, the additional step may leads the algorithm converge to an integrable solution. The matrix form of $\underline{\boldsymbol{P}}(\boldsymbol{x})$ is not necessary symmetric in our algorithm.

The matrix $\underline{\boldsymbol{P}}(\boldsymbol{x})$ is searched by minimizing a functional defined by

$$\Delta_J(\boldsymbol{M}^0, \underline{\boldsymbol{P}}) = \frac{1}{2} \iint_{\Omega_1} \left\| \underline{\boldsymbol{J}_{\boldsymbol{M}^0}} - \underline{\boldsymbol{P}} \right\|^2 \mathrm{d}\boldsymbol{x}, \tag{3.41}$$

where the Fröbenius norm is defined as $\|\underline{\boldsymbol{A}}\| = \sqrt{\sum_{ij}(A_{ij}^2)}$.

The algorithm also includes the boundary condition that the mapping of $\partial\Omega_1$ should be inside $\partial\Omega_2$, where $\partial\Omega_1$ and $\partial\Omega_2$ are the boundary of supports Ω_0 and Ω_1 respectively. With the initial mapping $\boldsymbol{M}^0$, the boundary condition is addressed by a second functional

$$\Delta_b(\boldsymbol{M}^0, \boldsymbol{M_b}) = \frac{1}{2}\oint_{\partial\Omega_1} \left|\boldsymbol{M}^0 - \boldsymbol{M_b}\right|^2 \mathrm{d}s. \tag{3.42}$$

When minimizing Eq. (3.42) over $\boldsymbol{M_b}$ from the set $\boldsymbol{M_b}(\boldsymbol{x}) \in \partial\Omega_1 \ \forall \boldsymbol{x} \in \partial\Omega_0$, the boundary condition of the mapping is satisfied.

Obtaining the matrix $\underline{\boldsymbol{P}}$ and the boundary points $\boldsymbol{M_b}$, the mapping is found by minimizing the following functional

$$\Delta(\boldsymbol{M}^0, \underline{\boldsymbol{P}}, \boldsymbol{M_b}) = (1-\alpha)\oint_{\partial\Omega_1} \left|\boldsymbol{M}^0 - \boldsymbol{M_b}\right|^2 \mathrm{d}s + \alpha \iint_{\Omega_1} \left\|\underline{\boldsymbol{J}_{\boldsymbol{M}^0}} - \underline{\boldsymbol{P}}\right\|^2 \mathrm{d}\boldsymbol{x}. \tag{3.43}$$

where the parameter α controls the weight of Δ_J to Δ_b.

So far, an updated mapping $\boldsymbol{M}^1(\boldsymbol{x})$ is concluded which can be used as the predefined input for the next iteration. However, one further procedure is introduced to improve the integrability of the mapping. Herein, a potential function $\phi(\boldsymbol{x})$ is calculated that minimizing a functional defined by

$$\Delta_\phi(\phi, \boldsymbol{M}^1) = \frac{1}{2}\iint_{\Omega_1} \left\|\nabla\phi - \boldsymbol{M}^1\right\|^2 \mathrm{d}\boldsymbol{x}. \tag{3.44}$$

Because the obtained mapping $\boldsymbol{M}^1(\boldsymbol{x})$ from the previous procedures may not be a conservative vector field, the minimization of Eq. (3.44) is aimed at searching for an optimum potential function for the mapping data. The numerical optimization is performed with the B-spline functions, where the $\phi(\boldsymbol{x})$ is represented as B-spline functions while the parameters of the function are found by solving Eq. (3.44) with its gradient formulas.

After $\phi(\boldsymbol{x})$ is obtained, its gradient $\nabla\phi(\boldsymbol{x})$ is used as the predefined mapping for the next iteration, with $\boldsymbol{M}^{1'}(\boldsymbol{x}) = \nabla\phi(\boldsymbol{x})$. Note that $\boldsymbol{M}^{1'}(\boldsymbol{x})$ is naturally an integrable vector field. Without knowing a strict mathematical proof, we have found that this step helps the algorithm converge to an integrable solution quickly. The reason for the convergence may be that in each iteration, an integrable mapping is set to start with.

In summary, the algorithm starts with an initial mapping $\boldsymbol{M}^0$ and performs the iteration subsequently:

$$\boldsymbol{M}_{\boldsymbol{b}}^n = \arg\min_{\boldsymbol{M}_{\boldsymbol{b}}} \Delta_b(\boldsymbol{M}^n, \boldsymbol{M}_{\boldsymbol{b}}), \tag{3.45a}$$

$$\underline{\boldsymbol{P}}^n = \arg\min_{\underline{\boldsymbol{P}}} \Delta_J(\boldsymbol{M}^n, \underline{\boldsymbol{P}}), \tag{3.45b}$$

$$\boldsymbol{M}^{n+1} = \arg\min_{\boldsymbol{M}} \Delta(\boldsymbol{M}^n, \boldsymbol{M}_{\boldsymbol{b}}^n, \underline{\boldsymbol{P}}^n), \tag{3.45c}$$

$$\phi = \arg\min_{\phi} \Delta_\phi(\phi, \boldsymbol{M}^{n+1}), \tag{3.45d}$$

$$\boldsymbol{M}^{n+1'} = \nabla\phi. \tag{3.45e}$$

Initializing mapping The initial mapping $\boldsymbol{M}^0(\boldsymbol{x})$ is set simply as a scaled map to the sampling of $D_1(\boldsymbol{x})$. Let's assume the sampling of $D_1(\boldsymbol{x})$ is $(x_{i,j}, y_{i,j})$ for $i = 1, 2, 3...m$ and $j = 1, 2, 3...n$ where M and N are the number of sampling points in each dimension. Then the initial map results in $(u_{i,j}, v_{i,j}) = \boldsymbol{M}^0(\boldsymbol{x}) = (ax_{i,j}, by_{i,j})$. The scaling factor a and b are chosen that the mapping coordinate (u, v) covers the support area of D_2.

Minimizing procedure for $\boldsymbol{M_b}$ With the known mapping $\boldsymbol{M}(\boldsymbol{x})$ in the current iteration, the boundary map $\boldsymbol{M_b}$ are found by projecting the boundary samples of $\boldsymbol{M}(\boldsymbol{x})$ to the support boundary of D_2. The support boundary $\partial\Omega_2$ is discretized as a polygon. The higher the number of line segments of the polygon, the more accurate the boundary $\partial\Omega_2$ is sampled. Then each sample of $\boldsymbol{M}(\boldsymbol{x}) \in \partial\Omega_2$ is projected by finding the position with the minimum distance to the polygon boundary. The obtained positions of all $\boldsymbol{M}(\boldsymbol{x}) \in \partial\Omega_2$ samples are set as $\boldsymbol{M_b}$. The process is done pointwise, and it is the same as the process described in Section 3.1 in [70].

A preprocessing is added for the first iteration. In the later iterations, the process illustrated here is not needed. In the first iteration, before starting to adapt the mapping to the density function $D_2(\boldsymbol{x})$, the interior samples of the initial mapping are simply morphed according to the projection of the boundary. The positions of the interior samples are calculated by

$$u_{i,j} = u_{1,j} + \frac{u_{m,j} - u_{1,j}}{m-1}, \tag{3.46a}$$

$$v_{i,j} = v_{i,1} + \frac{v_{i,n} - v_{i,1}}{n-1}. \tag{3.46b}$$

In this way, all the mapping samples $\boldsymbol{M}(\boldsymbol{x})$ are brought into the support area of D_2[1] and used as a better-suited initial map for the coming procedures.

Minimizing procedure for $\underline{\boldsymbol{P}}$ The $\underline{\boldsymbol{P}}$ are calculated analytically in [70], where the detailed mathematical derivation has been provided. The main idea of the derivation is reviewed as follows.

The procedure is performed pointwise. The Jacobian of the mapping in the current iteration is defined as

$$D\boldsymbol{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}, \tag{3.47}$$

with $m_{11} = \frac{\partial u}{\partial x}$, $m_{12} = \frac{\partial u}{\partial y}$, $m_{21} = \frac{\partial v}{\partial x}$, and $m_{22} = \frac{\partial v}{\partial y}$. $\underline{\boldsymbol{P}}$ is written explicitly as

$$\underline{\boldsymbol{P}} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}. \tag{3.48}$$

Note that $\underline{\boldsymbol{P}}$ is not necessarily defined as a symmetric matrix in our algorithm, whereas a symmetric matrix is used in the work of [70].

Next, the deviation between the matrices is defined by the function

$$H(p_{11}, p_{12}, p_{21}, p_{22}) = \left\| \underline{\boldsymbol{P}} - D\boldsymbol{M} \right\|^2. \tag{3.49}$$

Therefore, the minimization problem of Eq. (3.41) is then derived as

$$\underset{(p_{11},p_{12},p_{21},p_{22})\in\mathbb{R}^4}{\arg\max} \left\{ H(p_{11}, p_{12}, p_{21}, p_{22}) \middle| p_{11}p_{22} - p_{12}p_{21} = \frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}(\boldsymbol{x}))} \right\}. \tag{3.50}$$

The minimization of Eq. (3.50) is possibly found at the critical point of the Lagrange function

$$\Lambda(p_{11}, p_{12}, p_{21}, p_{22}, \lambda) = \frac{1}{2} \left\| \underline{\boldsymbol{P}} - D\boldsymbol{M} \right\|^2 + \lambda(\det(\underline{\boldsymbol{P}}) - \frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}(\boldsymbol{x}))}). \tag{3.51}$$

[1] This statement is true for most cases of D_2 with an arbitrary boundary shape. For those cases that the procedure cannot bring all the samples into the support area, the continuous procedures of the algorithm still can be performed.

where λ is the Lagrange multiplier. Setting all partial derivatives of Λ to be 0, we obtain the following algebraic system:

$$p_{11} + \lambda p_{22} = m_{11}, \tag{3.52a}$$
$$p_{12} - \lambda p_{21} = m_{12}, \tag{3.52b}$$
$$p_{21} - \lambda p_{12} = m_{21}, \tag{3.52c}$$
$$p_{22} + \lambda p_{11} = m_{22}, \tag{3.52d}$$
$$p_{11}p_{22} - p_{12}p_{21} = \frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}(\boldsymbol{x}))}. \tag{3.52e}$$

From the equations system in Eq. (3.52), each component of $\underline{\boldsymbol{P}}$ is reorganized as

$$p_{11} = \frac{\lambda m_{22} - m_{11}}{\lambda^2 - 1}, p_{12} = \frac{m_{12} + \lambda m_{21}}{1 - \lambda^2}, p_{21} = \frac{\lambda m_{12} + m_{21}}{1 - \lambda^2}, p_{22} = \frac{\lambda m_{11} - m_{22}}{\lambda^2 - 1}. \tag{3.53}$$

Substituting these expressions into Eq. (3.52e) yields the equation

$$a_4\lambda^4 + a_2\lambda^2 + a_1\lambda + a_0 = 0. \tag{3.54}$$

with

$$\begin{aligned}
a_4 &= \frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}(\boldsymbol{x}))}, \\
a_2 &= -2\frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}(\boldsymbol{x}))} - \det(D\boldsymbol{M}), \\
a_1 &= \|D\boldsymbol{M}\|^2, \\
a_0 &= \frac{D_1(\boldsymbol{x})}{D_2(\boldsymbol{M}(\boldsymbol{x}))} - \det(D\boldsymbol{M}).
\end{aligned}$$

Different considerations for solving λ have been discussed in detail in [70], which are followed in this work. Solving Eq. (3.54) leads to more than one possible result. Substitute all the possible solutions to Eq. (3.49) and select the minimizer so that the matrix $\underline{\boldsymbol{P}}$ is obtained through the processes.

Minimizing procedure for $\boldsymbol{M}$ Unlike the previous procedures, the minimizing for $\boldsymbol{M}$ can not be performed pointwise. Prins et al. [70] have derived a set of Poisson equations to solve the problem.

With Eq. (3.43), the first variation of Δ with respect to $\boldsymbol{M}$ is calculated as follows,

$$\begin{aligned}\delta\Delta(\boldsymbol{M},\underline{\boldsymbol{P}},\boldsymbol{M_b})(\boldsymbol{\eta}) &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\Delta(\boldsymbol{M}+\epsilon\boldsymbol{\eta},\underline{\boldsymbol{P}},\boldsymbol{M_b})-\Delta(\boldsymbol{M},\underline{\boldsymbol{P}},\boldsymbol{M_b})\right]\\ &= \lim_{\epsilon\to 0}\left[\frac{\alpha}{2}\iint_{\Omega_1} 2(D\boldsymbol{M}-\underline{\boldsymbol{P}})\colon D\boldsymbol{\eta}+\epsilon\left\|D\boldsymbol{\eta}\right\|^2\mathrm{d}\boldsymbol{x}\right.\\ &\qquad\left.+\frac{1-\alpha}{2}\oint_{\partial\Omega_1} 2(\boldsymbol{M}-\boldsymbol{M_b})\cdot\boldsymbol{\eta}+\epsilon\left|\boldsymbol{\eta}\right|^2\mathrm{d}s\right]\\ &= \alpha\iint_{\Omega_1}(D\boldsymbol{M}-\underline{\boldsymbol{P}})\colon D\boldsymbol{\eta}\mathrm{d}\boldsymbol{x}+(1-\alpha)\oint_{\partial\Omega_1}(\boldsymbol{M}-\boldsymbol{M_b})\cdot\boldsymbol{\eta}\mathrm{d}s.\end{aligned} \tag{3.55}$$

where $\boldsymbol{\eta}$ belongs to the set of two-dimension vector fields. The notation of $\underline{\boldsymbol{A}}\colon\underline{\boldsymbol{B}}$ denotes the inner product of two matrices, that is, $\underline{\boldsymbol{A}}\colon\underline{\boldsymbol{B}}=\sum_{ij}A_{ij}B_{ij}$. The minimizer is given by

$$\delta\Delta(\boldsymbol{M},\underline{\boldsymbol{P}},\boldsymbol{M_b})(\boldsymbol{\eta})=0,\qquad\forall\boldsymbol{\eta}\in\mathbb{R}^2. \tag{3.56}$$

Applying Gaussian's theorem and fundamental lemma of the calculus of variations [111], the above equation is derived into two decoupled Poisson equations with Robin boundary conditions:

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=\frac{\partial p_{11}}{\partial x}+\frac{\partial p_{12}}{\partial y}\qquad (x,y)\in\Omega_1\setminus\partial\Omega_1 \tag{3.57a}$$

$$(1-\alpha)u+\alpha\nabla u\cdot\hat{\boldsymbol{n}}=(1-\alpha)M_{bx}+\alpha\boldsymbol{P}_1\cdot\hat{\boldsymbol{n}}\qquad (x,y)\in\partial\Omega_1 \tag{3.57b}$$

and

$$\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}=\frac{\partial p_{12}}{\partial x}+\frac{\partial p_{22}}{\partial y}\qquad (x,y)\in\Omega_1\setminus\partial\Omega_1 \tag{3.58a}$$

$$(1-\alpha)v+\alpha\nabla v\cdot\hat{\boldsymbol{n}}=(1-\alpha)M_{by}+\alpha\boldsymbol{P}_2\cdot\hat{\boldsymbol{n}}\qquad (x,y)\in\partial\Omega_1 \tag{3.58b}$$

where we define $\boldsymbol{P}_1=(p_{11},p_{12})$, $\boldsymbol{P}_2=(p_{12},p_{22})$, and $\boldsymbol{M_b}=(M_{bx},M_{by})$ is the Cartesian coordinate of the boundary points. $\hat{\boldsymbol{n}}$ is the outward-pointing unit normal vector on $\partial\Omega_1$.

A numerical method is implemented to solve the mapping $(u,v)=\boldsymbol{M}(x,y)$ from the Poisson equations. The equations are discretized, and the five-point finite difference scheme is used to calculate all the derivatives in Eqs. (3.57) and Eqs. (3.58).

As illustrated in Fig. 3.13(a), for the interior points, the first order and second order derivative are defined as:

$$\frac{\partial z_{i,j}}{\partial x} = \frac{z_{i+1,j} - z_{i-1,j}}{2\Delta x}, \qquad \frac{\partial z_{i,j}}{\partial y} = \frac{z_{i,j+1} - z_{i,j-1}}{2\Delta y}, \tag{3.59}$$

$$\frac{\partial^2 z_{i,j}}{\partial x^2} = \frac{z_{i-1,j} - 2z_{i,j} + z_{i+1,j}}{\Delta x^2}, \qquad \frac{\partial^2 z_{i,j}}{\partial y^2} = \frac{z_{i,j-1} - 2z_{i,j} + z_{i,j+1}}{\Delta y^2}. \tag{3.60}$$

The forward (or backward) difference approximation scheme is applied to calculate the first derivative of the boundary points. For example, as shown in Fig. 3.13(b), the backward scheme is taken for calculating the $x-$derivative of the boundary points at x_{max} positions,

$$\frac{\partial z_{m,j}}{\partial x} = \frac{z_{m,j} - z_{m-1,j}}{\Delta x}, \qquad \frac{\partial z_{m,j}}{\partial y} = \frac{z_{m,j+1} - z_{m,j-1}}{2\Delta y}. \tag{3.61}$$

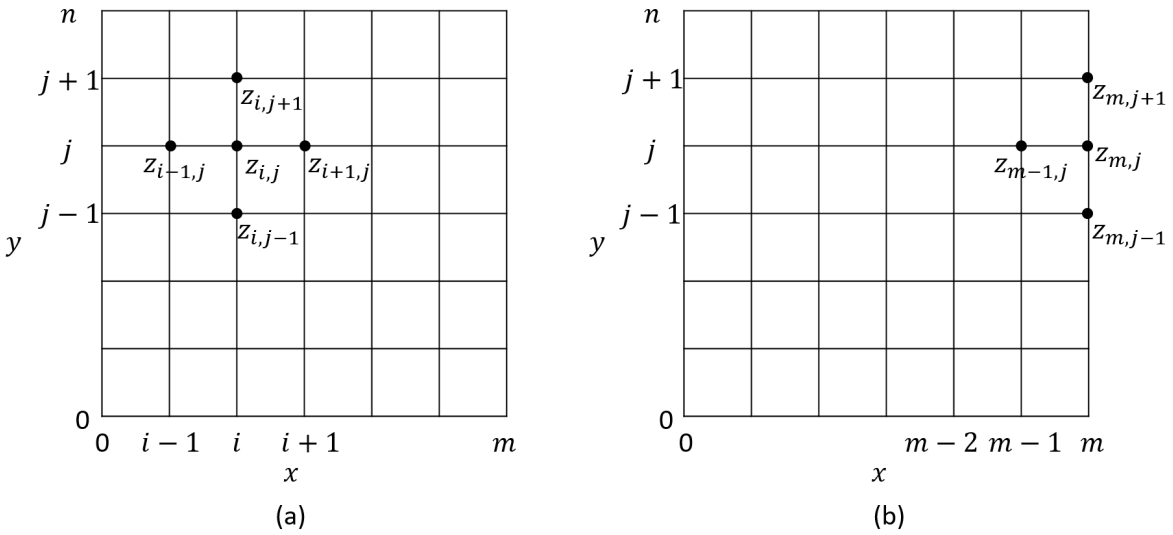

Fig. 3.13 (a) The 5-point finite difference scheme for the interior points and (b) the forward (or backward) difference approximation for the boundary points.

With this finite difference scheme, the Poisson equations can be derived into a set of linear equations. The linear equations system is then reorganized in matrix form, written as:

$$\underline{\boldsymbol{A}}\boldsymbol{u}^{\text{vec}} = \boldsymbol{F_1}, \tag{3.62a}$$

$$\underline{\boldsymbol{B}}\boldsymbol{v}^{\text{vec}} = \boldsymbol{F_2}. \tag{3.62b}$$

where the matrices $\underline{\boldsymbol{A}}$ and $\underline{\boldsymbol{B}}$ are the matrices of coefficients. And $\boldsymbol{u}^{\text{vec}}$, $\boldsymbol{v}^{\text{vec}}$ are the variables vector, which contain all the sampling points of u_{ij} and v_{ij}, respectively. $\boldsymbol{F_1}$, $\boldsymbol{F_2}$ are the

vectors of the function values. Solving the matrix equations of the linear system, one can obtain the coordinates of updated mapping $\boldsymbol{M}$.

Minimizing procedure for ϕ Solving a mapping for the L^2 Monge–Kantorovich problem, the mapping aims to fulfil the Jacobian equation; moreover, the integrability of the mapping is requested. However, the previous procedures are tackling the mapping for the Jacobian equation; the resulting mapping is not necessary integrable. One further step for modifying the mapping is performed to ensure the algorithm will converge to an integrable result before entering the next iteration.

In this step, one potential function is assumed, and the minimization is done between the gradient of the potential function and the mapping. Herein, the potential function is defined with B-spline functions:

$$\phi(\boldsymbol{x}) = \sum_k \sum_l C_{kl} N_k(x \mid \boldsymbol{X}) N_l(y \mid \boldsymbol{Y}) \tag{3.63}$$

where $N_k(x \mid \boldsymbol{X})$ and $N_l(y \mid \boldsymbol{Y})$ are the B-spline basis functions defined on the knot vector $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. The arbitrary choice of coefficients C_{kl} span a vector space of the potential function.

The minimization of $\Delta_\phi(\phi, \boldsymbol{M})$ in Eq. (3.44) is done with respect to the gradient of ϕ, where each component of its gradient is written as:

$$\frac{\partial \phi(\boldsymbol{x})}{\partial x} = \sum_k \sum_l C_{kl} \frac{\mathrm{d}N_k(x)}{\mathrm{d}x} N_l(y), \tag{3.64a}$$

$$\frac{\partial \phi(\boldsymbol{x})}{\partial y} = \sum_k \sum_l C_{kl} N_k(x) \frac{\mathrm{d}N_l(y)}{\mathrm{d}y}. \tag{3.64b}$$

Similarly, a least-squares method is applied to minimize $\|\nabla\phi - \boldsymbol{M}\|^2$ so that the coefficients C_{kl} of $\phi(\boldsymbol{x})$ are determined. In this way, $\phi(\boldsymbol{x})$ can be interpreted as the optimum potential function of the vector field $\boldsymbol{M}$.

With the obtained $\phi(\boldsymbol{x})$, a renewed mapping is set as the gradient of $\phi(\boldsymbol{x})$, which is naturally integrable and used as the predefined input for the next iteration.

$$\boldsymbol{M}'(\boldsymbol{x}) = \nabla\phi(\boldsymbol{x}). \tag{3.65}$$

3.4.3 Strategy for the mapping with arbitrary boundaries

The algorithm for solving a mapping from the two density functions presented above gives a practical implementation method. The discretization scheme for the derivative calculation is used in the algorithm, while the calculation is an approximation that is highly accurate only if the variables $\boldsymbol{x}$ are sampled on an equidistant grid. However, in a general application, both the density functions may have an arbitrary boundary contour of their supports. Sampling the variables on an equidistant grid for a bounded support with an arbitrary boundary is not trivial. Therefore, a flexible way is introduced in this work, that instead of considering the direct mapping between two density functions, a dummy density function is introduced. The two mappings are separately solved with the dummy function. Finally, the combination of the two maps gives the result for the problem. Because the homeomorphism is established in both separate map, the combinational map also reserves the homeomorphism.

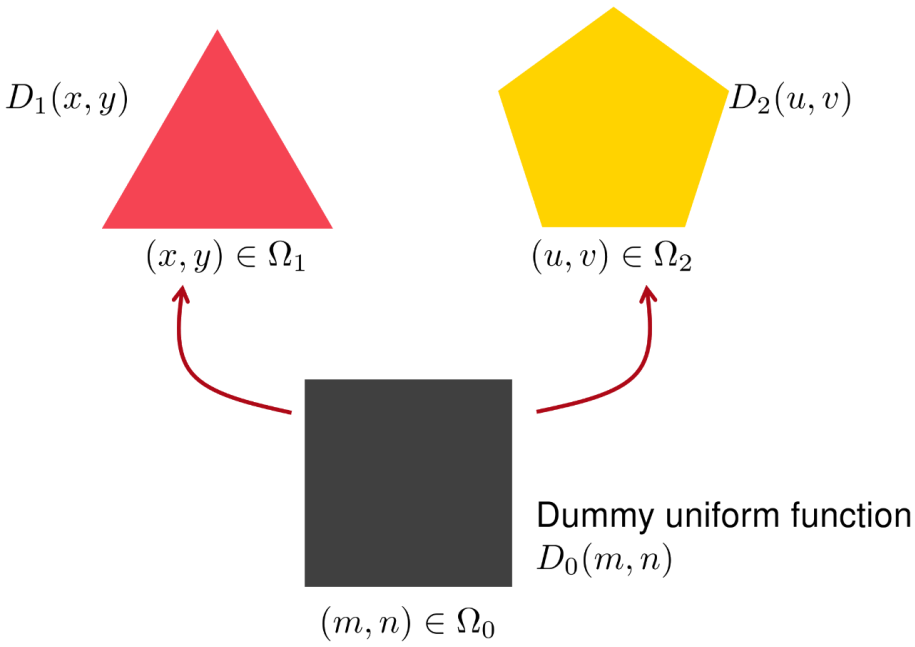

Fig. 3.14 Instead of mapping two density functions $D_1(\boldsymbol{x})$ and $D_2(\boldsymbol{u})$ directly, a dummy uniform function $D_0(\boldsymbol{m})$ is introduced. By solving the mapping between the dummy function and each density function separately, the mapping between $D_1(\boldsymbol{x})$ and $D_2(\boldsymbol{u})$ is then given by the combination of two separate maps.

The basic idea is shown in Fig. 3.14. The problem is to solve a one-to-one map $\boldsymbol{x} \leftrightarrow \boldsymbol{u}$ between the functions $D_1(\boldsymbol{x})$ and $D_2(\boldsymbol{u})$, where $\boldsymbol{x} = (x, y), \boldsymbol{u} = (u, v)$ are the coordinates. The supports Ω_1 and Ω_2 of the functions $D_1(\boldsymbol{x})$ and $D_2(\boldsymbol{u})$ are an arbitrary shape. A dummy density function $D_0(\boldsymbol{m})$ is introduced that is set as a constant function $D_0(\boldsymbol{m}) = C$ with a

squared shape of unit size. Therefore, the sample of the dummy function can be simply set as an equidistant grid. Both the maps $\boldsymbol{m} \leftrightarrow \boldsymbol{x}$ and $\boldsymbol{m} \leftrightarrow \boldsymbol{u}$ are solved using the algorithm from Section 3.4.2. The combination of the maps gives the solution for $D_1(\boldsymbol{x})$ and $D_2(\boldsymbol{u})$, $\boldsymbol{x} \leftrightarrow \boldsymbol{m} \leftrightarrow \boldsymbol{u}$.

The fact that the mappings $\boldsymbol{x}(\boldsymbol{m})$ and $\boldsymbol{u}(\boldsymbol{m})$ start from a uniform distribution $\boldsymbol{m}$ means that the resulting meshes $\boldsymbol{x}$ and $\boldsymbol{u}$ will be equi-density meshes and the density of the mesh nodes of $\boldsymbol{x}$ and $\boldsymbol{u}$ are proportional to their corresponding function values distribution.

However, one should realize that although solving both the maps $\boldsymbol{m} \leftrightarrow \boldsymbol{x}$ and $\boldsymbol{m} \leftrightarrow \boldsymbol{u}$ can result in integrable mapping functions $\boldsymbol{x}(\boldsymbol{m})$ and $\boldsymbol{u}(\boldsymbol{m})$, respectively, the combined mapping function $\boldsymbol{u}(\boldsymbol{x})$ is not necessarily integrable. A mathematical proof is given as follows.

Proof. For a vector field $\boldsymbol{F}(\boldsymbol{x}) = (F_x(\boldsymbol{x}), F_y(\boldsymbol{x}))$ defined on $\mathbb{R}^2$, its integrability is determined by its scalar curl [109, p. 458],

$$\mathrm{ScalarCurl} = \frac{\partial F_x(\boldsymbol{x})}{\partial y} - \frac{\partial F_y(\boldsymbol{x})}{\partial x}. \tag{3.66}$$

that if its scalar curl is equivalent to zero, $\boldsymbol{F}(\boldsymbol{x})$ is an integrable vector field. According to this theorem, because $\boldsymbol{x}(\boldsymbol{m})$ and $\boldsymbol{u}(\boldsymbol{m})$ are both integrable, we have

$$\frac{\partial x}{\partial n}(\boldsymbol{m}) = \frac{\partial y}{\partial m}(\boldsymbol{m}) \text{ or } \frac{\partial n}{\partial x}(\boldsymbol{x}) = \frac{\partial m}{\partial y}(\boldsymbol{x}). \tag{3.67}$$

$$\frac{\partial u}{\partial n}(\boldsymbol{m}) = \frac{\partial v}{\partial m}(\boldsymbol{m}). \tag{3.68}$$

For the function $\boldsymbol{u}(\boldsymbol{x})$, the corresponding derivative for calculating its scalar curl is derived as

$$\frac{\partial u}{\partial y}(\boldsymbol{x}) = \frac{\partial u}{\partial m}(\boldsymbol{m})\frac{\partial m}{\partial y}(\boldsymbol{x}) + \frac{\partial u}{\partial n}(\boldsymbol{m})\frac{\partial n}{\partial y}(\boldsymbol{x}), \tag{3.69}$$

$$\frac{\partial v}{\partial x}(\boldsymbol{x}) = \frac{\partial v}{\partial m}(\boldsymbol{m})\frac{\partial m}{\partial x}(\boldsymbol{x}) + \frac{\partial v}{\partial n}(\boldsymbol{m})\frac{\partial n}{\partial x}(\boldsymbol{x}), \tag{3.70}$$

Therefore, with all the equations above, Eq. (3.67) and Eq. (3.68) together are not a sufficient condition to conclude that the combined mapping is curl-free $\frac{\partial u}{\partial y}(\boldsymbol{x}) = \frac{\partial v}{\partial x}(\boldsymbol{x})$.

Remark. The resulting combination mapping $\boldsymbol{u}(\boldsymbol{x})$ should be re-optimized to obtain an integrable combination mapping with the method described above. The re-optimization is performed as follows.

First, the obtained mesh $\boldsymbol{x}$ is fixed. The mesh of $\boldsymbol{u}$ is to be modified so that $\boldsymbol{u}(\boldsymbol{x})$ is integrable, and $\boldsymbol{u}(\boldsymbol{m})$ satisfies the local mass conservation.

Second, to optimize the integrability of $\boldsymbol{u}(\boldsymbol{x})$, the same procedure as the minimizing procedure for ϕ in the last section is performed. Note that here, $\boldsymbol{x}$ is gridless data. Nevertheless, the B-spline technique can still be flexibly applied because the technique can also tackle the problem with gridless data and obtain an approximate potential function $\phi(\boldsymbol{x})$ for $\boldsymbol{u}(\boldsymbol{x})$. Moreover, because $\phi(\boldsymbol{x})$ is represented by B-spline functions, its gradient can be calculated analytically once it is obtained. Therefore, the gradient of $\phi(\boldsymbol{x})$ on the position of $\boldsymbol{x}$ is set as the predefined input for the next step, $\boldsymbol{u}'(\boldsymbol{x}) = \nabla\phi(\boldsymbol{x})$.

Third, the modified mesh $\boldsymbol{u}'$ from the last step may lose the local mass conservation with the mesh $\boldsymbol{m}$. Therefore, $\boldsymbol{u}'(\boldsymbol{m})$ is used as the initial mapping to solve the mapping between $D_0(\boldsymbol{m})$ and $D_2(\boldsymbol{u})$, so that the resulting mesh $\boldsymbol{u}$ meets the mass condition.

The second and third steps are iteratively performed until the $\boldsymbol{u}(\boldsymbol{x})$ satisfies both the integrability condition and the local mass conservation condition.

3.4.4 Integrability of the numerical result

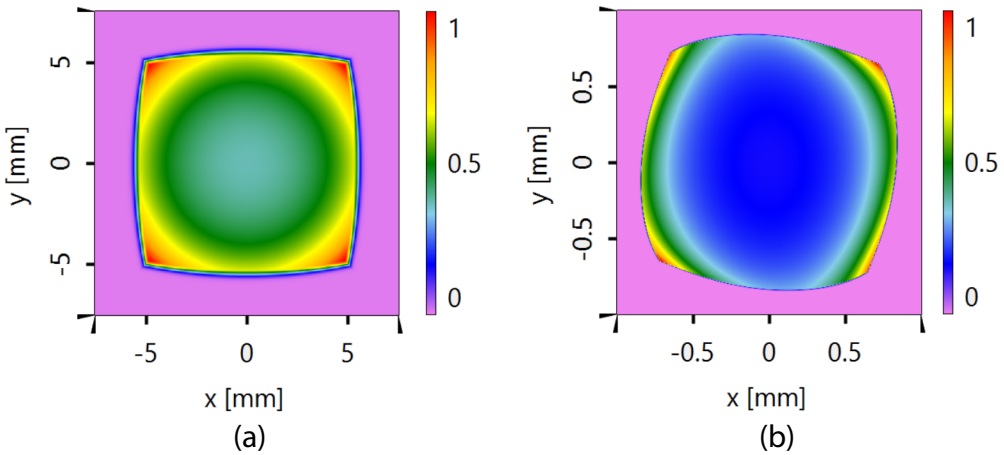

Fig. 3.15 Two density functions with special boundary contour. (a) $D_1(\boldsymbol{x})$: barrel shape, and (b) $D_2(\boldsymbol{u})$ distorted barrel shape.

A numerical experiment is shown to demonstrate the algorithm. Here, a barrel-shaped density function is mapped to another function with a distorted barrel shape. Both the functions are shown in Fig. 3.15. As described in the algorithm, both functions are mapped to a square uniform function to tackle the mapping problem for functions with arbitrary support boundary. Let $[-1, 1] \times [-1, 1]$ be the bounding size of the square dummy function. This

configuration of the dummy function can be fixed for any mapping problem, regardless of the physical domains in which the density functions are defined.

The algorithm solves the mappings between both density functions and the dummy function. Their homeomorphism is illustrated by the meshes in Fig. 3.16. The dummy function is sampled with 41×41 equidistant grid points, where the corresponding mapping meshes are shown in Fig. 3.16(a) and (b), respectively.

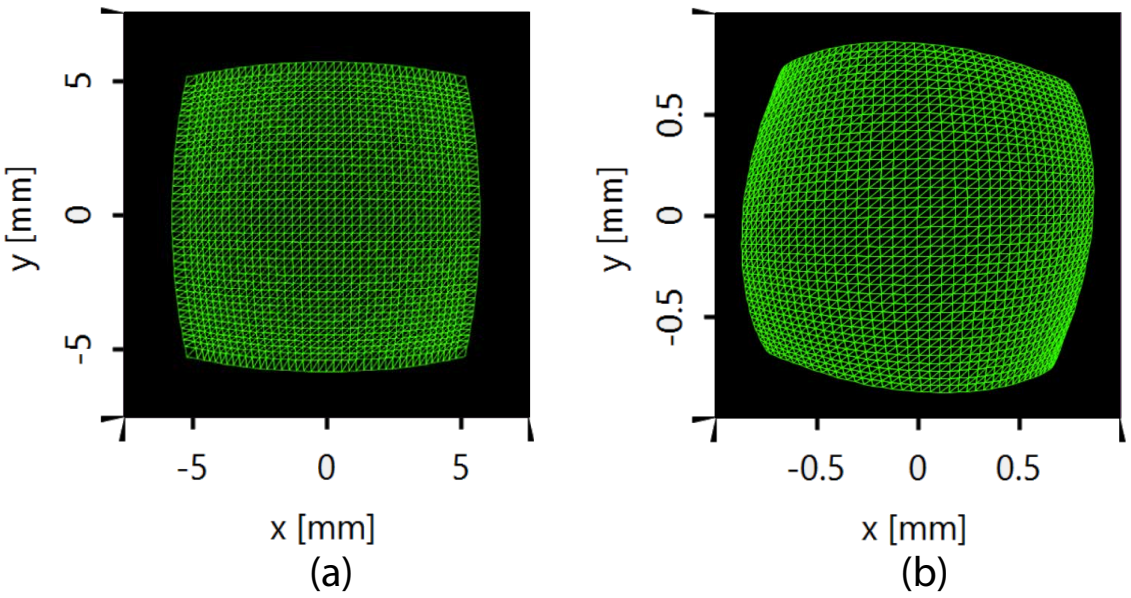

Fig. 3.16 Equi-mass meshes generated from a square equidistant starting grid. (a) mesh for $D_1(\boldsymbol{x})$: barrel shape, and (b) mesh for $D_2(\boldsymbol{u})$: distorted barrel shape.

Eq. (3.43) is calculated in every iteration in the algorithm and considered as the merit function for the problem. The convergence of the merit function is shown in Fig. 3.17(a). The integrability of the mapping function can be determined by its scalar curl, as denoted with Eq. (3.66). The relative root-mean-square (RMS) of the value of the scalar curl is calculated to illustrate the integrability.

$$f_{RRMS} = \sqrt{\frac{\sum_i \sum_j \left[\frac{\partial F_x(\boldsymbol{x}_{ij})}{\partial y} - \frac{\partial F_y(\boldsymbol{x}_{ij})}{\partial x}\right]^2}{\sum_i \sum_j \left[\frac{\partial F_x(\boldsymbol{x}_{ij})}{\partial y}\right]^2}}. \tag{3.71}$$

Fig. 3.17(b) shows the convergence of their integrability.

However, as discussed in Section 3.4.3, the combined mapping does not meet the integrability condition. For the combined mapping function, $f_{RRMS} = 28.5\%$, even though

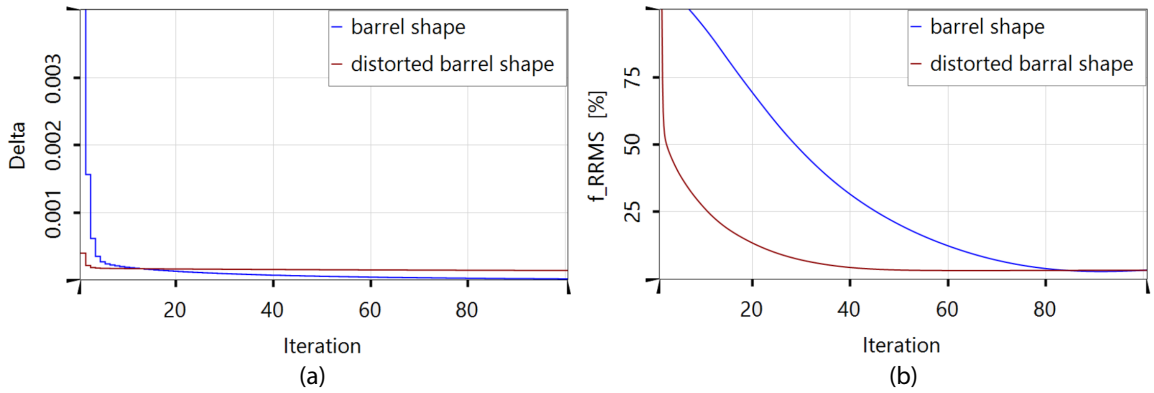

Fig. 3.17 Convergence analysis in each mapping solution for the two density functions. (a) Convergence of the mass conservation by the developing mapping. The deviation value in Eq. (3.43) is calculated with respect to iterations. (b) Convergence of the integrability of each separate mapping. The RRMS of the scalar curl is calculated for each iteration.

both separate mappings to about 3.3% and 3.2%, respectively. Further optimization of the combined mapping is performed. A further optimization of the combined mapping is performed. Fig. 3.18 shows the convergence of the integrability of the combined mapping in the optimization process.

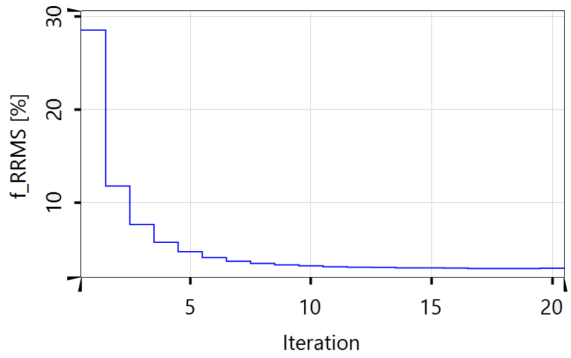

Fig. 3.18 Convergence of the integrability for the combined mapping.

In conclusion, the separate map-solving method provides a flexible way to calculate the mapping between two functions with different arbitrary support shapes. In optics, the method is applied to solve a mapping between two intensity functions. Two separate equi-flux meshes are obtained for the intensity functions using the algorithm illustrated in Section 3.4.2. The mapping between the functions is then obtained by the combination of the meshes, where the local energy conservation still holds. However, the integrability of the combined mapping is lost. A further optimization step proceeds the improvement of the integrability.

Table 3.1 The strategy of Fourier pair synthesis methods.

	Strategy
homeomorphic case	mapping-type method
non-homeomorphic case	mapping-type method + IFTA

3.5 Summary

In this chapter, the functional design for light shaping is discussed. The functionality of the required optical element is assumed as a WPR function. The retrieval of the output wavefront phase is the main step for attaining the WPR function. A Fourier pair synthesis is the key process for the output phase retrieval. A mapping-type method with the homeomorphic assumption between the fields in the spatial domain and the spatial-frequency domain is proposed for the Fourier pair synthesis, which solves the problem in homeomorphic cases also tackles the stagnation issue of IFTA in such cases. With the proposed mapping-type method, the strategy of the output wavefront retrieval is illustrated in Table 3.1. In the homeomorphic cases, the mapping-type method can provide an accurate result. In the cases where homeomorphism is not established, starting with the mapping-type method and continuing with the IFTA can obtain the proper result, also avoiding the speckle effect from IFTA. Because the homeomorphism cannot be known until the required output phase is retrieved, the mapping-type method should be used first.

The mapping-type Fourier pair synthesis method is also compared with another mapping-type algorithm in the spatial domain. The other method assumes a mapping between the irradiance distribution at the element plane and the distribution at the target plane, which is actually applying the constraint from the target on $E_\ell^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$ instead of $\tilde{E}_\ell^{\mathrm{out}}(\boldsymbol{\kappa})$. The gradient of the output wavefront phase is then derived from this mapping. However, the resulting gradient generally suffers from the integrability problem. At the same time, the gradient of the output wavefront phase derived from the mapping solution of the Fourier pair is integrable in any cases. The L^2 Monge–Kantorovich maths model is applied to solve the mapping from the Fourier pair. A least-squares-based algorithm is introduced to solve the problem. The integrability of the resulting mapping is well controlled in the algorithm.

The functional design by the physical-optics approaches provides a comprehensive view for the far-field light-shaping problem. The design methods in the literature for the diffractive optics and the refractive optics are unified. We learn from the diffractive optics method that a Fourier pair synthesis mainly required and from the refractive optics method that a homeomorphic assumption is usually applied. By introducing the homeomorphic assumption,

a mapping-type method enriches the class of Fourier pair synthesis methods. The physical-optics analysis and the unification of the approaches from different communities give a deeper insight and more strategies for the design. For example, where to apply the constraint from the target, or when to use the homeomorphic assumption. The resulting WPS function also helps to understand the system, for example, whether the homeomorphic situation is true in the system, before starting more complicated structural design.

Chapter 4

Design structural embodiment of the light-shaping element

When designing the functional embodiment of the light-shaping element, its functionality is analyzed, and a designed algorithm is proposed. For the far-field light-shaping problem, we know from the previous chapter that the functionality of the required element is mainly a WPR function that modifies the wavefront phase from the input into a required output. In the design of the structure embodiment, the design algorithm takes advantage of the conclusion from the WPR function, or more directly, from the required output wavefront phase.

Different element structures are considered to realize the element function, such as diffractive optics or refractive optics. CGH, as a type of diffractive optics, can be concluded from the WPR function because CGH can be modeled on a plane, and the local grating structure of the CGH is derived from the gradient of the WPR function. A freeform surface, due to its smoothness, which is considered an advantageous aspect, is also widely used for far-field light shaping. The required output wavefront phase obtained from the functional embodiment step is used as an objective for designing the freeform surface.

The physical model of the optical element is important in the design, either for developing the algorithm or for the analysis of the design results. In this chapter, the physical models of the HOE and freeform surface are presented. The design algorithms are developed to search for a proper configuration of each model that adapts the task. After the design, the structure of both types of elements is analyzed with physical-optics simulation. Their physical effects are presented. The extra effects that introduced errors to the design are discussed, and solutions are proposed for compensating the effects from their structures.

4.1 Design holography optical element (HOE) for light shaping

There are two different ways to fabricate the HOE. One is optically recorded, where two light beams interfere with each other, and the interference pattern is recorded as the hologram. The hologram registers both the amplitude and phase information. The hologram can also be calculated by knowing both the fields of the interference light. In this work, the latter method is used for the design. However, to model a HOE from a physical-optics point of view, the optically recording process is a key to understand its physical model.

4.1.1 Physical model of the HOE

The recording process of HOE is reviewed in the following. Fig. 4.1 shows the sketch of how the HOE is recorded. Let the electric field $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ be defined on the input and output plane of the HOE, respectively. Both the fields propagate to the HOE. Their interference pattern generate an intensity modulation, which can be recorded inside a material as an index modulation or absorption modulation pattern to form a HOE.

Here, the intensity is defined as

$$I(\boldsymbol{\rho}) = \boldsymbol{E}(\boldsymbol{\rho})\boldsymbol{E}^{*}(\boldsymbol{\rho}). \tag{4.1}$$

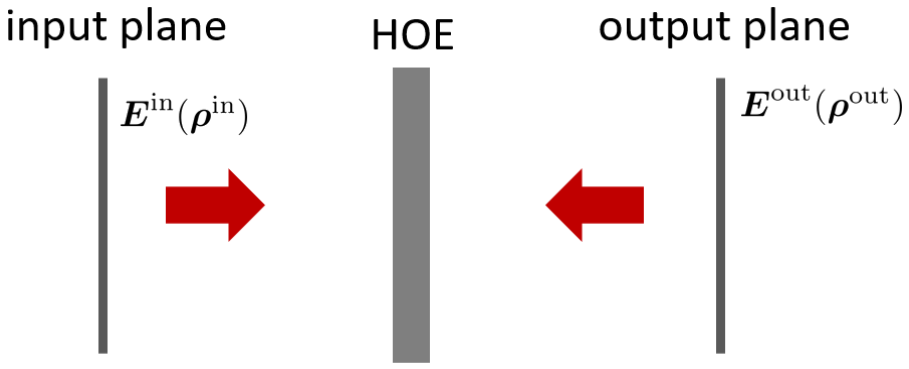

Fig. 4.1 Sketch of the design of HOE.

To simplify the discussion, both fields are assumed as a plane wave with the same polarization state first. Therefore, the fields are written explicitly as

$$\begin{aligned} \boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}}) &= \hat{\boldsymbol{E}}^{\mathrm{in}} \exp(\mathrm{i}\boldsymbol{\kappa}^{\mathrm{in}}\boldsymbol{\rho}^{\mathrm{in}}), \\ \boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}}) &= \hat{\boldsymbol{E}}^{\mathrm{out}} \exp(\mathrm{i}\boldsymbol{\kappa}^{\mathrm{out}}\boldsymbol{\rho}^{\mathrm{out}}). \end{aligned} \tag{4.2}$$

where $\hat{\boldsymbol{E}}^{\mathrm{in}}$ and $\hat{\boldsymbol{E}}^{\mathrm{out}}$ are the unit vector of the field. $\boldsymbol{\kappa} = (k_x, k_y)$ indicates the transversal components of the wave vector.

When both fields interfere at the HOE position, the interfere intensity is

$$\begin{aligned} I(\boldsymbol{\rho}) &= \left(\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}) + \boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho})\right)\left(\boldsymbol{E}^{\mathrm{in}^*}(\boldsymbol{\rho}) + \boldsymbol{E}^{\mathrm{out}^*}(\boldsymbol{\rho})\right) \\ &= 2 + 2\cos((\boldsymbol{\kappa}^{\mathrm{out}} - \boldsymbol{\kappa}^{\mathrm{in}})\boldsymbol{\rho}). \end{aligned} \tag{4.3}$$

Eq. (4.3) describes a grating pattern with its grating vector defined as

$$\boldsymbol{K} = \boldsymbol{\kappa}^{\mathrm{out}} - \boldsymbol{\kappa}^{\mathrm{in}}. \tag{4.4}$$

Therefore, the intensity modulation generates a HOE as a linear grating. The HOE modifies the wave vector of the input field with the grating equation Eq. (4.4).

So far, an HOE design with two plane waves or two planar wavefronts is derived, and the HOE is concluded as a grating. In general, if both fields $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ are in their homeomorphic field zone, that is, by Fourier transform, the field values in the $\boldsymbol{\rho}-$domain contribute to a single field value in its $\boldsymbol{\kappa}-$domain, and they can be decomposed as local plane waves. Both sets of local plane waves interfere at the HOE plane. Locally, the interference pattern generates a local linear grating. In the far-field light-shaping task, the smoothness of each wavefront usually can be fulfilled. Therefore, the local linear grating model can be applied for HOEs designed for light shaping.

With the local linear grating approximation, the functionality of the HOE mainly is to modulate the input wavefront and provide another smooth output wavefront. The local linear grating approximation (LLGA) is shown in Fig. 4.2, where the local period of the gratings can be calculated from its grating vector or the phase function of the HOE.

$$\Lambda(\boldsymbol{\rho}) = \frac{2\pi}{|\boldsymbol{K}(\boldsymbol{\rho})|} = \frac{2\pi}{|\nabla(\Delta\psi(\boldsymbol{\rho}))|}, \tag{4.5}$$

where $\Delta\psi(\boldsymbol{\rho})$ is the phase function of the HOE.

At an arbitrary position at the HOE, the interaction with the input field is assumed as a linear grating interacts with a plane wave. The interaction can be rigorously solved by the Fourier modal method (FMM) [112–114]. Once the HOE material is given and the period of the local grating $\Lambda(\boldsymbol{\rho})$ is derived, FMM calculates the field-related diffraction property of the HOE.

The FMM decomposes the grating into layers, with each layer considered as a periodic medium. The implementation of the method contains two main procedures: 1. An eigenmode solver is developed for each periodically modulated layer. The permittivity distribution of the

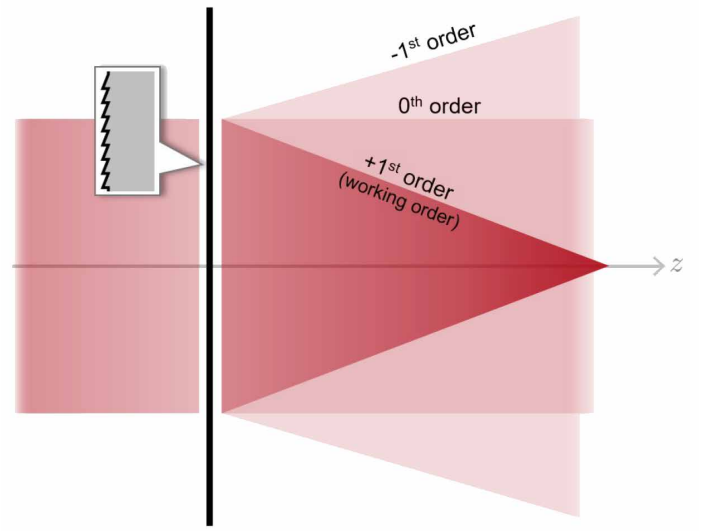

Fig. 4.2 Sketch of local linear grating approximation (LLGA). Figure from the website:(http://www.lighttrans.com/technology-whitepapers/local-linear-grating-approximation-llga.html)

layer is expanded into a Fourier series regarding its spatial frequency. The Maxwell equations are also derived in their Fourier transformation form. Therefore, the FMM solver is working in the spatial-frequency domain ($\boldsymbol{\kappa}-$domain). The Maxwell equations are then reorganized into a matrix equation with the structure-related parameters set as the eigenmodes of the layer. 2. An S-matrix for each layer is computed to connect the field in front of and behind the layer by recursively matching the boundary conditions. With these two procedures, for a given input field, FMM provides the diffractive field, reflective field and even the field inside the grating component.

Note that the diffractive field contains different orders for each local grating, which may introduce noise for the light shaping. The compensation and optimization of the redundant orders will be discussed in Section.4.1.3 with examples.

4.1.2 Design of the HOE

The smoothness of the required output wavefront phase is asked for to obtain a smooth phase function for the HOE. Retrieval of the output wavefront phase from the target field is usually difficult because the amplitude or the phase of the target field is not captured during the measurement.

Different algorithms are proposed to retrieve the output phase with the given information from the task. IFTA is one of the commonly used approaches for calculating the required

output phase. However, as discussed in Section.3.1, IFTA may end up with a stagnation problem if the output field exhibits a homeomorphic situation. Besides, in general, without additional constraints, the result of IFTA usually contains phase dislocation. The smoothness of the output phase cannot be preserved, and speckles may appear in the generated pattern.

The mapping-type Fourier synthesis method proposed in Section 3.2 provides a tool to achieve a smooth output wavefront phase. The algorithm is performed in a single step without iterative processes. It is derived under the homeomorphic assumption. However, for those situations where homeomorphism does not exist, the obtained phase can be used as an initial guess for the IFTA, while the speckle problem in IFTA can also be avoided with this initial guess.

After the required output wavefront phase is obtained, the element function for the HOE is assumed as a WPR function, which is calculated by simply subtracting the input wavefront phase from the out phase. The LLGA model of the HOE is established with its WPR function, using Eq. (4.5). The LLGA model is then applied to analyze the designed HOE and further optimization. In addition, the WPR function is wrapped with the 2π module and used for the fabrication of the HOE component or directly stored for the spatial light modulator (SLM).

4.1.3 Demonstration with examples

The next example demonstrates the design of HOE for light shaping. Here, a Gaussian wave with its waist radius as $500\mu\mathrm{m}$ and wavelength of $532\mathrm{nm}$ is given as the source field. The target irradiance is a top-hat profile defined by a separable super Gaussian field.

$$\|\boldsymbol{E}(x,y)\| = \exp\left[-\left(\frac{x}{\omega_{0,x}}\right)^{m_x} - \left(\frac{y}{\omega_{0,y}}\right)^{m_y}\right], \tag{4.6}$$

where $\omega_{0,x}$ and $\omega_{0,y}$ are the waist radii for both $x-$ and $y-$dimensions, and m_x, m_y is each order. In this example, the waist radii of the super Gaussian fields are chosen as $500\mathrm{mm}$ for both the $x-$ and $y-$dimension, and both the orders are 10. The target plane is set $1\mathrm{m}$ away from the HOE and perpendicular to the optical axis. The input and target irradiance distributions are shown in Fig.4.3.

The HOE is predefined as a thin slab with double planar surfaces. The medium in between has a refractive index of 1.8. The required output wavefront phase is retrieved by the mapping-type Fourier synthesis method. When it is retrieved, the WPR function of the HOE is calculated by subtraction from the wavefront phase of the input field. The WPR is presented in Fig. 4.4(a). It mainly provides a spherical phase function that governs the free space propagation behind the HOE to achieve the required divergence and phase modulation

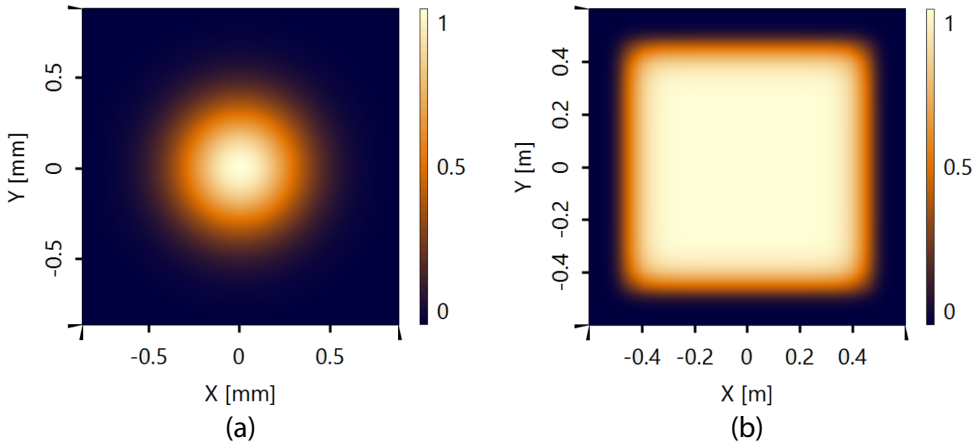

Fig. 4.3 Irradiance distribution (normalized) of the input and target. (a) Irradiance of the Gaussian profile on the input plane of the HOE that was obtained with the given source field. (b) Top-hat shape irradiance distribution on the target plane.

to redistribute the interior irradiance. The residual phase after abstracting a best fit spherical phase is shown in Fig. 4.4(b).

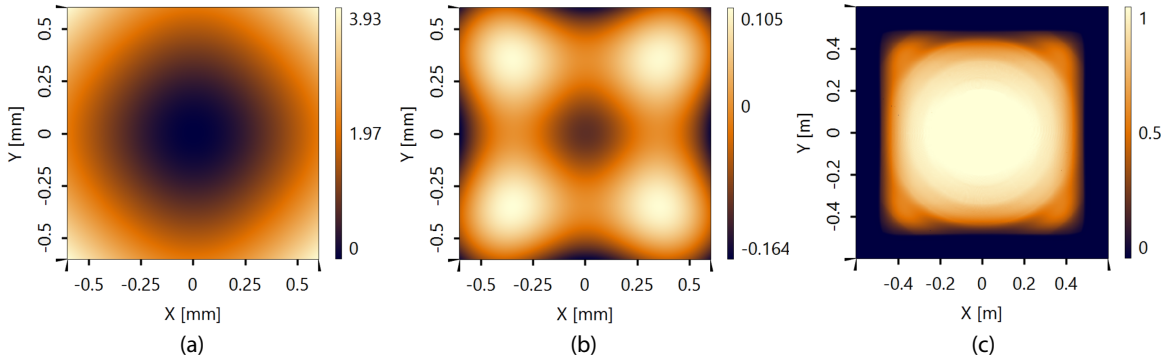

Fig. 4.4 (a) The WPR function of the HOE. (b) The residual phase after subtracting a best fit spherical phase. (c) Irradiance distribution (normalized): simulation result with the HOE component.

The resulting HOE is modeled and analyzed in the software VirtualLab Fusion [95]. After the HOE is configured in the software, the simulation is done following the field tracing techniques illustrated in Fig. 2.2. In this case, the $\mathcal{B}$ is specified as the LLGA operator, and the local gratings are set as the sawtooth type. The detector set at the target plane gives the irradiance distribution, as shown in Fig. 4.4 (c). Errors are indicated compared to the target pattern shown in Fig. 4.3 (b). The errors in the irradiance distribution are due to the HOE structure. In addition to the wavefront phase response, the HOE structure also modulates the amplitude of the input field. Two main aspects are introduced by the HOE structure:

(1) additional diffractive orders and (2) irregular amplitude modulation from the Rayleigh coefficients with respect to grating periods.

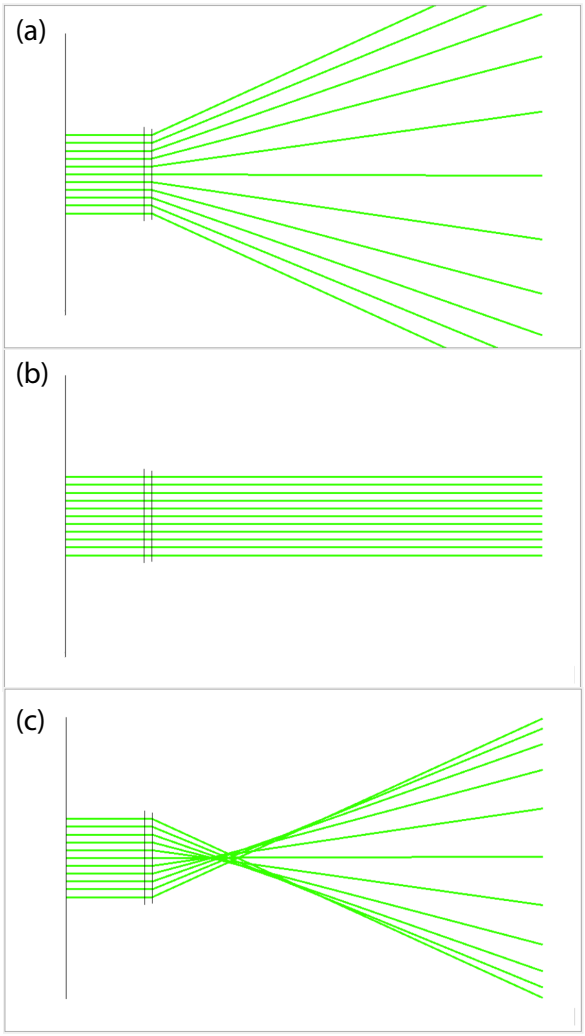

Fig. 4.5 Ray-tracing demonstration of the propagation of different orders: (a) $+1$ order, (b) 0 order, (c) -1 order.

With the LLGA model, different diffractive orders of the HOE are rigorously calculated. The ray-tracing diagram in Fig. 4.5 illustrates the propagation of the $+1$, 0 and -1 order. The $+1$ order of the HOE is the working order that contributes to the required shaping. However, due to the diffractive effect of the HOE structure, the additional orders are introduced as stray light and influence the detector irradiance.

The local gratings of the HOE also introduce modulations of the amplitude of the input field, which are indicated by the Rayleigh coefficients of the gratings. The period of the local gratings is a function with respect to its local coordinate in the HOE. Therefore, the Rayleigh coefficients of different grating periods introduce amplitude modulations through the whole

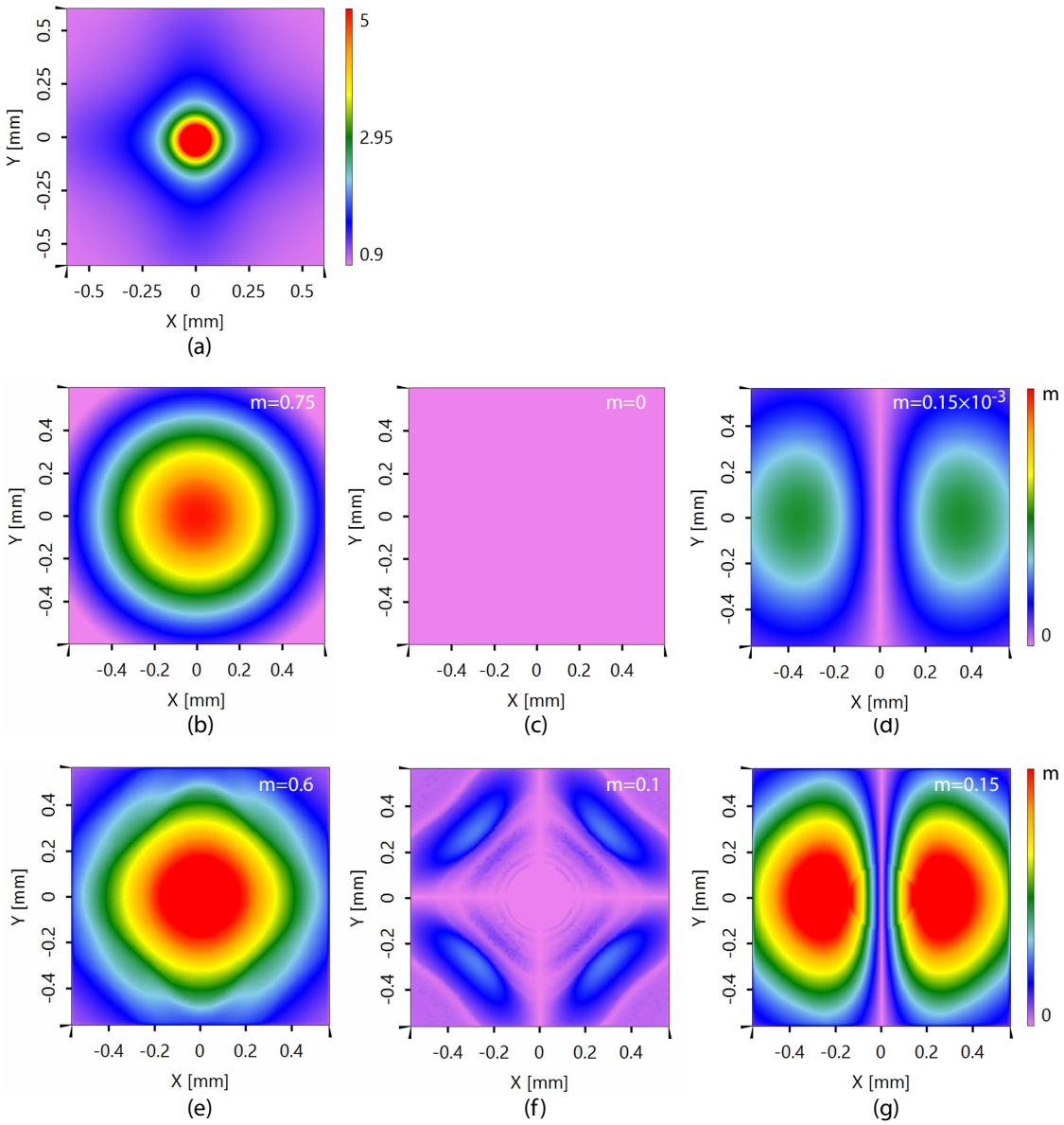

Fig. 4.6 (a) The period distribution of the local gratings of the HOE (unit: μm). (b) $x-$, (c) $y-$ and (d) $z-$ component of the field in front of the HOE. (e) $x-$, (f) $y-$ and (g) $z-$ component of the field behind the HOE.

HOE. Fig. 4.6 (a) shows the period distribution of the local gratings. To achieve a required output wavefront phase value at a certain position, or more precisely, a required local output wave vector, the grating period varies roughly from $900\mathrm{nm}$ to $30\mu\mathrm{m}$. The maximum value in Fig. 4.6 (a) is clipped to $5\mu\mathrm{m}$ for better visualization.

The field in front of the HOE is shown in the second row of Fig. 4.6 with the field components. The $+1$ order of the field behind the HOE is also shown in the third row to

indicate the amplitude modulation from the HOE. The comparison of the fields in front and behind shows the different modulation regarding the local grating positions. At the area near the edge of the HOE, where the grating period is small, the modulation shows some irregular effect.

The effects of (1) additional orders and (2) amplitude modulations with respect to different grating periods must be overcome to improve the HOE performance for more accurate light shaping.

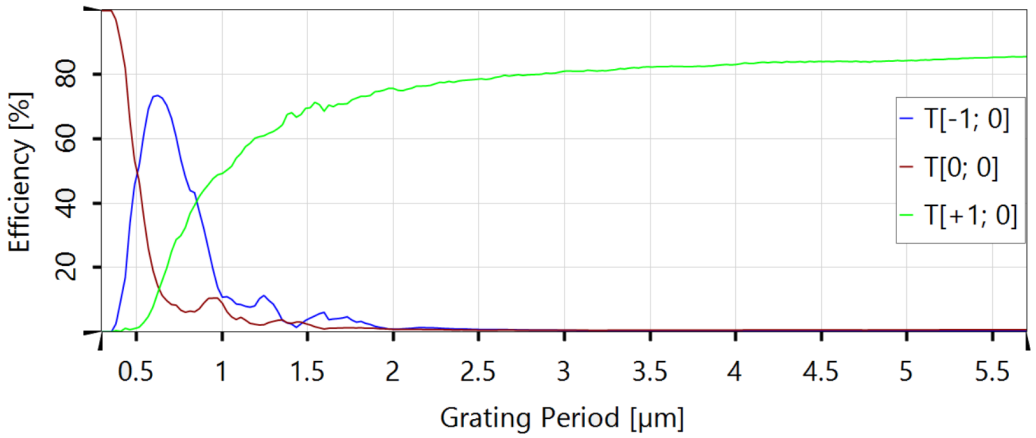

Fig. 4.7 The diffraction efficiency of different orders with respect to grating periods.

Fig. 4.7 indicates the diffraction efficiency of a sawtooth grating, with a normal incident plane wave of 532nm for the investigation. The curves in the figure illustrate the diffraction efficiency of the three orders, respectively, developed with respect to the grating periods. Fig. 4.7 shows that with increasing period value, the efficiency of the $+1$ order also increases. In the small period area, where the numbers are comparable to the wavelength, the 0 order and -1 order contribute the main diffraction efficiency. However, in the area where the period values are above $2\mu\text{m}$, the diffraction efficiency of the 0 order and -1 order decrease to a value close to 0%, thus the $+1$ order dominates the diffraction lights. Moreover, the efficiency of the $+1$ order almost converges to a constant number when the period is up to $5\mu\text{m}$. Therefore, we can conclude that if the period distribution of the local gratings in the HOE is above the number of $5\mu\text{m}$, the diffraction light will be dominated by the working order ($+1$ order), the stray light from the additional orders is suppressed, and the amplitude modulation of the local gratings is close to a constant value t hrough the whole HOE.

According to Eq. (4.5), to relax the local grating period, one should reduce the gradient of the WPR function of the HOE, which means reducing $\boldsymbol{K}(\boldsymbol{\rho})$, the difference between the input and output wave vectors of the local plane waves. To this end, one idea for the fixed output wavefront phase that is the requirement for the design is to introduce a wavefront

phase for the input field that is similar to the required output phase, so that locally their difference $\Delta\psi(\boldsymbol{\rho})$ is reduced and, therefore, $\boldsymbol{K}(\boldsymbol{\rho})$ is also reduced.

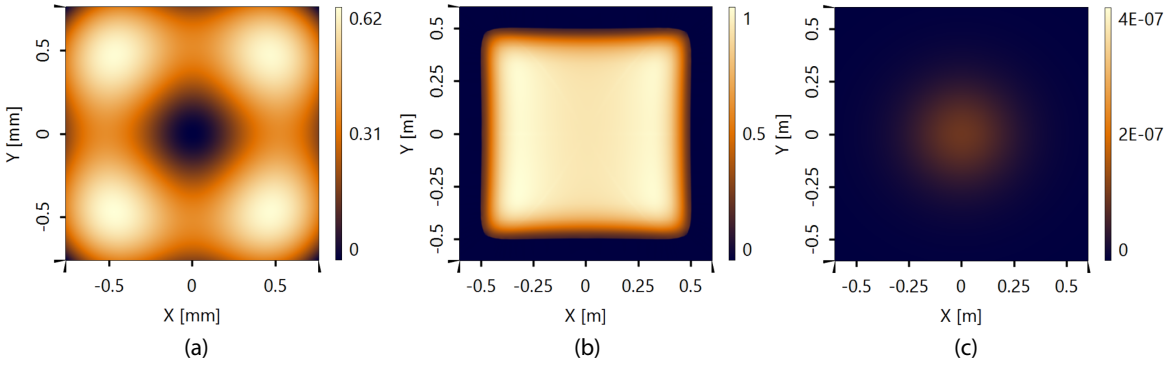

Fig. 4.8 (a) Recalculated WPR function of the HOE with a spherical surface. (b) Irradiance distribution (normalized): simulation result with the hybrid component. (c) Irradiance distribution from the 0 order (normalized with the maximum number of (b)).

A simple approach is increasing the HOE thickness and setting the front surface of the HOE as a spherical surface. In this way, a quadratic wavefront phase is introduced inside the HOE. In this example, the radius of curvature of the spherical surface is chosen as $1\mathrm{mm}$ so that a wavefront phase approximate to the required output phase is generated. Without changing its medium, the HOE thickness is set as $100\mu\mathrm{m}$. Therefore, a hybrid component with refractive and diffractive optics is established. The WPR function of the HOE is recalculated with the generated wavefront phase in the HOE, as shown in Fig. 4.8 (a).

With the hybrid component, the simulation result shows the irradiance distribution at the target plane, as shown in Fig. 4.8 (b). Compared to the result shown in Fig. 4.4 (c), the detected irradiance this time is much more homogeneous and coincides with the target pattern. With the spherical surface, the local grating period of the designed HOE is released to a large number and therefore reduces the stray light from the additional orders. The irradiance from the 0 order is shown as an example to demonstrate the stray light, as shown in Fig. 4.8 (c). The small values indicate the stray light is a tiny portion that can be neglected.

In summary, the LLGA operator is applied for the modeling of HOEs. The HOE is designed by the WPR function obtained from the functional design step. However, the structure of the HOE may introduce an additional amplitude modulation effect to the input field. For reducing the extra effect, the period of the local gratings in the HOE should be approximately ten times higher than the wavelength. A hybrid solution is suggested for enlarging the period by introducing an additional surface for the HOE.

4.2 Design freeform surface for light shaping

Similar to the HOE, the modeling of the freeform surface is also important for developing the design algorithm because the modelling techniques reveal the relationship between the parameter of the freeform surface and its response to the electric field.

Modeling the electric field propagation through the surface can be solved rigorously by Maxwell's equations with boundary conditions. However, if the size of the surface is beyond millimeter magnitude, or even hundreds of micrometer magnitude, the numerical effort of those rigorous solvers is expensive even with current computer processing. The local plane interface approximation (LPIA) is a more practical method to model the field propagation through a large surface.

In general, the LPIA operator is an integral operator that acts in the $\boldsymbol{\kappa}-$ domain. However, for the cases where the input field can be decomposed into local plane waves, which is fulfilled for most light-shaping systems with the freeform surface, the LPIA operator can be applied in the $\boldsymbol{\rho}-$domain simply with a pointwise calculation [90]. In the following subsections, the detail of the modeling method is presented first, and then the design algorithm of the freeform surface is proposed. Finally, the algorithm is demonstrated with examples.

4.2.1 Physical model of the freeform surface

To illustrate the LPIA method, a sketch about field propagation through a surface is shown in Fig. 4.9. The input and out field of the surface $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ are defined on two parallel planes. In general, the planes can be non-parallel [115]. For modeling, $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ are connected by two free-space propagation operators the LPIA operator of the surface [90]. The sequence of the operators is written explicitly as,

$$\boldsymbol{E}^{\mathrm{out}} = \boldsymbol{\mathcal{P}}^{\mathrm{out}}\boldsymbol{\mathcal{B}}^{\mathrm{LPIA}}\boldsymbol{\mathcal{P}}^{\mathrm{in}}\boldsymbol{E}^{\mathrm{in}}, \tag{4.7}$$

where $\boldsymbol{\mathcal{P}}^{\mathrm{in}}$ is the operator for propagating the field from the input plane to the surface, and $\boldsymbol{\mathcal{P}}^{\mathrm{out}}$ propagates the field from the surface to the output plane.

In the case that the input field can be approximately decomposed into local plane waves, the $\boldsymbol{\mathcal{P}}^{\mathrm{in}}$ and $\boldsymbol{\mathcal{P}}^{\mathrm{out}}$ represent the local plane wave propagation [116, 117]. The LPIA method consider a local boundary condition at the surface for each local plane wave, which is defined by $\boldsymbol{\mathcal{B}}^{\mathrm{LPIA}}$. Therefore, the $\boldsymbol{\mathcal{B}}^{\mathrm{LPIA}}$ operator is a pointwise calculation.

$\boldsymbol{\mathcal{B}}^{\mathrm{LPIA}}$ contains three sub-operators, which are two coordinate transformations and an operator for the boundary condition. If the global coordinate system is defined with the unit vector of its axes as $[\hat{x}, \hat{y}, \hat{z}]$, the first coordinate transformation operator $\boldsymbol{\mathcal{Q}}$ transforms the

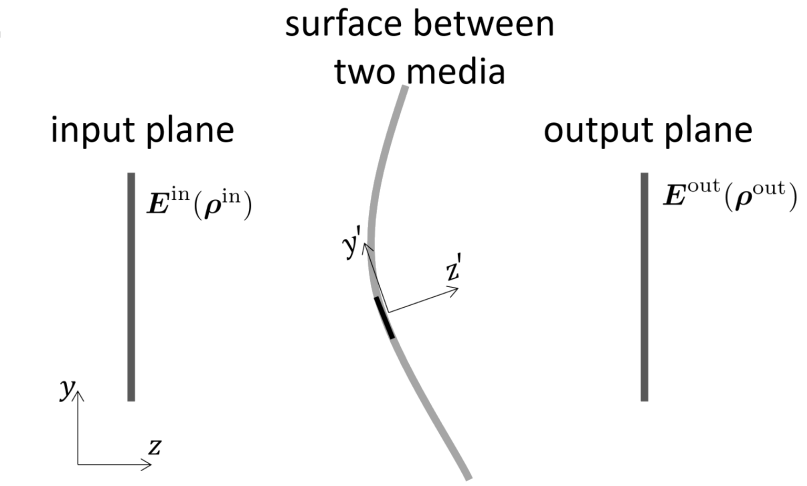

Fig. 4.9 Illustration of the field propagation through surface.

global coordinate system to the local coordinate systems $[\hat{x}', \hat{y}', \hat{z}']$ on the surface,

$$[\hat{x}', \hat{y}', \hat{z}']^{\mathrm{T}} = \boldsymbol{\mathcal{Q}}[\hat{x}, \hat{y}, \hat{z}]^{\mathrm{T}}. \tag{4.8}$$

Therefore, with $\boldsymbol{\mathcal{Q}}$, all the local plane waves from the input field can be transformed into their local coordinate system at the surface

$$\boldsymbol{E}'^{\mathrm{in_c}} = \boldsymbol{\mathcal{Q}}\boldsymbol{E}^{\mathrm{in_c}}, \tag{4.9}$$

where $\boldsymbol{E}^{\mathrm{in_c}}$ and $\boldsymbol{E}'^{\mathrm{in_c}}$ are the input local plane wave fields at the surface, in the global and local coordinate system respectively.

After the field passes through the surface, the output local plane wave fields are then transformed back to the global coordinate system by

$$\boldsymbol{E}^{\mathrm{out_c}} = \boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{E}'^{\mathrm{out_c}}. \tag{4.10}$$

Similarly, $\boldsymbol{E}^{\mathrm{out_c}}$ and $\boldsymbol{E}'^{\mathrm{out_c}}$ are the output fields at the surface in the global and local coordinate system, respectively.

At the local positions of the surface, the boundary condition is applied to connect the input and output fields:

$$\boldsymbol{E}'^{\mathrm{out_c}} = \boldsymbol{\mathcal{C}}^{\mathrm{Fres}}\boldsymbol{E}'^{\mathrm{in_c}}, \tag{4.11}$$

where $\boldsymbol{\mathcal{C}}^{\mathrm{Fres}}$ the Fresnel matrix of each local planes at the surface.

From Eqs. (4.9), (4.10) and (4.11), the $\boldsymbol{\mathcal{B}}^{\mathrm{LPIA}}$ is written explicitly as

$$\boldsymbol{\mathcal{B}}^{\mathrm{LPIA}} = \boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{\mathcal{C}}^{\mathrm{Fres}}\boldsymbol{\mathcal{Q}}. \tag{4.12}$$

The time complexity of the LPIA is $\mathcal{O}(n)$, where n is the sampling points of the field, mainly the wavefront. Therefore, the computational effort of the LPIA is much more efficient than the rigorous Maxwell's solver.

So far, the LPIA method is described for modeling field propagation through the surface. In the following, the principle of the modeling technique is used for the design of a freeform surface.

4.2.2 Design of the freeform surface

Remember that in the functional embodiment design, the task has been analyzed with the field tracing techniques. The functionality of the shaper is mainly a WPR function, such that the required output wavefront phase is achieved by the shaper. The required output wavefront phase from the functional design is used in the design of the freeform surface. In the functional design, the element is assumed to be a phase-only response. However, once the real structure of the freeform lens is introduced, the amplitude of the input field will be modulated. Although the required output wavefront phase can be reconstructed by the freeform lens with the amplitude modulation, the output field is changed from the one in the functional design process so that the amplitude constraint for the design is broken. Therefore, the target field or the target irradiance distribution cannot be achieved.

Feng et al. [107, 118] have proposed an iterative wavefront phase tailoring method to address the above problem. With the amplitude modulation from the freeform surface, the required output wavefront phase is recalculated again. Then, the retrieval of the required output wavefront phase and the construction of the freeform lens alternately proceeds. In this way, the design of the freeform lens converges to a result that a proper output field is reached for light shaping. In their method, a complex nonlinear partial differential equation is used to retrieve the required output phase with a known freeform lens. Actually, by the mapping-type Fourier pair synthesis method proposed in the functional design (Section 3.2), the output wavefront phase can be retrieved with the mapping of the electric fields in a Fourier pair without solving a complex differential equation.

In our work, an iterative approach is also developed for the design of the freeform surface. The functional design procedure described in the previous chapter is used in every iteration for the output wavefront phase retrieval.

For the introduction of the design algorithm, the field tracing diagram of a far-field light-shaping system is recalled here, as shown in Fig. 4.10.

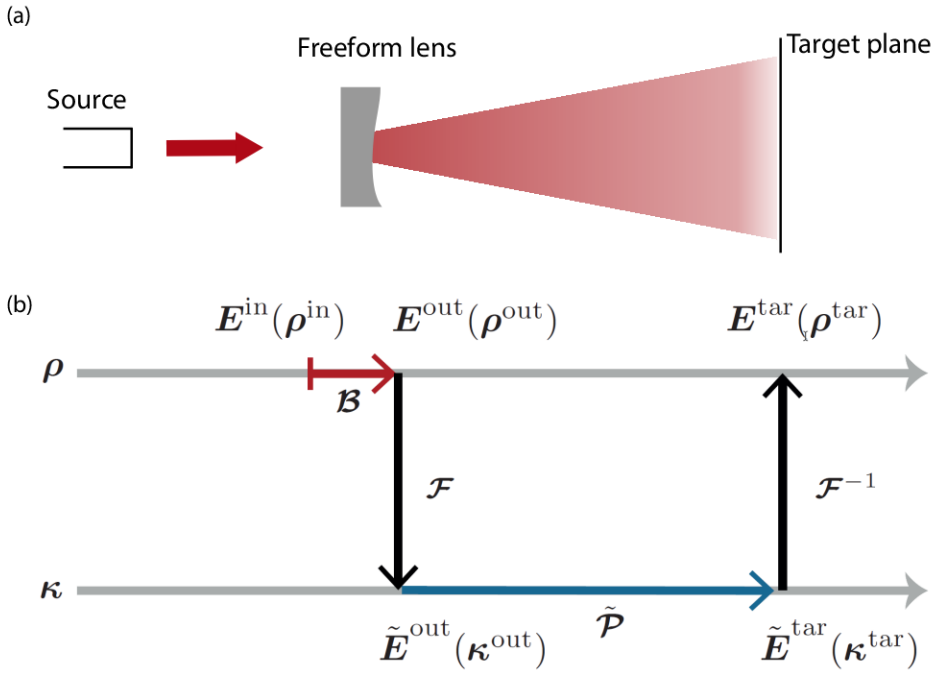

Fig. 4.10 The field tracing diagram illustrates the modeling techniques for light-shaping system with the freeform lens.

In the field tracing diagram, $\mathcal{B}$ is the LPIA operator used to model the functionality of the freeform lens, as described in the last subsection. Therefore, with the freeform lens, $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ can be calculated by

$$\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}}) = \mathcal{B}^{\mathrm{LPIA}}\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}}). \tag{4.13}$$

Following the same inverse thinking logic in the functional design procedures, the field $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ is calculated by inverse field tracing from $\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})$:

$$\begin{aligned}\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}}) &= \tilde{\mathcal{P}}^{-1}\mathcal{F}\left\{\boldsymbol{E}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})\right\} \\ &= \exp(-\mathrm{i}\check{k}_z(\boldsymbol{\kappa}^{\mathrm{out}})\Delta z)\cdot\mathcal{F}\left\{\tilde{\boldsymbol{E}}^{\mathrm{tar}}(\boldsymbol{\rho}^{\mathrm{tar}})\right\}.\end{aligned} \tag{4.14}$$

In the design, the amplitude of $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ is considered as the amplitude constraint. Therefore, a proper freeform surface for the task would generate $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$, which satisfies the constraint of $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ via a Fourier transform.

The design algorithm starts with an initial freeform lens. The freeform surface of the lens is adapted for generating a required output field $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ under the amplitude constraint of $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$. An iterative way is introduced for searching the freeform surface. In each iteration, $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is obtained by Eq. (4.13). The amplitude of both $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ is applied to recalculated a wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ for $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$. Then the freeform surface is reconstructed with the updated $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$. The procedure is repeated until the amplitude constraint of $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ is satisfied.

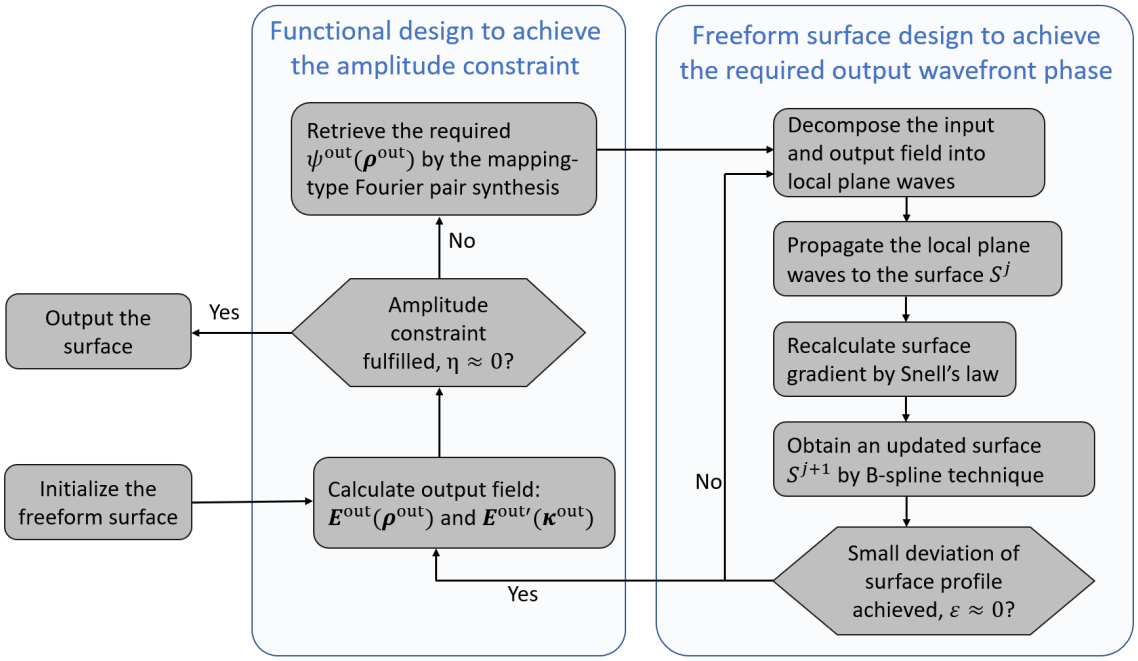

Fig. 4.11 The flow chart of the freeform surface design algorithm.

Freeform surface design algorithm Fig. 4.11 provides a flow chart illustrating the work flow of the design. The design algorithm is described in detail as follows:

1. A initial surface is set in the system, which is defined as $S^0 = \{\boldsymbol{r} \in \mathbb{R}^3 | \boldsymbol{r} = (\boldsymbol{\rho}, z^0(\boldsymbol{\rho}))\}$.
2. Both $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ are calculated by Eq. (4.13) and Eq. (4.14), respectively. Then the amplitudes of both fields are extracted for the retrieval of $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$, where the mapping-type Fourier pair synthesis method is applied.
3. An internal iterative process is introduced to update the freeform surface to achieve the obtained output wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$. Here, the local plane wave approximation (LPWA) is applied for the fields $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$, and the LPIA is assumed for the freeform lens.

(a) In the $j-$iteration, the height profile of the freeform surface is indicated by $z^{0,j}(\boldsymbol{\rho})$.

(b) On both the input and output plane of the freeform lens, the fields $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ are decomposed as local plane waves, with the transversal components of each wave vector calculated by

$$\boldsymbol{\kappa}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}}) = \nabla\psi^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}}), \tag{4.15}$$

$$\boldsymbol{\kappa}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}}) = \nabla\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}}), \tag{4.16}$$

where $\psi^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ and $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ are the wavefront phase of both fields. Their wave vector is then derived with

$$\boldsymbol{k}^{\mathrm{in}} = \left(\boldsymbol{\kappa}^{\mathrm{in}}, \sqrt{{k_0}^2(n^{\mathrm{in}})^2 - (\boldsymbol{\kappa}^{\mathrm{in}})^2}\right), \tag{4.17}$$

$$\boldsymbol{k}^{\mathrm{out}} = \left(\boldsymbol{\kappa}^{\mathrm{out}}, \sqrt{{k_0}^2(n^{\mathrm{out}})^2 - (\boldsymbol{\kappa}^{\mathrm{out}})^2}\right), \tag{4.18}$$

where n^{in} and n^{out} are the refractive indices in front of and behind the lens, respectively.

(c) Both sets of local plane waves are propagated to the freeform surface, where the intersection positions are indicated by $\boldsymbol{r}\left(\boldsymbol{\rho}^{\mathrm{in}}\right) = \left(\boldsymbol{\rho}(\boldsymbol{\rho}^{\mathrm{in}}), z^{0,j}(\boldsymbol{\rho}(\boldsymbol{\rho}^{\mathrm{in}}))\right)$ and $\boldsymbol{r}\left(\boldsymbol{\rho}^{\mathrm{out}}\right) = (\boldsymbol{\rho}(\boldsymbol{\rho}^{\mathrm{out}}), z^{0,j}(\boldsymbol{\rho}(\boldsymbol{\rho}^{\mathrm{out}})))$, respectively. The wave vectors of each local plane at the intersection positions are indicated as $\boldsymbol{k}^{\mathrm{in}}\left(\boldsymbol{r}(\boldsymbol{\rho}^{\mathrm{in}})\right)$ and $\boldsymbol{k}^{\mathrm{out}}\left(\boldsymbol{r}(\boldsymbol{\rho}^{\mathrm{out}})\right)$.

(d) On the position of $\boldsymbol{r} = (\boldsymbol{\rho}, z^{0,j}(\boldsymbol{\rho}))$, the normal vector $\boldsymbol{N}(\boldsymbol{r})$ of the freeform surface is recalculated with respect to the local plane waves from the front and the back using Snell's law. If the freeform surface is the front surface of the lens,

$$\boldsymbol{N}(\boldsymbol{r}) = n^{\mathrm{lens}}\boldsymbol{k}^{\mathrm{out}}(\boldsymbol{r}) - n^{\mathrm{in}}\boldsymbol{k}^{\mathrm{in}}(\boldsymbol{r}). \tag{4.19}$$

If the freeform surface is the back surface of the lens,

$$\boldsymbol{N}(\boldsymbol{r}) = n^{\mathrm{out}}\boldsymbol{k}^{\mathrm{out}}(\boldsymbol{r}) - n^{\mathrm{lens}}\boldsymbol{k}^{\mathrm{in}}(\boldsymbol{r}). \tag{4.20}$$

n^{lens} is the refractive index in the freeform lens. In practice, the samples of local plane waves from front and back may intersect on different positions on the surface. To apply Snell's law, the wave vectors of both sets of local plane waves should be interpolated at the same positions. The B-spline interpolation technique is applied here for the interpolation of $\boldsymbol{k}^{\mathrm{in}}\left(\boldsymbol{r}(\boldsymbol{\rho}^{\mathrm{in}})\right)$ and $\boldsymbol{k}^{\mathrm{out}}\left(\boldsymbol{r}(\boldsymbol{\rho}^{\mathrm{out}})\right)$.

(e) The gradient of the freeform surface is derived from its recalculated normal vector by

$$\nabla z^{0,j+1}(\boldsymbol{\rho}) = \left(-\frac{N_x(\boldsymbol{\rho})}{N_z(\boldsymbol{\rho})}, -\frac{N_y(\boldsymbol{\rho})}{N_z(\boldsymbol{\rho})} \right), \tag{4.21}$$

where N_x, N_y, N_z are three components of the vector $\boldsymbol{N}$

(f) The B-spline integration technique is applied to obtain $\nabla z^{0,j+1}(\boldsymbol{\rho})$ from its gradient data, and constructing an updated surface profile $z^{0,j+1}(\boldsymbol{\rho})$, which is represented by B-spline functions.

(g) The root-mean-square (RMS) of the deviation between the height profiles of j and $j+1$ iteration is calculated by

$$\varepsilon = \sqrt{\frac{\sum_{\boldsymbol{\rho}} \left(z^{0,j+1}(\boldsymbol{\rho}) - z^{0,j}(\boldsymbol{\rho}) \right)^2}{N_{\boldsymbol{\rho}}}}, \tag{4.22}$$

where $N_{\boldsymbol{\rho}}$ is the sampling number of $\boldsymbol{\rho}$.

(h) Iteratively perform from step (c) to (g) until a small value of ε is reached.

So far, an updated freeform surface S^1 is found for realizing the required output wavefront phase $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$.

4. The field $\boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}})$ is traced through the updated freeform lens and the output field in the $\boldsymbol{\kappa}$ domain is recalculated as

$$\tilde{\boldsymbol{E}}^{\mathrm{out}'}(\boldsymbol{\kappa}^{\mathrm{out}}) = \boldsymbol{\mathcal{F}} \left\{ \boldsymbol{\mathcal{B}}^{\mathrm{LPIA}} \boldsymbol{E}^{\mathrm{in}}(\boldsymbol{\rho}^{\mathrm{in}}) \right\}. \tag{4.23}$$

5. The fulfillment of the design constraint is checked with the freeform lens. We use the relative RMS deviation between the amplitude of $\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ and $\tilde{\boldsymbol{E}}^{\mathrm{out}'}(\boldsymbol{\kappa}^{\mathrm{out}})$ as the merit function,

$$\eta = \sqrt{\frac{\sum_{\boldsymbol{\kappa}^{\mathrm{out}}} \left(\left\| \tilde{\boldsymbol{E}}^{\mathrm{out}'}(\boldsymbol{\kappa}^{\mathrm{out}}) \right\| - \alpha \left\| \tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}}) \right\| \right)^2}{\sum_{\boldsymbol{\kappa}^{\mathrm{out}}} \alpha \left\| \tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}}) \right\|}}, \tag{4.24}$$

where $N_{\boldsymbol{\kappa}^{\mathrm{out}}}$ is the sampling number of $\boldsymbol{\kappa}^{\mathrm{out}}$. α is a scaling factor given by

$$\alpha = \frac{\sum_{\boldsymbol{\kappa}^{\mathrm{out}}} \left\| \tilde{\boldsymbol{E}}^{\mathrm{out}'}(\boldsymbol{\kappa}^{\mathrm{out}}) \right\| \left\| \tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}}) \right\|}{\sum_{\boldsymbol{\kappa}^{\mathrm{out}}} \left\| \tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}}) \right\|^2}.$$

If the value of η is close to zero, the amplitude constraint is achieved, and the freeform surface is obtained. Otherwise, the freeform surface is altered by performing the procedure from step (1) to (4) iteratively until a small number of η is achieved.

The algorithm here is to design a lens with a freeform surface and a predefined surface. It has no restriction about the source or the shape of the predefined surface. Moreover, the freeform surface can be designed flexibly as either the front surface or back surface of the lens. In fact, by the iterative procedure, the output amplitude obtained with the freeform lens includes the Fresnel effect from the lens. Therefore, the algorithm also takes the Fresnel effect into account and compensates for it during the design. The iterative procedure can also be used in the design of the HOE to compensate for the grating effect from the HOE.

4.2.3 Demonstration with examples

A light-shaping example is demonstrated in the following to verify the algorithm. In the example, the given source is a plane wave, with a beam size of $1 \times 1\mathrm{mm}$ and a wavelength of $532\mathrm{nm}$. The target irradiance distribution is shown as in Fig. 4.12 (a), with size of $1.2\mathrm{m} \times 0.8\mathrm{m}$, which is located at a vertical plane $1\mathrm{m}$ distance away from the freeform lens. Therefore, the system is in a non-paraxial situation with the outgoing light having the divergent angle of $62\,^\circ \times 44\,^\circ$. The freeform lens is set with a predefined planar surface as its front surface, and the back surface is to be designed. The refractive index of the medium in the freeform lens is set as 1.52.

For the design of the freeform surface, the algorithm starts with an initial surface designed by the ray mapping method [48]. Because the system is non-paraxial, the surface gradient obtained from the ray mapping method suffers from the non-integrable problem. To construct a surface with the non-integrable gradient data, a surface is represented by B-spline functions, with the same formula of Eq. (3.63). The derivative formulas of the B-spline functions are derived as shown in Eq. (3.64). The surface gradient data is fitted with the B-spline derivative formulas by a standard least-squares optimization method, and the coefficients of the B-spline functions are concluded. In this way, an approximated freeform surface is constructed.

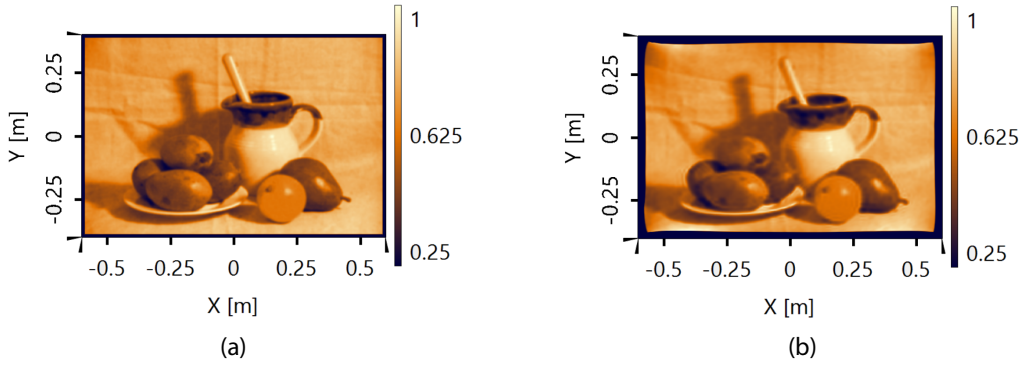

Fig. 4.12 (a) The target irradiance distribution (normalized). (b) Irradiance distribution obtained by field-tracing simulation with the initial freeform lens, where the freeform surface is designed by the ray mapping method.

A field-tracing simulation with the initial freeform lens is performed, and the detector gives the irradiance distribution shown as Fig. 4.12 (b). From the result, one can see that the irradiance pattern is deformed compared to the target, especially at the corners of the pattern. Therefore, although the freeform surface is an optimum result from the gradient data, it exhibits errors in its profile. However, this approximate surface is adequate for use as an initial surface to proceed with the design algorithm presented in the last section.

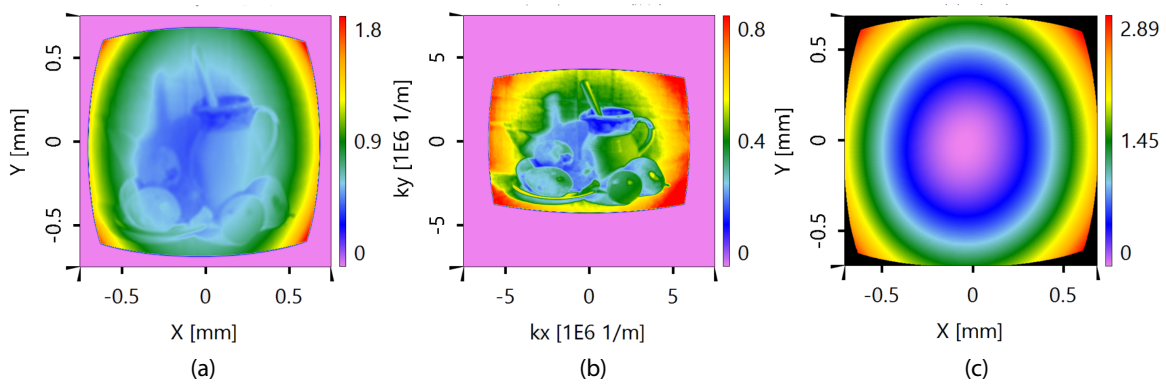

Fig. 4.13 (a) $\left\|\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})\right\|^2$ (unit: $\mathrm{V}^2/\mathrm{m}^2$): obtained by propagating the input field through the initial freeform lens. (b) $\left\|\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})\right\|^2$ (unit: $\mathrm{V}^2/\mathrm{m}^2$): obtained by inverse propagation from the target field. (c) $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ (unit: $10^3\mathrm{rad}$): required output wavefront phase obtained by the mapping-type Fourier synthesis method.

The design logic in this dissertation is followed to search for a proper freeform surface for the task. The functional design is considered, and then the structural design is based on the information from the functional design. Therefore, the required output wavefront phase is required to realize the WPR function. The retrieval of the output wavefront phase is discussed

in the previous chapter for the functional design. Under the homeomorphic assumption, the required output wavefront phase is obtained by the mapping-type Fourier pair synthesis method. The electric field behind the freeform lens in both domains are requested for that purpose.

The field behind the freeform lens in the $\boldsymbol{\rho}-$domain $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is obtained by field tracing with the initial freeform lens. Fig. 4.13 (a) shows the squared amplitude $\left\|\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})\right\|^2$. $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is coupled to the structure of the freeform lens so that the changing of the freeform surface will modulate $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$. Therefore, the iterative algorithm tends to minimize the changing of the freeform surface profile.

$\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})$ is obtained from the target field, and its amplitude is given as a constraint for the design because its phase can be considered as freedom under the far-field assumption of the target field. Fig. 4.13 (b) shows the squared amplitude $\left\|\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})\right\|^2$. A mapping $\boldsymbol{\kappa}(\boldsymbol{\rho})$ between $\left\|\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})\right\|^2$ and $\left\|\tilde{\boldsymbol{E}}^{\mathrm{out}}(\boldsymbol{\kappa}^{\mathrm{out}})\right\|^2$ is found by the mapping-type Fourier pair synthesis method. The required output wavefront phase for $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is then calculated based on the mapping function $\boldsymbol{\kappa}(\boldsymbol{\rho})$, as shown in Fig. 4.13 (c).

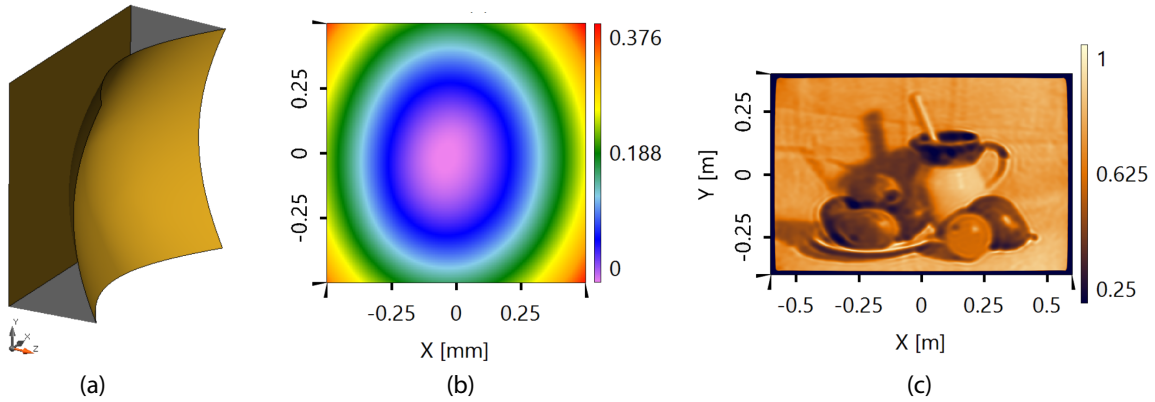

Fig. 4.14 (a) The 3D view of the designed freeform lens. (b) The height profile of the freeform surface (unit: mm). (c) Detector result: the irradiance distribution obtained by field tracing with the designed freeform lens.

The obtained $\psi^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is used for the design of the freeform surface by the method with the LPIA modeling technique. An internal iterative procedure is applied to construct the freeform surface from both the input/output wavefront phases. In this example, the procedure converges quickly so that, with 6 iterations, the deviation of the height profile between two iterations converges to nanometer magnitude. The new freeform surface is then applied for the next phase retrieval. The alternative process between the functional and the structural design converges to a surface that meets the design constraint with $\eta = 0.06$. Fig. 4.14 (a) shows the result of the final freeform lens in a 3D view. The height profile of the freeform

surface is shown in Fig. 4.14 (b). A simulation is performed in the software VirtualLab Fusion [95] to investigate the designed freeform lens. Fig. 4.14 (c) is the detector result, which provides the simulated irradiance distribution with the freeform lens. The result is coincident with the target pattern in Fig. 4.12, with a relative root-mean-square deviation (RRMSD) of 0.05 in the irradiance. Compared to the distorted pattern shown in Fig. 4.12 (b), the irradiance distribution here indicates that the performance of the freeform surface has been improved from the surface designed by the ray mapping method.

It is worth to be noted that, the small deviation from the target irradiance also indicated that the Fresnel effect of the freeform lens has been compensated in the design. In each iteration of the algorithm, for constructing a new freeform surface, the required $\boldsymbol{E}^{\mathrm{out}}(\boldsymbol{\rho}^{\mathrm{out}})$ is calculated from Eq. (4.13), where the LPIA operator in the calculation includes the Fresnel effect. Therefore, the design naturally considers and compensates all the physical effect from the freeform lens.

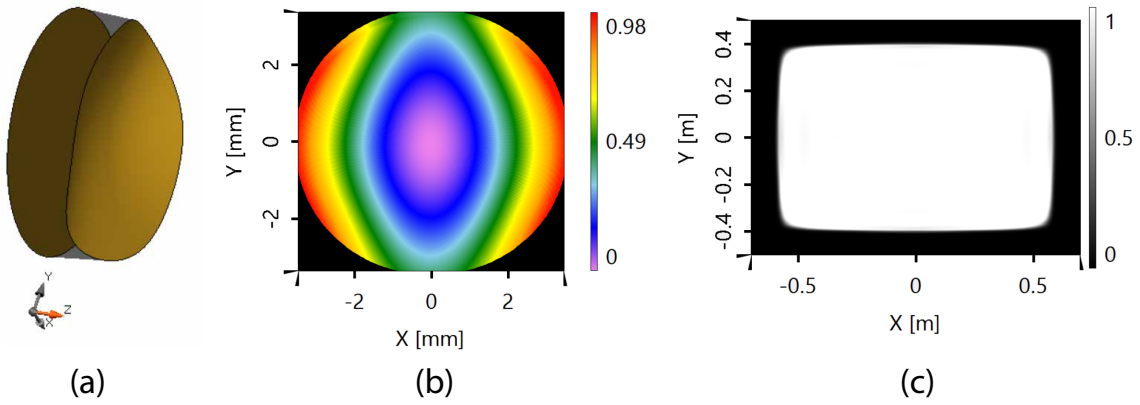

Fig. 4.15 (a) The designed freeform lens. (b) The height profile of the freeform surface (Unit: mm). (c) The simulated irradiance (normalized) with the freeform lens.

The design algorithm of the freeform surface has no restrictions on the input light source. It can be applied to the input field with any type of wavefront phase and arbitrary aperture shape. Another example is illustrated in the following to demonstrate the design. Here, the example is the same as the second example in Section 3.3.2 for functional design discussion. A spherical wave with a circular aperture is to be shaped to a homogeneous rectangle pattern in the far field. The full divergence angle of the spherical wave is $30°$, with a wavelength of $532\mathrm{nm}$. The target pattern is located on a vertical plane $1\ \mathrm{m}$ away from the optical element, with a size of $1.2 \times 0.8\ \mathrm{m}$.

The freeform lens is also set with a predefined planar surface as its front surface. The algorithm in the previous subsection is adopted for the design of the freeform surface as the second surface of the lens. The thickness of the lens is set as $1\ \mathrm{mm}$. The initial surface for the

design is set as a spherical surface. Then the alternative process of the output wavefront phase retrieval and the surface construction is performed. With 6 iterations, the surface converges with $\eta = 0.012$. Fig. 4.15 (a) shows the 3D view of the designed freeform lens. The freeform surface of the lens is shown in Fig. 4.15 (b). A field tracing simulation is performed with the spherical wave and the designed freeform lens. The irradiance distribution illustrated in Fig. 4.15 (c) indicates that the lens designed based on the output phase solve the shaping problem in a good way.

4.3 Summary

In summary, the functional design provides the basic result for the design of the structure. The algorithm for constructing the HOE and freeform lens is developed according to the physical models to realize the required output wavefront phase discussed in the functional design. The physical structure of the HOE and the freeform lens naturally break the phase-only assumption in the functional design. Therefore, the compensation of the effect introduced by the structure is considered in the structural design step.

Due to the grating effect of the HOE, the designed HOE from the WPR function may introduce amplitude modulation to the input field. Additional diffraction orders from the HOE will also introduce stray light to the target signal. The grating behavior indicates that in the large period area, the transmission efficiency of the $+1$ order is independent of the period, and the $+1$ order will dominate the diffractive light. Therefore, a hybrid component is suggested as a solution to compensate for the grating effect. The hybrid component contains a curved surface that introduces a wavefront phase similar to the required one, and the HOE contributes to the residual phase. In this way, the grating period of the HOE can be released so that the grating effect from the HOE is reduced.

For the freeform lens, the 3D topology of both surfaces of the lens also modulates the amplitude of the input field and even changes the shape of the field. Although the freeform lens can be calculated realizing the required output wavefront phase obtained from the functional design step, the amplitude of the output field no longer satisfies the constraint from the target. Therefore, an iterative algorithm is developed by alternatively performing the output wavefront phase retrieval and freeform surface construction. In this way, both the required output wavefront phase and the freeform surface are adjusted so that the output field meets the constraint from the target.

The design examples in this chapter demonstrate that the strategy for the HOE design and the algorithm developed for the freeform surface design provide well-suited solutions for the light-shaping tasks.

Chapter 5

Summary and outlook

In this work, the far-field light-shaping problem is analyzed and tackled from a physical-optics point of view.

Within the field tracing framework, the electric field of light passing through the light-shaping system is connected with different operators, where the operators together build an algorithm to trace the field from the source plane to the target plane. A homeomorphism between the field of the source and the target can be established when all the operators in the system can be performed pointwise. In typical light-shaping systems, the most critical point for this homeomorphism is on the Fourier transform for the free space propagation behind the optical element. The homeomorphism is established if the Fourier transform can be accurately performed by the HFT method. The homeomorphic assumption is usually used in those geometric-optics-based algorithms to design light-shaping elements, such as the ray-mapping method for designing a freeform surface, where a one-to-one map between the irradiance of the source and target is assumed. However, the validity of the homeomorphic assumption should be tested with a physical-optics modeling method rather than a geometric-optics, ray-tracing simulation.

An example shows that if the design and modeling are based on the same assumption, extra physical effects, such as diffraction, may not be revealed in the simulation. The example makes it obvious that for a system that does not appear in a homeomorphic situation, a geometrics-optics design cannot fulfill the task. As a rule of thumb, the more non-paraxial the system is, the more accurate the homeomorphic assumption will be.

To systematically tackle the light-shaping problem, the design is performed with a two-step strategy. The functionality of the required optical element is considered first, followed with the structural design to realize its functionality. The functional embodiment is determined as a WPR function because controlling the input field phase is more achievable for most optical elements. With inverse thinking of the modeling techniques, the design mainly

requires a Fourier pair synthesis such that the output phase is determined. The IFTA is one of the methods for phase retrieval. However, it may lead to a stagnation issue when the fields of the Fourier pair are closed to a homeomorphic situation. To this end, a mapping-type Fourier pair synthesis method is proposed. Via this method, a mapping between the Fourier pair is solved, and the required wavefront phase is derived from the mapping. A similar algorithm in literature has been proposed for the phase retrieval also based on a mapping relationship. However, the mapping is solved between the irradiance from energy conservation. Due to the nonlinear relation between the wavefront phase gradient and the mapping, the gradient data usually suffers from the integrable problem, especially in a non-paraxial case. For our method, the mapping is solved between the electric field from Parseval's equation such that the resulting mapping function is equivalent to the gradient data of the wavefront phase. Therefore, the curl-free characteristic of the mapping functions guarantees the obtained gradient data is integrable in both paraxial and non-paraxial situations. After the required output wavefront phase is retrieved, the WPR is nothing other than subtracting the input and output wavefront phase.

The design of the structure embodiment of the optical element often takes advantage of the results from the functional embodiment. The design of both a HOE and a freeform lens for light shaping is demonstrated. The algorithms are based on their physical models, in which the LLGA is addressed for the HOE, and the LPIA is applied for the freeform lens.

For the HOE design, its element function is obtained first, which is the same as the WPR function from the functional design step. The period of the local gratings is then derived from its element function. A Gaussian-to-top-hat shaping task is taken as an example to demonstrate the algorithm. The amplitude modulation from the HOE structure is an extra effect from the design, which introduces errors and stray light into the target pattern. A hybrid component by adding a curved surface to the HOE is suggested to reduce this effect. The curved surface contributes most of the required output wavefront phase, while the hologram part of the hybrid component contributes to the small special modulations. In this way, the period of the local gratings is increased, whereby the amplitude modulation is relaxed.

For the design of a freeform lens, the freeform lens is set as a combination of a predefined surface and a freeform surface to be designed. The freeform surface is constructed using the obtained required output wavefront phase in the functional design step, with an inverse LPIA method. The algorithm starts with an initial surface profile, and the retrieval of the output wavefront phase and the construction of the freeform surface alternatively proceed until a merit function is achieved. Here the merit function is defined as the amplitude constraint for the field in the $\boldsymbol{\kappa}-$domain behind the freeform lens, where the field is concluded from the target irradiance distribution. In this way, the designed freeform surface is a smooth surface

Table 5.1 The outlook from the work in this thesis.

	Source	**Target irradiance distribution**
current work	single source	located in far-field zone
future work	extended source	located with limited distance

that fulfills the task even in a non-paraxial situation. The algorithm has no restriction on the input wavefront and the shape of the predefined surface. Moreover, the freeform surface can be designed flexibly as either the front or back surface of the lens.

The work in this thesis tackles the light-shaping problem with a single-mode source and target irradiance in the far-field zone. In future research, the work can be extended in two aspects. As shown in Table 5.1, for the source aspect, the shaping problem can be addressed for extended source, which is usually modeled with multi-mode; for the target aspect, the shaping problem can be reformulated without the far-field assumption that the target irradiance distribution is located in a limited distance. This thesis provides basic knowledge for the extended work, and new algorithms can be developed based on the current work.

List of figures

List of tables

References

[1] R. Winston, L. Jiang, and M. Ricketts, "Nonimaging optics: a tutorial," *Adv. Opt. Photon.*, vol. 10, pp. 484–511, Jun 2018.

[2] P.-A. Blanche, *Optical Holography: Materials, Theory and Applications*. Elsevier, 2020.

[3] J.-H. Park, "Recent progress in computer-generated holography for three-dimensional scenes," *Journal of Information Display*, vol. 18, no. 1, pp. 1–12, 2017.

[4] T. Zhan, J. Xiong, J. Zou, and S.-T. Wu, "Multifocal displays: review and prospect," *PhotoniX*, vol. 1, 12 2020.

[5] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[6] M. Oikawa, T. Shimobaba, N. Masuda, and T. Ito, "Computer-generated hologram using an approximate Fresnel integral," *Journal of Optics*, vol. 13, p. 075405, jun 2011.

[7] A. D. Stein, Z. Wang, and J. S. Leigh, "Computer-generated holograms: A simplified ray-tracing approach," *Computers in Physics*, vol. 6, no. 4, pp. 389–392, 1992.

[8] K. Matsushima and M. Takai, "Recurrence formulas for fast creation of synthetic three-dimensional holograms," *Appl. Opt.*, vol. 39, pp. 6587–6594, Dec 2000.

[9] M. E. Lucente, "Interactive computation of holograms using a look-up table," *Journal of Electronic Imaging*, vol. 2, no. 1, pp. 28 – 34, 1993.

[10] T. Shimobaba, H. Nakayama, N. Masuda, and T. Ito, "Rapid calculation algorithm of fresnel computer-generated-hologram using look-up table and wavefront-recording plane methods for three-dimensional display," *Opt. Express*, vol. 18, pp. 19504–19509, Sep 2010.

[11] J. Jia, Y. Wang, J. Liu, X. Li, Y. Pan, Z. Sun, B. Zhang, Q. Zhao, and W. Jiang, "Reducing the memory usage for effectivecomputer-generated hologram calculation using compressed look-up table in full-color holographic display," *Appl. Opt.*, vol. 52, pp. 1404–1412, Mar 2013.

[12] S. Jiao, Z. Zhuang, and W. Zou, "Fast computer generated hologram calculation with a mini look-up table incorporated with radial symmetric interpolation," *Opt. Express*, vol. 25, pp. 112–123, Jan 2017.

[13] Y.-L. Li, D. Wang, N.-N. Li, and Q.-H. Wang, "Fast hologram generation method based on the optimal segmentation of a sub-cgh," *Opt. Express*, vol. 28, pp. 32185–32198, Oct 2020.

[14] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[15] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, pp. 2758–2769, Aug 1982.

[16] F. Wyrowski and O. Bryngdahl, "Iterative fourier-transform algorithm applied to computer holography," *J. Opt. Soc. Am. A*, vol. 5, pp. 1058–1065, Jul 1988.

[17] F. Wyrowski, "Diffractive optical elements: iterative calculation of quantized, blazed phase structures," *J. Opt. Soc. Am. A*, vol. 7, pp. 961–969, Jun 1990.

[18] D. Prongué, H. P. Herzig, R. Dändliker, and M. T. Gale, "Optimized kinoform structures for highly efficient fan-out elements," *Appl. Opt.*, vol. 31, pp. 5706–5711, Sep 1992.

[19] V. Arrizón, M. Testorf, S. Sinzinger, and J. Jahns, "Iterative optimization of phase-only diffractive optical elements based on a lenslet array," *J. Opt. Soc. Am. A*, vol. 17, pp. 2157–2164, Dec 2000.

[20] K.-H. Brenner, "Method for designing arbitrary two-dimensional continuous phase elements," *Opt. Lett.*, vol. 25, pp. 31–33, Jan 2000.

[21] H.-E. Hwang, H. T. Chang, and W.-N. Lie, "Fast double-phase retrieval in fresnel domain using modified gerchberg-saxton algorithm for lensless optical security systems," *Opt. Express*, vol. 17, pp. 13700–13710, Aug 2009.

[22] F. Zhang, J. Zhu, W. Yue, J. Wang, Q. Song, G. Situ, F. Wyrowski, and H. Huang, "An approach to increase efficiency of DOE based pupil shaping technique for off-axis illumination in optical lithography," *Opt. Express*, vol. 23, pp. 4482–4493, Feb 2015.

[23] H. Pang, J. Wang, A. Cao, and Q. Deng, "High-accuracy method for holographic image projection with suppressed speckle noise," *Opt. Express*, vol. 24, pp. 22766–22776, Oct 2016.

[24] F. J. Salgado-Remacha, "Reducing the variability in random-phase initialized gerchberg-saxton algorithm," *Optics and Laser Technology*, vol. 85, pp. 30–34, 2016.

[25] C.-Y. Chen, W.-C. Li, H.-T. Chang, C.-H. Chuang, and T.-J. Chang, "3-d modified gerchberg–saxton algorithm developed for panoramic computer-generated phase-only holographic display," *J. Opt. Soc. Am. B*, vol. 34, pp. B42–B48, May 2017.

[26] H. Wang, W. Yue, Q. Song, J. Liu, and G. Situ, "A hybrid Gerchberg-Saxton-like algorithm for DOE and CGH calculation," *Optics and Lasers in Engineering*, vol. 89, pp. 109 – 115, 2017. 3DIM-DS 2015: Optical Image Processing in the context of 3D Imaging, Metrology, and Data Security.

[27] M. Golub, I. Sisakyan, and V. Soifer, "Infra-red radiation focusators," *Optics and Lasers in Engineering*, vol. 15, no. 5, pp. 297 – 309, 1991. Special Issue on Computer Optics in the USSR.

[28] T. Dresel, M. Beyerlein, and J. Schwider, "Design and fabrication of computer-generated beam-shaping holograms," *Appl. Opt.*, vol. 35, pp. 4615–4621, Aug 1996.

[29] Y. Arieli, N. Eisenberg, A. Lewis, and I. Glaser, "Geometrical-transformation approach to optical two-dimensional beam shaping," *Appl. Opt.*, vol. 36, pp. 9129–9131, Dec 1997.

[30] H. Aagedal, M. Schmid, S. Egner, J. Müller-Quade, T. Beth, and F. Wyrowski, "Analytical beam shaping with application to laser-diode arrays," *J. Opt. Soc. Am. A*, vol. 14, pp. 1549–1553, Jul 1997.

[31] J. Turunen and F. Wyrowski, *Diffractive optics for industrial and commercial applications*. Akademie Verlag, Berlin, 1997. book.

[32] A. Hermerschmidt, H. J. Eichler, S. Teiwes, and J. Schwartz, "Design of diffractive beam-shaping elements for nonuniform illumination waves," *Proc. SPIE*, vol. 3291, pp. 40–48, 1998.

[33] T. Kaempfe, E.-B. Kley, and A. Tuennermann, "Hybrid approach to the design of refractive beam shaping elements," in *Laser Beam Shaping VI* (F. M. Dickey and D. L. Shealy, eds.), vol. 5876, pp. 163 – 175, International Society for Optics and Photonics, SPIE, 2005.

[34] F. M. Dickey, *Laser beam shaping: theory and techniques*. CRC press, 2018.

[35] F. Fang, X. Zhang, A. Weckenmann, G. Zhang, and C. Evans, "Manufacturing and measurement of freeform optics," *CIRP Annals*, vol. 62, no. 2, pp. 823 – 846, 2013.

[36] R. Wu, Z. Feng, Z. Zheng, R. Liang, P. Benítez, J. C. Miñano, and F. Duerr, "Design of freeform illumination optics," *Laser & Photonics Reviews*, vol. 12, no. 7, p. 1700310, 2018.

[37] H. Ries and J. Muschaweck, "Tailored freeform optical surfaces," *J. Opt. Soc. Am. A*, vol. 19, pp. 590–595, Mar 2002.

[38] R. Wu, K. Li, P. Liu, Z. Zheng, H. Li, and X. Liu, "Conceptual design of dedicated road lighting for city park and housing estate," *Appl. Opt.*, vol. 52, pp. 5272–5278, Jul 2013.

[39] R. Wu, P. Liu, Y. Zhang, Z. Zheng, H. Li, and X. Liu, "A mathematical model of the single freeform surface design for collimated beam shaping," *Opt. Express*, vol. 21, pp. 20974–20989, Sep 2013.

[40] K. Brix, Y. Hafizogullari, and A. Platen, "Designing illumination lenses and mirrors by the numerical solution of Monge Ampère equations," *J. Opt. Soc. Am. A*, vol. 32, pp. 2227–2236, Nov 2015.

[41] S. Chang, R. Wu, L. An, and Z. Zheng, "Design beam shapers with double freeform surfaces to form a desired wavefront with prescribed illumination pattern by solving a Monge-Ampère type equation," *Journal of Optics*, vol. 18, no. 12, p. 125602, 2016.

[42] L. L. Doskolovich, A. A. Mingazov, D. A. Bykov, E. S. Andreev, and E. A. Bezus, "Variational approach to calculation of light field eikonal function for illuminating a prescribed region," *Opt. Express*, vol. 25, pp. 26378–26392, Oct 2017.

[43] C. Bosel and H. Gross, "Single freeform surface design for prescribed input wavefront and target irradiance," *J. Opt. Soc. Am. A*, vol. 34, pp. 1490–1499, Sep 2017.

[44] L. B. Romijn, J. H. M. ten Thije Boonkkamp, and W. L. IJzerman, "Freeform lens design for a point source and far-field target," *J. Opt. Soc. Am. A*, vol. 36, pp. 1926–1939, Nov 2019.

[45] D. Bykov, L. Doskolovich, A. Mingazov, and E. Bezus, "Optics mass transportation problem in the design of freeform optical elements generating far-field irradiance distributions for plane incident beam," *Applied Optics*, vol. 58, p. 9131, 11 2019.

[46] L. Wang, K. Qian, and Y. Luo, "Discontinuous free-form lens design for prescribed irradiance," *Appl. Opt.*, vol. 46, pp. 3716–3723, Jun 2007.

[47] R. Wu, H. Li, Z. Zheng, and X. Liu, "Freeform lens arrays for off-axis illumination in an optical lithography system," *Appl. Opt.*, vol. 50, pp. 725–732, Feb 2011.

[48] A. Bauerle, A. Bruneton, R. Wester, J. Stollenwerk, and P. Loosen, "Algorithm for irradiance tailoring using multiple freeform optical surfaces," *Opt. Express*, vol. 20, pp. 14477–14485, Jun 2012.

[49] Z. Feng, L. Huang, G. Jin, and M. Gong, "Designing double freeform optical surfaces for controlling both irradiance and wavefront," *Opt. Express*, vol. 21, pp. 28693–28701, Nov 2013.

[50] A. Bruneton, A. Bäuerle, R. Wester, J. Stollenwerk, and P. Loosen, "High resolution irradiance tailoring using multiple freeform surfaces," *Opt. Express*, vol. 21, pp. 10563–10571, May 2013.

[51] Z. Feng, L. Huang, M. Gong, and G. Jin, "Beam shaping system design using double freeform optical surfaces," *Opt. Express*, vol. 21, pp. 14728–14735, Jun 2013.

[52] Y. Schwartzburg, R. Testuz, A. Tagliasacchi, and M. Pauly, "High-contrast computational caustic design," *ACM Trans. Graph.*, vol. 33, pp. 74:1–74:11, July 2014.

[53] X. Mao, H. Li, Y. Han, and Y. Luo, "Polar-grids based source-target mapping construction method for designing freeform illumination system for a lighting target with arbitrary shape," *Opt. Express*, vol. 23, pp. 4313–4328, Feb 2015.

[54] C. Bosel and H. Gross, "Ray mapping approach for the efficient design of continuous freeform surfaces," *Opt. Express*, vol. 24, pp. 14271–14282, Jun 2016.

[55] Z. Feng, B. D. Froese, R. Liang, D. Cheng, and Y. Wang, "Simplified freeform optics design for complicated laser beam shaping," *Appl. Opt.*, vol. 56, pp. 9308–9314, Nov 2017.

[56] C. Gannon and R. Liang, "Ray mapping with surface information for freeform illumination design," *Opt. Express*, vol. 25, pp. 9426–9434, Apr 2017.

[57] K. Desnijder, P. Hanselaer, and Y. Meuret, "Ray mapping method for off-axis and non-paraxial freeform illumination lens design," *Opt. Lett.*, vol. 44, pp. 771–774, Feb 2019.

[58] V. Oliker, "Mathematical aspects of design of beam shaping surfaces in geometrical optics," in *Trends in Nonlinear Analysis* (M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, eds.), (Berlin, Heidelberg), pp. 193–224, Springer Berlin Heidelberg, 2003.

[59] D. Michaelis, P. Schreiber, and A. Bräuer, "Cartesian oval representation of freeform optics in illumination systems," *Opt. Lett.*, vol. 36, pp. 918–920, Mar 2011.

[60] C. Canavesi, W. J. Cassarly, and J. P. Rolland, "Target flux estimation by calculating intersections between neighboring conic reflector patches," *Opt. Lett.*, vol. 38, pp. 5012–5015, Dec 2013.

[61] L. L. Doskolovich, K. V. Borisova, M. A. Moiseev, and N. L. Kazanskiy, "Design of mirrors for generating prescribed continuous illuminance distributions on the basis of the supporting quadric method," *Appl. Opt.*, vol. 55, pp. 687–695, Feb 2016.

[62] V. Oliker, "Controlling light with freeform multifocal lens designed with supporting quadric method(SQM)," *Opt. Express*, vol. 25, pp. A58–A72, Feb 2017.

[63] J. S. Schruben, "Formulation of a reflector-design problem for a lighting fixture," *J. Opt. Soc. Am.*, vol. 62, pp. 1498–1501, Dec 1972.

[64] R. Wu, L. Xu, P. Liu, Y. Zhang, Z. Zheng, H. Li, and X. Liu, "Freeform illumination design: a nonlinear boundary problem for the elliptic Monge Ampère equation," *Opt. Lett.*, vol. 38, pp. 229–231, Jan 2013.

[65] F. R. Fournier, W. J. Cassarly, and J. P. Rolland, "Fast freeform reflector generation using source-target maps," *Opt. Express*, vol. 18, pp. 5295–5304, Mar 2010.

[66] C. Villani, *Topics in optimal transportation*. No. 58, American Mathematical Soc., 2003.

[67] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *International Journal of Computer Vision*, vol. 60, pp. 225–240, Dec 2004.

[68] M. M. Sulman, J. Williams, and R. D. Russell, "An efficient approach for the numerical solution of the Monge-Ampère equation," *Applied Numerical Mathematics*, vol. 61, no. 3, pp. 298 – 307, 2011.

[69] R. Wu, Y. Zhang, M. M. Sulman, Z. Zheng, P. Benitez, and J. C. Minano, "Initial design with l2 monge-kantorovich theory for the Monge-Ampère equation method in freeform surface illumination design," *Opt. Express*, vol. 22, pp. 16161–16177, Jun 2014.

[70] C. Prins, R. Beltman, J. ten Thije Boonkkamp, W. IJzerman, and T. Tukker, "A least-squares method for optimal transport using the Monge-Ampère equation," *SIAM Journal on Scientific Computing*, vol. 37, no. 6, pp. B937–B961, 2015.

[71] A. V. Pfeil and F. Wyrowski, "Wave-optical structure design with the local plane-interface approximation," *Journal of Modern Optics*, vol. 47, no. 13, pp. 2335–2350, 2000.

[72] A. Bruneton, A. Bäuerle, R. Wester, J. Stollenwerk, and P. Loosen, "Limitations of the ray mapping approach in freeform optics design," *Opt. Lett.*, vol. 38, pp. 1945–1947, Jun 2013.

[73] S. Wei, D. Ma, Z. Zhengbo, and Z. Fan, "Least-squares ray mapping method for freeform illumination optics design," *Optics Express*, vol. 28, 01 2020.

[74] J. C. Minano and J. C. Gonzalez, "New method of design of nonimaging concentrators," *Appl. Opt.*, vol. 31, pp. 3051–3060, Jun 1992.

[75] P. Benitez, J. C. Minano, J. Blen, R. Mohedano, J. Chaves, O. Dross, M. Hernandez, and W. Falicoff, "Simultaneous multiple surface optical design method in three dimensions," *Optical Engineering*, vol. 43, no. 7, pp. 1489–1502, 2004.

[76] O. Dross, R. Mohedano, P. Benitez, J. C. Minano, J. Chaves, J. Blen, M. Hernandez, and F. Munoz, "Review of SMS design methods and real world applications," in *Nonimaging Optics and Efficient Illumination Systems* (R. Winston and R. J. Koshel, eds.), vol. 5529, pp. 35 – 47, International Society for Optics and Photonics, SPIE, 2004.

[77] J. C. Minano, P. Benítez, J. Blen, and A. Santamaría, "High-efficiency free-form condenser overcoming rotational symmetry limitations," *Opt. Express*, vol. 16, pp. 20193–20205, Dec 2008.

[78] S. Sorgato, R. Mohedano, J. Chaves, M. Hernández, J. Blen, D. Grabovičkić, P. Benítez, J. C. M. nano, H. Thienpont, and F. Duerr, "Compact illumination optic with three freeform surfaces for improved beam control," *Opt. Express*, vol. 25, pp. 29627–29641, Nov 2017.

[79] D. Infante, "Design, simulation, and quality evaluation of micro-optical freeform beam shapers at different illumination conditions," *Applied Optics*, vol. 55, p. 8340, 10 2016.

[80] S. Schmidt, S. Thiele, A. Toulouse, C. Bösel, T. Tiess, A. Herkommer, H. Gross, and H. Giessen, "Tailored micro-optical freeform holograms for integrated complex beam shaping," *Optica*, vol. 7, pp. 1279–1286, Oct 2020.

[81] M. Kuhn, F. Wyrowski, and C. Hellmann, "Non-sequential optical field tracing," in *Advanced Finite Element Methods and Applications* (T. Apel and O. Steinbach, eds.), vol. 66 of *Lecture Notes in Applied and Computational Mechanics*, pp. 257–273, Springer Berlin Heidelberg, 2013.

[82] F. Wyrowski, "Unification of the geometric and diffractive theories of electromagnetic fields," *Proc. DGaO*, 2017.

[83] L. Rabiner, R. Schafer, and C. Rader, "The chirp z-transform algorithm," *IEEE Transactions on Audio and Electroacoustics*, vol. 17, pp. 86–92, June 1969.

[84] Z. Wang, S. Zhang, O. Baladron-Zorita, C. Hellmann, and F. Wyrowski, "Application of the semi-analytical Fourier transform to electromagnetic modeling," *Opt. Express*, vol. 27, no. 11, pp. 15335–15350, 2019.

[85] O. Bryngdahl, "Optical map transformations," *Optics Commun.*, vol. 10, no. 2, pp. 164–168, 1974. test.

[86] Z. Wang, O. Baladron-Zorita, C. Hellmann, and F. Wyrowski, "Theory and algorithm of the homeomorphic Fourier transform for optical simulations," *Opt. Express*, vol. 28, pp. 10552–10571, Mar 2020.

[87] O. Baladron-Zorita, Z. Wang, C. Hellmann, and F. Wyrowski, "Isolating the Gouy phase shift in a full physical-optics solution to the propagation problem," *J. Opt. Soc. Am. A*, vol. 36, no. 9, pp. 1551–1558, 2019.

[88] M. Born and E. Wolf, *Principles of Optics*. Cambridge University Press, 7th ed., 1999.

[89] N. C. Roberts, "Beam shaping by holographic filters," *Appl. Opt.*, vol. 28, pp. 31–32, Jan 1989.

[90] R. Shi, C. Hellmann, and F. Wyrowski, "Physical-optics propagation through curved surfaces," *J. Opt. Soc. Am. A*, vol. 36, pp. 1252–1260, Jul 2019.

[91] J. J. Stamnes, "Waves, rays, and the method of stationary phase," *Opt. Express*, vol. 10, pp. 740–751, Aug 2002.

[92] L. Yang, R. Knoth, C. Hellmann, and F. Wyrowski, "Non-paraxial diffractive and refractive laser beam shaping," *Proc. SPIE*, vol. 10518, 2018.

[93] L. Piegl and W. Tiller, *The NURBS book*. Springer Science & Business Media, 2012.

[94] S. Zwick, R. Feßler, J. Jegorov, and G. Notni, "Resolution limitations for tailored picture-generating freeform surfaces," *Opt. Express*, vol. 20, pp. 3642–3653, Feb 2012.

[95] Physical optics simulation and design software "Wyrowski VirtualLab Fusion", developed by Wyrowski Photonics GmbH, distributed by LightTrans International UG, Jena, Germany.

[96] T. Dresel, M. Beyerlein, and J. Schwider, "Design of computer-generated beam-shaping holograms by iterative finite-element mesh adaption," *Appl. Opt.*, vol. 35, pp. 6865–6874, Dec 1996.

[97] H. Aagedal, M. Schmid, T. Beth, S. Teiwes, and F. Wyrowski, "Theory of speckles in diffractive optics and its application to beam shaping," *Journal of Modern Optics*, vol. 43, no. 7, pp. 1409–1421, 1996.

[98] R. Bräuer, F. Wyrowski, and O. Bryngdahl, "Diffusers in digital holography," *J. Opt. Soc. Am. A*, vol. 8, pp. 572–578, Mar 1991.

[99] O. Ripoll, V. Kettunen, and H. P. Herzig, "Review of iterative Fourier-transform algorithms for beam shaping applications," *Optical Engineering*, vol. 43, no. 11, 2004.

[100] P. M. Hirsch, J. A. Jordan Jr, and L. B. Lesem, "Method of making an object dependent diffuser," Nov. 9 1971.

[101] J. S. Liu and M. R. Taghizadeh, "Iterative algorithm for the design of diffractive phase elements for laser beam shaping," *Opt. Lett.*, vol. 27, pp. 1463–1465, Aug 2002.

[102] D. Engström, A. Frank, J. Backsten, M. Goksör, and J. Bengtsson, "Grid-free 3D multiple spot generation with an efficient single-plane FFT-based algorithm," *Opt. Express*, vol. 17, pp. 9989–10000, Jun 2009.

[103] F. Wyrowski, H. Aagedal, H. Thienpont, P. Chavel, and D. A. B. Miller, "Wave transformation by physical-optics system design, conference, optics in computing," in *INTERNATIONAL JOURNAL OF OPTOELECTRONICS, Optics in computing, Conference, Optics in computing*, vol. 12, pp. 127–144, SPIE, 1998.

[104] L. Yang, I. Badar, C. Hellmann, and F. Wyrowski, "Light-shaping design by a Fourier pair synthesis: the homeomorphic case," *Opt. Express*, vol. 29, pp. 3621–3630, Feb 2021.

[105] K. Desnijder, P. Hanselaer, and Y. Meuret, "Flexible design method for freeform lenses with an arbitrary lens contour," *Opt. Lett.*, vol. 42, pp. 5238–5241, Dec 2017.

[106] Z. Feng, B. D. Froese, and R. Liang, "Composite method for precise freeform optical beam shaping," *Appl. Opt.*, vol. 54, pp. 9364–9369, Nov 2015.

[107] Z. Feng, D. Cheng, and Y. Wang, "Transferring freeform lens design into phase retrieval through intermediate irradiance transport," *Optics Letters*, vol. 44, p. 5501, 11 2019.

[108] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on Pure and Applied Mathematics*, vol. 44, no. 4, pp. 375–417, 1991.

[109] J. E. Marsden and A. Tromba, *Vector calculus [6th ed.]*. New York: W.H. Freeman, 6th ed. int. ed. ed., 2012.

[110] M. Sulman, J. Williams, M. F. Beg, and R. Russell, "Volumetric image registration methods based on solving the Monge-Ampère equation," *The Canadian Applied Mathematics Quarterly*, vol. 17, 01 2009.

[111] R. Courant and D. Hilbert, *Methods of Mathematical Physics: Partial Differential Equations*. Wiley Classics Library, Wiley, 2008.

[112] L. Li, "Use of fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A*, vol. 13, pp. 1870–1876, Sep 1996.

[113] L. Li, "New formulation of the fourier modal method for crossed surface-relief gratings," *J. Opt. Soc. Am. A*, vol. 14, pp. 2758–2767, Oct 1997.

[114] E. Popov and M. Nevière, "Maxwell equations in fourier space: fast-converging formulation for diffraction by arbitrary shaped, periodic, anisotropic media," *J. Opt. Soc. Am. A*, vol. 18, pp. 2886–2894, Nov 2001.

[115] S. Zhang, D. Asoubar, C. Hellmann, and F. Wyrowski, "Propagation of electromagnetic fields between non-parallel planes: a fully vectorial formulation and an efficient implementation," *Appl. Opt.*, vol. 55, pp. 529–538, Jan 2016.

[116] A. V. Pfeil, F. Wyrowski, A. Drauschke, and H. Aagedal, "Analysis of optical elements with the local plane-interface approximation," *Appl. Opt.*, vol. 39, pp. 3304–3313, Jul 2000.

[117] M. R. Foreman and P. Török, "Computational methods in vectorial imaging," *Journal of Modern Optics*, vol. 58, no. 5-6, pp. 339–364, 2011.

[118] Z. Feng, D. Cheng, and Y. Wang, "Iterative wavefront tailoring to simplify freeform optical design for prescribed irradiance," *Optics Letters*, vol. 44, p. 2274, 05 2019.

Acknowledgements

Throughout the work of this dissertation I have received a great deal of support and assistance. I would like to acknowledge the following people, without whom I would not have been able to complete this research.

I would first like to thank my supervisor, Prof. Dr. Frank Wyrowski, whose expertise was invaluable in formulating the research questions and methodology. I like him for accepting me as a Phd student working in his research group. Frank is a passionate mentor who always encourages me to conduct my research with my own thinking. However, through my research, his insight and knowledge into the subject also help me a lot. I always remember while in the discussion with him, our opinions, sometimes even conflicting ideas, were sharing and valuable conclusions were made for boosting the research work. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I also like to thank all my colleagues, who are also my good friends, in the research group of "Applied Computational Optics". Working together with these talented and lovely colleagues not only helps me in the research work, but also enriches my PhD journey with a delightful memory. I enjoy the working atmosphere in this group that we always support and enlighten each other with our typical character.

Next, I like to thank all the other colleagues in LightTrans International GmbH and Wyrowski Photonics UG. The research work of this dissertation is conducted with the software VirtualLab Fusion. I thank all the optical engineers in LightTrans International GmbH and Wyrowski Photonics UG that offer me the background knowledge of the software. Their technical support for the software help me accelerating my research and gaining more knowledge besides physics, especially the programming skill.

Last but not least, I like to thanks my dear parents and my family for their patience and love of me. It was their love helped me went through some difficult time in my PhD journey.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Die aus anderen Quellen direkt order indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. Herr Prof. Dr. Frank Wyrowski als Betreuer dieser Arbeit.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungsbzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen.

Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten ergalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die geltende Promotionsordnung der Physikalisch-Astronomishcen Fakultät ist mir bekannt.

Ich versichere ehrenwörtlich, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

_____ _____

Ort, Datum Unterschrift