Computation of 2D Fourier transforms and diffraction integrals using Gaussian radial basis functions

A. Martínez-Finkelshtein, D. Ramos-López and D.R. Iskander

Applied and Computational Harmonic Analysis

2016

Computation of 2D Fourier transforms and diffraction integrals using Gaussian radial basis functions

A. Martínez-Finkelshtein, D. Ramos-López and D.R. Iskander

A. Martínez-Finkelshtein, D. Ramos-López, D.R. Iskander. *Computation of 2D Fourier transforms and diffraction integrals using Gaussian radial basis functions*. Applied and Computational Harmonic Analysis, ISSN 1063-5203 (2016).

http://dx.doi.org/10.1016/j.acha.2016.01.007

http://www.sciencedirect.com/science/article/pii/S1063520316000087

Computation of 2D Fourier transforms and diffraction integrals using Gaussian radial basis functions

A. Martínez–Finkelshtein[a,b,*], D. Ramos-López[a], D. R. Iskander[c]

[a]*Department of Mathematics, University of Almería, Spain*
[b]*Instituto Carlos I de Física Teórica y Computacional, Granada University, Spain*
[c]*Department of Biomedical Engineering, Wroclaw University of Technology, Wroclaw, Poland*

Abstract

We implement an efficient method of computation of two dimensional Fourier-type integrals based on approximation of the integrand by Gaussian radial basis functions, which constitute a standard tool in approximation theory. As a result, we obtain a rapidly converging series expansion for the integrals, allowing for their accurate calculation. We apply this idea to the evaluation of diffraction integrals, used for the computation of the through-focus characteristics of an optical system. We implement this method and compare it performance in terms of complexity, accuracy and execution time with several alternative approaches, especially with the extended Nijboer-Zernike theory, which is also outlined in the text for the reader's convenience. The proposed method yields a reliable and fast scheme for simultaneous evaluation of such kind of integrals for several values of the defocus parameter, as required in the characterization of the through-focus optics.

Keywords: 2D Fourier transform, Diffraction integrals, Radial Basis Functions, Extended Nijboer–Zernike theory, Through-focus characteristics of an optical system

*Corresponding author.
Email addresses: andrei@ual.es (A. Martínez–Finkelshtein), drl012@ual.es (D. Ramos-López), robert.iskander@pwr.edu.pl (D. R. Iskander)

1. Introduction

The importance of the 2D Fourier transform in mathematical imaging and vision is difficult to overestimate. For a function g given on $\mathbb{R}^2$ in polar coordinates (ρ,θ) it reduces to calculating the integrals of the form

$$F(r,\phi)=\frac{1}{\pi}\int_0^{\infty}\int_0^{2\pi} g(\rho,\theta)e^{2\pi i r\rho\cos(\theta-\phi)}\rho\, d\rho\, d\theta, \tag{1}$$

in which the Fourier transform is expressed in polar coordinates (r,ϕ). For instance, the impulse response of an optical system (referred to as the point-spread function (PSF)) can be defined in terms of diffraction integrals of the form (1). The PSF uniquely defines a linear optical system and it is usually calculated for a single value of the focus parameter [1].

Calculating through-focus characteristics of an optical system has a wide variety of important applications including phase-diverse phase retrieval [2, 3], wavefront sensing [4], aberration retrieval in lithography, microscopy, and extreme ultraviolet light optics [5, 6], as well as in physiological optics, where such calculation can been used to assess the efficacy of intraocular lenses [7, 8, 9], study the depth-of-focus of the human eye [10, 11] or determine optimal pupil size in retinal imaging instruments such as the confocal scanning laser ophthalmoscope [12].

Another rapidly developing field in imaging is the digital holography [13, 14, 15, 16], which finds applications in the quantitative visualization of phase objects such as living cells using microscopic objectives in a digital holographic set-up. This technology of acquiring and processing holographic measurement data usually is comprised of two steps, the recording of an interference pattern produced by a real object on a CCD, followed by the numerical reconstruction of the hologram by simulating the back-propagation of the image. Fourier-type integrals are an essential part of several algorithms used in digital holography, for instance when calculating the Fresnel Transforms.

In this context, the numerical algorithms used to calculate the 2D Fourier transform play a central role. Unfortunately, such integrals can be explicitly evaluated in terms of standard special functions only in a limited number of situations. On the other hand, purely numerical procedures for these evaluations are computationally cumbersome and require substantial computational resources. Thus, the most popular alternative are the semi-analytic methods that use an approximation of the integrand with functions from a certain class, for which some forms of closed expressions are available. This is the paradigm of the extended Nijboer–Zernike (ENZ) approach [17, 18], explained briefly in Section 4. The ENZ theory constitutes an important step forward in comparison to the direct quadrature integration. However, it has some important limitations acknowledged by its creators, such as the restriction to circular pupils and symmetric aberrations containing only the even terms of Zernike polynomial expansion (i.e., only cosine dependence), or its poor performance for large numerical apertures.

In order to overcome these limitations we have developed a technique for calculating integrals of the form (1) by combining the essence of the ENZ approach with the approximation power of the radial basis function (RBF), which are standard tools in the approximation theory and numerical analysis, appearing in applications ranging from engineering to solutions of partial derivative equations, see for example [21]. The Gaussian radial basis functions (GRBF) have been also used in the context of ophthalmic optics [22, 25, 26].

The structure of the paper is as follows. Section 2 is devoted to the derivation of the main formulas that can be used for the calculation of general Fourier integrals (1). The detailed algorithm is developed in Section 3 for the diffraction integrals defined therein, including its efficient implementation and error estimates. An assessment of accuracy and efficiency of this procedure is carried out in Section 5, where it is compared also with the extended Nijboer-Zernike theory (outlined in

Section 4) and with the standard 2D fast Fourier transform.

Some of the ideas underlying this approach have been presented in [27]. In this work we develop further the algorithmic aspects of this procedure, discuss its efficient implementation, and compare its performance with some standard alternatives.

2. Fourier integrals for Gaussian RBF

Consider the case when g is a Gaussian RBF with the shape parameter $\lambda > 0$ and center at a point with cartesian coordinates (a,b), multiplied by a radially-symmetric complex-valued function h; in other words, let

$$g(\rho,\theta) = h(\rho)e^{-\lambda((x-a)^2+(y-b)^2)}, \quad x = \rho\cos(\theta), \quad y = \rho\sin(\theta).$$

Using the polar coordinates for the center,

$$a = q\cos(\alpha), \quad b = q\sin(\alpha),$$

we can rewrite it as

$$g(\rho,\theta) = h(\rho)e^{-\lambda(q^2+\rho^2-2\rho q\cos(\theta-\alpha))}. \tag{2}$$

In this section we discuss an expression for (1) with g as in (2):

$$\begin{aligned} F(r,\phi;q,\alpha) =& \frac{1}{\pi}\int_0^\infty \int_0^{2\pi} e^{-\lambda(q^2+\rho^2-2\rho q\cos(\theta-\alpha))} e^{2\pi i r\rho\cos(\theta-\phi)} h(\rho)\rho\, d\rho\, d\theta \\ =& \frac{1}{\pi} e^{-\lambda q^2} \int_0^\infty d\rho\, h(\rho)\rho e^{-\lambda\rho^2} \int_0^{2\pi} e^{2\lambda\rho q\cos(\theta-\alpha)} e^{2\pi i r\rho\cos(\theta-\phi)} d\theta. \end{aligned} \tag{3}$$

For arbitrary constants $A, B \in \mathbb{C}$ we have the following identity:

$$\int_0^{2\pi} e^{2A\cos(\theta-\alpha)} e^{2B\cos(\theta)} d\theta = 2\pi I_0\left(2\sqrt{A^2+2AB\cos\alpha+B^2}\right), \tag{4}$$

where I_0 is the modified Bessel function of the first kind (see e.g. [33, Chapter 9]). Observe that I_0 is an entire even function, so that the value in the right hand side of (4) is independent of the choice of the branch of the square root.

Identity (4) is a straightforward consequence of the fact that

$$A\cos(\theta-\alpha)+B\cos(\theta)=\sqrt{A^2+2AB\cos\alpha+B^2}\cos(\theta-\gamma),$$

where

$$\gamma=\arctan\left(\frac{A\sin\alpha}{B+A\sin\alpha}\right),$$

(with an appropriate choice of the branch of the square root), along with the identity ([33], formula (9.6.16)):

$$I_0(z)=\frac{1}{\pi}\int_0^\pi e^{\pm z\cos\theta}d\theta.$$

Using (4) in (3) we conclude that

$$\begin{aligned}F(r,\phi;q,\alpha)&=2e^{-\lambda q^2}\int_0^\infty e^{-\lambda\rho^2}I_0\left(2\rho\sqrt{\Omega}\right)h(\rho)\rho\,d\rho\\&=e^{-\lambda q^2}\int_0^\infty e^{-\lambda\rho}I_0\left(2\sqrt{\rho\Omega}\right)h(\sqrt{\rho})\,d\rho\\&=e^{-\lambda q^2}\mathscr{L}\left[I_0\left(2\sqrt{\cdot\,\Omega}\right)h(\sqrt{\cdot})\right](\lambda),\end{aligned}\tag{5}$$

where

$$\Omega=\Omega(r,\phi;q,\alpha)=\lambda^2q^2+2\pi ir\lambda q\cos(\phi-\alpha)-\pi^2r^2,\tag{6}$$

and $\mathscr{L}[\cdot]$ denotes the Laplace transform. Formulas (5)–(6) are the basic building blocks for the algorithm proposed below.

Let us particularize these identities to the case of a circular exit pupil of an optical system, when h can be taken as the characteristic function of the interval $[0,1]$,

$$h(\rho)=\chi_{[0,1]}(\rho)=\begin{cases}1, & \text{if } 0\le\rho\le1,\\ 0, & \text{otherwise.}\end{cases}$$

We see that the crucial step is the evaluation of the integral

$$\int_0^1 e^{-\lambda\rho} I_0\left(2\sqrt{\rho\Omega}\right) d\rho.$$

The Taylor series for the Bessel function (see [33], formula (9.6.12)),

$$I_0\left(2\sqrt{\rho\Omega}\right) = \sum_{s=0}^{\infty} \frac{\Omega^s}{(s!)^2}\rho^s,$$

is locally uniformly convergent on the whole plane, and thus

$$\int_0^1 e^{-\lambda\rho} I_0\left(2\sqrt{\rho\Omega}\right) d\rho = \sum_{s=0}^{\infty} \frac{(m_s(\lambda)\Omega)^s}{(s!)^2}, \tag{7}$$

where

$$m_s(\lambda) = \begin{cases} \displaystyle\int_0^1 e^{-\lambda\rho}\rho^s d\rho, & \lambda \neq 0, \\ (1+s)^{-1}, & \lambda = 0. \end{cases} \tag{8}$$

Integration by parts and straightforward calculations show that for $s = 0, 1, 2, \ldots$ and $\lambda \neq 0$,

$$m_0(\lambda) = \frac{1-e^{-\lambda}}{\lambda}, \quad m_{s+1}(\lambda) = \frac{(s+1)m_s(\lambda) - e^{-\lambda}}{\lambda}. \tag{9}$$

Notice also that

$$m_0(\lambda) = \mathscr{L}[\chi_{[0,1]}](\lambda), \quad m_s(\lambda) = \left(-\frac{d}{d\lambda}\right)^{s-1} m_0(\lambda), \quad s \geq 1. \tag{10}$$

Clearly, these formulas can be easily extended to the case when g is a linear combination of functions of the form (2), that is,

$$g(\rho,\theta) = h(\rho)\sum_{k=1}^{K} c_k e^{-\lambda_k(q_k^2+\rho^2-2\rho q_k\cos(\theta-\alpha_k))}, \quad h(\rho) = \chi_{[0,1]}(\rho),$$

where $c_k \in \mathbb{C}$ are certain coefficients, $\lambda_k > 0$ are the shape parameters, and (q_k, α_k) are the polar coordinates of the corresponding centers of the Gaussian RBFs. Formulas (5)–(7) show that the corresponding Fourier transform F, defined in (1), can be written as

$$F(r,\phi) = \sum_{k=1}^{K} c_k e^{-\lambda_k q_k^2} \sum_{s=0}^{\infty} \frac{m_s(\lambda_k)}{(s!)^2}\,\Omega(r,\phi;q_k,\alpha_k)^s,$$

with Ω defined in (6).

If we assume additionally that all the shape parameters are equal, $\lambda_1 = \cdots = \lambda_K = \lambda$, we can reorganize the calculations as follows:

$$F(r,\phi) = \sum_{s=0}^{\infty} \frac{m_s(\lambda)}{(s!)^2} H_s(r,\phi), \quad H_s(r,\phi) = \sum_{k=1}^{K} c_k e^{-\lambda q_k^2} \Omega(r,\phi;q_k,\alpha_k)^s.$$

3. Calculation of diffraction integrals

Optical systems can be described by means of the complex-valued pupil function $P(\rho,\theta)$ that, in the case of a circular pupil, can be expressed in (normalized) polar coordinates as

$$P(\rho,\theta) = A(\rho,\theta)\exp\left(-i\frac{2\pi n}{\lambda}W(\rho,\theta)\right), \tag{11}$$

where $A(\rho,\theta)$ is the aperture or amplitude transmittance function of the optical system, $W(\rho,\theta)$ is the wavefront error, and constants n and λ denoting the refractive index and the wavelength of the light, respectively. According to Fourier optics [20], the complex-valued point-spread function of such a system is given by the diffraction integral

$$U(r,\phi;f) = \frac{1}{\pi}\int_0^1 \int_0^{2\pi} \exp(if\rho^2)P(\rho,\theta)\exp\left(2\pi i\rho r\cos(\theta-\phi)\right)\rho d\theta d\rho, \tag{12}$$

where f represents the defocusing ($f = \pi/2$ corresponds to one focal depth), and (r,ϕ) denote the polar coordinates in the image plane. The corresponding optical impulse response or point-spread function of the system can be evaluated by

$$PSF(r,\phi;f) = |U(r,\phi;f)|^2.$$

In this section we discuss the computational framework and the efficient implementation of the algorithm for calculation of such diffraction integrals. Observe that formally the defocusing term ($\exp(if\rho^2)$) in (12) could be absorbed in the

aberration term. However, from the physical point of view, the aberration term is an intrinsic error of the optical system, while the defocusing is a deliberately introduced defect, which can take on several values that are relatively large with respect to the aberrations of the system. For this reason, a separate treatment is commonly preferred, see [18].

3.1. Computational framework

In practice, the wavefront is sampled at a discrete and finite set of points on the pupil (the unit disk). In other words, the input data set has the form (x_j, y_j, W_j), with $W_j = W(x_j, y_j)$. Values W_j can be measured directly or obtained by standard procedures.

All semi-analytic methods for calculation of Fourier (1) or diffraction integrals (12) are based on an approximation of the integrand by a function of a suitable chosen form. We also start our calculation of (12) by approximating the complex pupil function P in (11) by a linear combination of Gaussian radial basis functions (GRBF), which yields an expression of the form

$$P(\rho,\theta) = \sum_{k=1}^{K} c_k g_k(\rho,\theta), \quad g_k(\rho,\theta) = e^{-\lambda_k(q_k^2+\rho^2-2\rho q_k\cos(\theta-\alpha_k))}, \tag{13}$$

where $c_k \in \mathbb{C}$ are certain coefficients, $\lambda_k > 0$ are the shape parameters, and (q_k, α_k) are the polar coordinates of the centers of the Gaussian RBFs. In order to obtain (13) we first fix a basis $\{g_k\}_{k=1}^N$ of GRBF, or equivalently, choose, for each index k, values for q_k, α_k and λ_k. The optimal choice of the shape parameter based on the given data is a highly non-linear problem, usually solved by cross validation, see [23, 24], but the following rule of thumb has been tested in practice: take the same shape parameter for all k's,

$$\lambda_1 = \cdots = \lambda_K = \lambda \in [1, 20], \tag{14}$$

select $r \in \mathbb{N}$ and create a grid of $r \times r$ equally spaced points in the square $[-1.2, 1.2]^2$. The goal of covering an area larger than the unit disk is to deal with the Gibbs phenomenon at the boundary of the disk. The polar coordinates of these $K = r^2$ points will be the pairs (q_k, α_k) in (13).

With the basis $\{g_k\}$ chosen, we calculate the coefficients c_k fitting the pupil function by means of the linear least squares. For a high number of centers K this problem is ill-conditioned, and the use of Tikhonov regularization [29] is recommended, with an appropriate choice of regularization parameter, for instance, by the L-curve method, as described for example in [30].

For an alternative approach to the 2D nonlinear approximation of functions by sums of exponentials see [31].

With all the parameters in the representation (13) calculated we can use the formulas obtained in Section 2, namely

$$U(r,\phi;f) = \sum_{k=1}^{K} c_k\, e^{-\lambda_k q_k^2} \sum_{s=0}^{\infty} \frac{m_s(\lambda_k - if)}{(s!)^2}\, \Omega(r,\phi,q_k,\alpha_k)^s, \tag{15}$$

with

$$\Omega(r,\phi;q,\alpha) = \lambda^2 q^2 + 2\pi i r \lambda q \cos(\phi - \alpha) - \pi^2 r^2.$$

Furthermore, with the additional assumption (14) we can rewrite this formula as

$$U(r,\phi;f) = \sum_{s=0}^{\infty} \frac{m_s(\lambda_k - if)}{(s!)^2}\, H_s(r,\phi), \quad H_s(r,\phi) = \sum_{k=1}^{K} c_k\, e^{-\lambda q_k^2}\, \Omega(r,\phi,q_k,\alpha_k)^s. \tag{16}$$

3.2. Efficient implementation of the calculations

The recurrence relation (9) can present some numerical instability, especially for large value of the imaginary part of the argument; these considerations suggest to replace the coefficients $m_s(\lambda)$, defined in (8), by their normalized counterparts,

$$\widehat{m}_s(\lambda) = \frac{m_s(\lambda)}{(s!)^2},$$

that can be efficiently generated for $\lambda \neq 0$ by the following double recurrence:

$$\widehat{m}_{s+1}(\lambda) = \frac{1}{\lambda}\left[\frac{\widehat{m}_s(\lambda)}{s+1} - \tau_s(\lambda)\right], \quad \tau_{s+1}(\lambda) = \frac{\tau_s(\lambda)}{(s+2)^2}, \quad s = 0,1,2,\ldots \tag{17}$$

with the initial conditions

$$\tau_0(\lambda) = e^{-\lambda}, \quad \widehat{m}_0(\lambda) = \frac{1-\tau_0(\lambda)}{\lambda}. \tag{18}$$

Since formulas in (15)–(16) contain an infinite sum, we must choose a *cut-off parameter*, $S \in \mathbb{N}$, so that the expression in (15) is replaced by

$$U(r,\phi;f) = \sum_{k=1}^{K} c_k\, e^{-\lambda_k q_k^2} \sum_{s=0}^{S} \widehat{m}_s(\lambda_k - if)\, \Omega(r,\phi,q_k,\alpha_k)^s, \tag{19}$$

while (16) becomes

$$U(r,\phi;f) = \sum_{s=0}^{S} \widehat{m}_s(\lambda_k - if)\, H_s(r,\phi), \tag{20}$$

where we denote

$$H_s(r,\phi) = \sum_{k=1}^{K} c_k\, e^{-\lambda q_k^2}\, \Omega(r,\phi,q_k,\alpha_k)^s.$$

It is worth observing that in these formulas the defocus parameter f and the coordinates r and ϕ are independent: f does not appear in Ω, while $m_s(\lambda_k - if)$ need to be calculated only once regardless the number of points at which we evaluate the functions.

In practice, we want to find the values of $U(r,\phi;f)$ at a vector of J points on the plane, given by their polar coordinates (ρ,ϕ), and for a vector of defocus parameters f. In other words, we need to compute efficiently the matrix

$$\boldsymbol{U} = \left(U(r_j,\phi_j;f_m)\right)_{m,j} \in \mathbb{C}^{M\times J}.$$

Let us discuss the algorithm under the assumption (14), so we will use formula (16).

Applying (17)–(18) to the vector $\boldsymbol{f}$, we obtain the matrix

$$\boldsymbol{M} = \left(\widehat{m}_s(\lambda - if_m)\right)_{m=1,\ldots,M}^{s=0,1,\ldots,S} \in \mathbb{C}^{M\times(S+1)},$$

which, as it was mentioned, is computed once for all (r_j, ϕ_j)'s.

Another observation that will speed up computations is that if P and Q are $\mathbb{R}^{2\times 1}$ vectors of Cartesian coordinates of points on $\mathbb{R}^2$ with polar coordinates (r,ϕ) and (q,α), respectively, then by (6),

$$\Omega(r,\phi;q,\alpha) = \lambda^2 q^2 + 2\pi i r\lambda q\cos(\phi-\alpha) - \pi^2 r^2 = (\lambda Q + \pi i P)^T(\lambda Q + \pi i P), \tag{21}$$

where the superscript T denotes the transpose of a matrix. In other words, $\Omega(r,\phi;q,\alpha)$ is the square of the euclidean distance (in $\mathbb{R}^2$) from λQ to $-\pi i P$. It allows us to encode efficiently the evaluation of $\mathbf{\Omega} = (\Omega_{k,j}) \in \mathbb{R}^{K\times J}$,

$$\Omega_{k,j} = \Omega(r_j, \phi_j, q_k, \alpha_k),$$

by first finding the Cartesian coordinates of the centers,

$$a_k = q_k\cos\alpha_k, \quad b_k = q_k\sin\alpha_k, \quad k = 1,\dots,K.$$

Finally, we find the column vector

$$\boldsymbol{d} = (c_k\, e^{-\lambda q_k^2})_{k=1,\dots,K} \in \mathbb{R}^{K\times 1}. \tag{22}$$

We describe the rest of the procedure in the Algorithm 1.

It should be noticed finally that we can slightly increase the accuracy adding a constant term to the approximation of the complex pupil function, that is, instead of (13) using the representation

$$P(\rho,\theta) = c_0 + \sum_{k=1}^{K} c_k g_k(\rho,\theta), \quad g_k(\rho,\theta) = e^{-\lambda_k(q_k^2+\rho^2-2\rho q_k\cos(\theta-\alpha_k))},$$

where c_0 can be taken as the average of the values of P at the given points. Notice that the contribution to U corresponding to this term can be computed using Algorithm 1 with $\lambda = 0$.

input : Matrices $\boldsymbol{M} \in \mathbb{C}^{M\times(S+1)}$, $\boldsymbol{\Omega} = (\Omega_{k,j}) \in \mathbb{R}^{K\times J}$, and $\boldsymbol{d} \in \mathbb{R}^{K\times 1}$

output: Matrix $\boldsymbol{U} \in \mathbb{C}^{M\times J}$

Initialization:

Set $\boldsymbol{U} = \boldsymbol{0} \in \mathbb{R}^{M\times J}$

Define $\boldsymbol{R} \in \mathbb{R}^{K\times J}$ as the matrix of 1's, and compute

$$\boldsymbol{H} = \boldsymbol{d}^T \boldsymbol{R}. \tag{23}$$

`// At this stage,` $\boldsymbol{H} = H_0(r_j, \phi_j)_{j=1,\ldots,J^2} \in \mathbb{C}^{1\times J}$

for $s \leftarrow 0$ **to** $S-1$ **do**

Calculate

$$\boldsymbol{U} \leftarrow \boldsymbol{U} + \boldsymbol{M}(:, s+1)\boldsymbol{H},$$

Update $\boldsymbol{R}$ by

$$\boldsymbol{R} \leftarrow \boldsymbol{R}. * \boldsymbol{\Omega}$$

`// We use the Matlab notation (.*) for the term-by-term multiplication;`

Compute $\boldsymbol{H}$ using (23);

end

Algorithm 1: Implementation of formula (16).

3.3. Error estimates and convergence analysis

The algorithm described above presents two main sources of errors (besides the standard truncation and machine arithmetic errors that we do not discuss here):

- the fitting error in (13);
- the truncation error, given by the choice of S, in replacing the infinite series in (15) or (16) by finite sums.

The former is difficult to estimate and constitutes an active field of research within the radial basis functions community, see e.g. [21]. Concerning the truncation error, observe that from the definition (8) it follows that $|m_s(\lambda)| \leq 1$ for $\mathrm{Re}(\lambda) \geq 0$. Indeed, for purely imaginary λ this is trivial, while for $\mathrm{Re}(\lambda) > 0$,

$$|m_s(\lambda)| \leq |m_s(\mathrm{Re}(\lambda))| \leq \int_0^1 e^{-\mathrm{Re}(\lambda)\rho} d\rho = |m_0(\mathrm{Re}(\lambda))| = \frac{1 - e^{-\mathrm{Re}(\lambda)}}{\mathrm{Re}(\lambda)} \leq 1.$$

In particular, for $\lambda > 0$,

$$|m_s(\lambda - if)| \leq 1.$$

On the other hand, by expression (6),

$$|\Omega| = |\Omega(r, \phi; q, \alpha)| = \sqrt{\left(\lambda^2 q^2 - \pi^2 r^2\right)^2 + \left(2\pi r \lambda q \cos(\phi - \alpha)\right)^2} = \lambda^2 q^2 + \pi^2 r^2.$$

According to our algorithm and (14), $q \leq 1.7$, $r \leq 2$, $\lambda \leq 20$, so that

$$|\Omega| \leq 1200, \tag{24}$$

where the upper bound is not tight.

In consequence, the truncation error in (15) is given by

$$\left| \sum_{s=S+1}^{\infty} \frac{m_s(\lambda_k - if)}{(s!)^2} \Omega(r, \phi, q_k, \alpha_k)^s \right| \leq \sum_{s=S+1}^{\infty} \frac{1}{(s!)^2} \left(\lambda_k^2 q_k^2 + \pi^2 r^2\right)^s,$$

or in other words, the remainder of the Taylor expansion for the modified Bessel function

$$I_0\left(2\sqrt{\lambda_k^2 q_k^2+\pi^2 r^2}\right),$$

and thus,

$$\left|\sum_{s=S+1}^{\infty}\frac{m_s(\lambda_k-if)}{(s!)^2}\,\Omega(r,\phi,q_k,\alpha_k)^s\right|\leq\frac{\left(\lambda_k^2 q_k^2+\pi^2 r^2\right)^{S+1}}{(S+1)!}\max_{0\leq t\leq\lambda_k^2 q_k^2+\pi^2 r^2}\left|I_0^{(S+1)}(t)\right|.$$

The bound (24) above is very conservative, and shows that in the worst case scenario $S=100$ provides a truncation error of order 10^{-9}.

Although, as the numerical experiments explained in Section 5 show, the method performs well even for reasonably large values of r and f in (12), we can slightly improve accuracy by using the scheme (15), and replacing the infinite series by its Padé approximant of a suitable order with center at the origin (instead of a mere truncation, as we did above), see e.g. [32].

4. Outline of the extended Nijboer-Zernike theory

In Section 5 we are going to produce the results of some numerical experiments regarding the behavior of the algorithm described above, and to compare them with the standard procedures used for the computation of diffraction integrals.

One of the best-known semi-analytic methods of calculation of these integrals is the so-called Extended Nijboer–Zernike (ENZ) theory [17, 18], similar in spirit to the method presented in this paper, with the main difference that the pupil function P is expanded in terms of Zernike polynomials, instead of Gaussian radial basis functions:

$$P(\rho,\theta)=\sum_{n,m}c_n^m Z_n^m(\rho,\theta) \tag{25}$$

(see e.g. [20, Section 9.2] for the definition of Zernike polynomials).

For the reader's convenience, in this section we present an outline of the ENZ formulas.

If the coefficients c_n^m in the expression (25) are known, then the integral U in (12) takes the form

$$U(r,\phi;f) = \sum_{n,m} c_n^m U_n^m(r,\phi;f), \tag{26}$$

where

$$U_n^m(r,\phi;f) = \frac{1}{\pi}\int_0^1 \int_0^{2\pi} \left[\exp(if\rho^2) Z_n^m(\rho,\theta) \exp\left((2\pi i \rho r \cos(\theta-\phi)\right)\right] \rho\, d\theta\, d\rho.$$

According to ENZ theory, for the double-index (n,m) with $m \geq 0$ (a necessary condition for the applicability of these formulas) we have

$$U_n^m(r,\phi;f) = 2i^m V_n^m(r,f)\cos(m\phi), \tag{27}$$

with

$$V_n^m(r,f) = \exp(if) \sum_{k=0}^{\infty} \left(\frac{-if}{\pi r}\right)^k B_k(r) \tag{28}$$

and

$$B_k(r) = \sum_{j=0}^{p} u_{kj} \frac{J_{m+k+2j+1}(2\pi r)}{2\pi r}, \tag{29}$$

with the coefficients

$$u_{kj} = (-1)^p \frac{m+k+2j+1}{q+k+j+1} \binom{m+k+j}{k} \binom{k+j}{k} \binom{k}{p-j} \Big/ \binom{q+k+j}{k} \tag{30}$$

where $p = (n-m)/2$ and $q = (n+m)/2$. Here J_ν is the Bessel function of the first kind and order ν, see e.g. [33, Chapter 9].

Observe that these formulas contain a removable singularity at $r = 0$, so some extra care when organizing the calculations should be put. Moreover, they also present difficulties when the defocus parameter f is large. According to [18], approximately $3|f|$ terms are needed in general to accurately evaluate expression (28), and the practical use of that formula is limited to a range $|f| \leq 5\pi$.

In order to overcome these difficulties an alternative approach has been put forward in [19]. The main idea is the use of Bauer's identity:

$$\exp(if\rho^2) = \exp\left(\frac{1}{2}if\right)\sum_{k=0}^{\infty}(2k+1)i^k j_k\left(\frac{1}{2}f\right)R_{2k}^0(\rho), \tag{31}$$

where R_n^m is the radial part of the Zernike polynomials, and

$$j_k(z) = \left(\frac{\pi}{2z}\right)^{1/2} J_{k+(1/2)}(z), \quad k = 0,1,\ldots \tag{32}$$

Using (25) and (31) in (12) we reduce the problem to the computation of integrals containing products of two radial Zernike polynomials, $R_{n_1}^{m_1}$ and $R_{n_2}^{m_2}$. The crucial idea is to use the linearization formulas for these products of the form

$$R_{m_1+2p_1}^{m_1} R_{m_2+2p_2}^{m_2} = \sum_l c_l R_{m_1+m_2+2l}^{m_1+m_2},$$

where coefficients c_l defined as

$$c_l = \sum_{s_1,s_2,t} f_{p_1,s_1}^{m_1} f_{p_2,s_2}^{m_2} g_{s_1+s_2-2t,l}^{m_1+m_2}$$

are expressed in terms of the quantities $f_{p,s}^m$ and $g_{u,l}^m$ given explicitly by

$$f_{p,s}^m = (-1)^{p-s}\frac{2s+1}{p+s+1}\left[\binom{m+p-s-1}{m-1}\binom{m+p+s}{s}\Big/\binom{p+s}{s}\right], \tag{33}$$

for $s = 0,\ldots,p$;

$$g_{u,l}^m = \frac{m+2l+1}{m+u+l+1}\left[\binom{m}{u-l}\binom{u+l}{l}\Big/\binom{m+l+u}{m+l}\right], \tag{34}$$

for $u = l,\ldots,l+m$; and

$$b_{s_1,s_2,t} = \frac{2s_1+2s_2-4t+1}{2s_1+2s_2-2t+1}\left(\frac{A_{s_1-t}A_tA_{s_2-t}}{A_{s_1+s_2-t}}\right), \tag{35}$$

for $t = 0,\ldots,\min\{s_1,s_2\}$ and $A_k = \binom{2k}{k}$. For $m = 0$, expressions (33) and (34) boil down to

$$f_{p,s}^0 = \delta_{p,s}, \quad g_{u,l}^0 = \delta u,l,$$

where δ is the Kronecker's delta.

As a result, we obtain the so-called Bessel-Bessel series expression for V_n^m in (28): for n, m nonnegative integers with $n-m \geq 0$ and even, with $p = \frac{1}{2}(n-m)$ and $q = \frac{1}{2}(n+m)$,

$$V_n^m(r,f) = \exp\left(\frac{1}{2}if\right) \sum_{k=0}^{\infty} (2k+1) i^k j_k\left(\frac{1}{2}f\right) \sum_{l=l_0}^{k+p} (-1)^l w_{k,l} \frac{J_{m+2l+1}(2\pi r)}{2\pi r}, \quad (36)$$

where $l_0 = \max\{0, k-q, p-k\}$ and

$$w_{k,l} = \sum_{s=0}^{p} \sum_{t=0}^{\min\{k,s\}} f_{p,s}^m b_{k,s,t} g_{k+s-2t,l}^m. \quad (37)$$

In the special case that $m = 0$ we have that

$$w_{k,k+p-2j} = b_{k,p,j}, \quad j = 0, 1, \ldots, \min\{k, p\},$$

while all other $w_{k,l}$ vanish. In this way, all $w_{k,l} \geq 0$ and $\sum_l w_{k,l} = 1$.

The conclusion of [19] is that this new scheme is valid and accurate even for very large values of f. By analyticity, the expression (36) remains valid also for complex values of f, which is important in a number of practical applications, with the downside in the removable singularity of (32) at $f = 0$. In practice, it is convenient to use both ENZ schemes if the defocus ranges from 0 to large values of f.

5. Comparative assessment of accuracy and efficiency

In this section we analyze the performance of the methods based on the Gaussian radial basis functions (GRBF) proposed in this paper is analyzed and compared to the popular alternative approaches.

Since the closed analytic expression for a Fourier transform type integrals like (1) and (12) is possible only for most elementary integrands, for their computation we must rely either on numerical or on semi-analytical methods (or analytical

approximations). The first group comprises quadrature-type numerical procedures in which the integration over a 2D domain is replaced by the calculation of a discrete sum based on evaluation of the integrand at a discrete set of points, with the particular challenge of integrating a highly oscillatory function. The most popular approach is the bi-dimensional discrete Fourier transform (we will refer to it as the FFT2-based approach), calculated via the Fast Fourier Transform algorithm. Its efficiency can be substantially enhanced using the fractional Fourier transform [34] or a "butterfly diagram" ideas [35], see also [36].

The best known representative of the second group is the Extended Nijboer–Zernike (ENZ) theory that was briefly exposed in Section 4.

In this section, these alternatives are discussed and compared with our method, paying special attention to their computational complexity, precision, accuracy and speed. We use the "naive" notion of complexity, understanding by this the number of real floating point operations (*flops*) needed to run the algorithm. Since the exact number of flops is in general difficult or not feasible to calculate, the leading term for large values of the parameters is used.

The assessments are performed under assumptions allowing the application of all the methods explained above. In other words, we will evaluate the diffraction integrals $U = U(r,\phi;f)$, defined in (12), in a grid of points (in polar coordinates) $(r,\phi) \in \mathbb{R}^{N\times N}$, and for a vector $\boldsymbol{f} \in \mathbb{R}^M$ of values of the defocus parameters f.

5.1. Some remarks about FFT vs semi-analytic methods

In the FFT2-based scheme, the value of $U(r,\phi;f)$ is computed by means of the bi-dimensional fast Fourier transform. The crucial step is the substitution of the double integral in (12) by a discrete sum. Notice that each new value of the defocus parameter f obliges to calculate the values of U completely, at a computationally high cost. Furthermore, the use of the FFT requires re-sampling the wavefront at

a regular Cartesian grid covering the pupil; for convenience, the length of the grid should be an integer power of 2 in each direction.

Another remarkable issue to be addressed when using the FFT2 scheme is the aliasing, a typical phenomenon that can appear due to discontinuities of the integrand. In order to prevent this, the pupil must be small in comparison with the sampled area (or in other words, we must extend the pupil to a larger region, setting the pupil function to zero in the complementary domain), resulting in a large area where $U(r,\phi;f)$ is negligible [36]. Thus, a big portion of the computational load of this scheme is useless, and in general the spatial resolution needed with this method will be much higher than that required for the semi-analytic methods like discussed here. However, FFT2 is the standard method used in commercial ray tracing packages, such as Zemax or Code V.

The advantage of the semi-analytic approaches, such as the ENZ theory or the method proposed here, is that they reduce the computation of U in (12) to evaluation of more or less complex explicit expressions in terms of some elementary or special functions. One of the benefits of having these formulas is a better control of the image domain being computed, increasing the precision. Another important advantage is the huge boost in performance gained when a parallelization of calculations is done for multiple values of the defocus parameter f.

5.2. Computational complexity

Let us analyze and compare the computational cost of evaluating the diffraction integral using the ENZ theory and by the GRBF scheme, for a grid of values $\boldsymbol{r} \in \mathbb{R}^N$, $\boldsymbol{\phi} \in \mathbb{R}^N$ (so that with the notation of Section 3.2, $J = N^2$), and for M distinct defocuses, gathered in a vector $\boldsymbol{f} \in \mathbb{R}^M$.

ENZ method

We start with the basic ENZ method, corresponding to formulas (26)-(30). Observe that in (29) we must compute values of the Bessel functions of the first kind J_ν, usually by their power series

$$J_\nu(z) = \left(\tfrac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} (-1)^k \frac{(\frac{1}{4}z^2)^k}{k!\,\Gamma(\nu+k+1)} \tag{38}$$

(see e.g. [formula (10.2.2)][28]). In practice this series must be truncated at some term $B-1$. For the values of variables and parameters usually appearing in this context, we have determined experimentally that we can ensure the truncation error of 10^{-8} taking $B = 15$. The computational complexity of computing the Gamma function at a value of $\mathscr{O}(10^t)$ is $\mathscr{O}\left(t^2\log(t)^2\right)$. With $B = 15$ and with the usual Zernike fit (up to the 8th order) it is sufficient to take $t \leq 3$, which yields a maximum of 11 operations for the Gamma function evaluation. Thus, for N different values of r we need roughly $\mathscr{O}(3N+25)$ operation to to evaluate a single term in the series (38), and in consequence, the computational cost of the evaluating a single Bessel function is $\mathscr{O}(3BN)$.

With the same assumptions, the evaluation of each coefficient in (30) is at most $\mathscr{O}(150)$ flops. Consequently, the complexity of computation of each B_k defined in (29) is

$$\mathscr{O}((p+1)(30B+186))$$

operations.

In order to evaluate expression (28), we truncate the infinite series at some term $S-1$, in which case the number of operations required to evaluate V_n^m by (28) is

$$\mathscr{O}(S((p+1)(30B+186)+2)+S+8) = \mathscr{O}(30BS(p+1)).$$

Propagating this estimate to (27), computed also at N different values of ϕ, we need about $\mathscr{O}\left(30BS(p+1)+2N^2M+2N\right)$ flops for its evaluation.

In order to estimate the overall cost for expression (26) we have to take into account the indices n and m therein: n ranges from 0 to a certain n^* (the maximum radial order), and for each fixed n, m takes even/odd non-negative integer values $\leq n$. Thus, at each level n, only $\mathscr{O}(n/2)$ Zernike polynomials are used (recall the restriction of $m \geq 0$ for ENZ formulas). Then, $p = \frac{1}{2}(n-m) = \mathscr{O}(n/4)$ in average.

In addition, the maximum radial order n^* yields a total of $\frac{1}{2}n^*(n^*+1)$ Zernike polynomials, from which only those with the cosine are used, so that the effective number of polynomials is approximately $K = \frac{1}{4}n^*(n^*+1)$.

Combining all the previous assumptions, the complexity of computing U at a grid of $N \times N \times M$ points for (r, ϕ, f) by the ENZ formulas can be estimated to be

$$cost(U_{ENZ}) = \mathscr{O}\left(K^{3/2}BNS + K\left(N^2M + NMS + BNS\right)\right),$$

where B and S are the truncation values for the series in (38) and (28), respectively, and K is the total number of Zernike polynomials included in the pupil representation (25).

Improved ENZ method

We can perform a similar analysis for the improved ENZ scheme, corresponding to equations (33)-(37). The only difference with respect to the previous discussion lies in the computation of V_n^m, so we only need to re-estimate the computational cost of V_n^m in equation (36).

With the same assumptions used before, the evaluation cost of coefficients in equations (33), (34), (35) are constant and of at most of 110 flops. Thus, the complexity of coefficients $w_{k,l}$ in formula (37) is of $\mathscr{O}(330pk)$ flops. Recalling the estimate of $\mathscr{O}(3BN)$ flops for each Bessel function J_ν, the cost of the finite sum (that with index l) in expression (36) is $\mathscr{O}((k+p)(3BN+330pk+3N))$.

The cost of evaluation of the Bessel spherical function j_k is $\mathscr{O}(3BM+3M) = \mathscr{O}(3BM)$. Thus, if the improved formula for V_n^m is used with the truncation of the

infinite series at the term $S-1$, its complexity is

$$\mathscr{O}\left(S\left(3BM+BN+(S/2+p)(3BN+165pS)\right)+2M\right).$$

Assuming as before that $p=\frac{1}{2}(n-m)=\mathscr{O}(n/4)$ in average, it can be simplified to

$$\mathscr{O}\left(nS^2\left(21S+10n\right)+1.5BNS^2+\frac{3}{4}nSBN+NMS+BMS\right).$$

Since the rest of the calculations is the same as in the basic ENZ scheme, the complexity of evaluating U in a grid of $N\times N\times M$ points (r,ϕ,f) by the improved ENZ formulas (33)-(37) is

$$\begin{aligned} cost(U_{ENZ\text{ improved}}) = \mathscr{O}\Big(&K^2S^2+K^{3/2}\left(S^3+BNS\right)+\\ &+\left(K\left(N^2M+NMS+BNS+S^3+S^2BN\right)+N^2M\right), \end{aligned}$$

where B and S are the truncation values for the series in (38) and (28), respectively, and K is the total number of Zernike polynomials included in the pupil representation (25).

Now we switch to the computational complexity of the GRBF based formulas described in Section 3. Again, we are interested in evaluating expression (19) at a grid $\boldsymbol{r}\in\mathbb{R}^N$, $\boldsymbol{\phi}\in\mathbb{R}^N$ and $\boldsymbol{f}\in\mathbb{R}^M$, with the additional assumption (14), so that formula (20) is used.

According to the discussion in Section 3.2, once the cut-off parameter S has been chosen, the double recurrence (17)-(18) must be evaluated for $s\leq S$, with an estimated computational cost of $\mathscr{O}(4SM)$ flops. After that, the calculation of Ω in (21) takes about $\mathscr{O}\left(7N^2K\right)$ flops. The computational load of finding $\boldsymbol{d}$ by (22) is almost negligible, requiring only $\mathscr{O}(4K)$ flops. Recall that Ω and $\boldsymbol{d}$ are computed only once during the initialization of the algorithm.

The computational complexity of evaluating each H_s in the mesh is of $\mathscr{O}\left(K+2N^2\right)$

Method	Complexity (single f)	Complexity (vector of f)
FFT2	$\mathcal{O}(N^2\log(N))$	$\mathcal{O}(MN^2\log(N))$
ENZ	$\mathcal{O}(N^2K+NK^{3/2})$	$\mathcal{O}(N^2M+N^2KM+NK^{3/2})$
ENZ improved	$\mathcal{O}(N^2M+NK^{3/2}+K^2)$	$\mathcal{O}(N^2M+N^2KM+NK^{3/2}+K^2)$
GRBF	$\mathcal{O}(N^2K)$	$\mathcal{O}(N^2M+N^2K)$

Table 1: Estimates of the minimal computational complexity of the methods when evaluating U at a $N\times N$ grid of nodes (r,ϕ) and for M values for the defocus parameter f, using K functions in the corresponding series expansions (for ENZ and GRBF).

operations, yielding a total of $\mathcal{O}\left(KS+2SN^2\right)$ flops for the evaluation of all the required H_s terms.

Summarizing, the cost of calculating U in (19) according to Algorithm 1 is of

$$\mathcal{O}\left(2N^2MS+7KN^2+N^2M\right).$$

Recall also that the minimal estimated computational complexity of the FFT2 method for a single value of f, even for optimal implementation, is of $\mathcal{O}(N^2\log(N))$, which corresponds to the cost of the FFT, the most computationally demanding part.

Table 5.2 summarizes the leading terms of the computational cost of each method, with the assumption that the same values of B and S for all semi-analytical methods have been chosen. Comparison between rows two and three shows that the GRBF approach is much more efficient than the ENZ theory, especially for a large amount of values for f. The FFT2-based method seems to be of similar complexity to GRBF, but in practice the number of sample points N needed to achieve a reasonable accuracy for FFT2 will be much larger than that required for GRBF, which in our experiments was set to 400 (see the description in the next section).

5.3. Execution time

Since the complexity estimates give only a rough idea of the computational demand of a method, we have looked also at the execution time, which is a simpler and a more informative assessment.

For the beginning, we run the ENZ, the improved ENZ and the GRBF algorithms evaluating U at an 100×100 mesh of nodes for a single value of the parameter f, recording the execution time in dependence of the value of functions used in the corresponding series expansions. The experiments were performed on a standard PC running Matlab v. 8.1. Figure 1 shows the results, along with the corresponding regression curves. The value of the slope of these curves at the origin are approximately 0.021 seconds/function for ENZ and 0.0028 seconds/function for GRBF. This gives a ratio of at least 7.5 times faster for the GRBF approach, with the same number of functions, or reversely, one can use 7.5 times more functions in the GRBF scheme, for the same execution time. For comparative purposes, the execution time to evaluate function U numerically making use of the two-dimensional fast Fourier transform was of approximately 0.25 seconds, matching the execution time for ENZ using about 12 Zernike terms, or for GRBF with approximately 90 Gaussian RBFs.

In order to achieve a reasonable comparison, in the rest of our experiments we implemented a standard setting for each of the methods. Namely, the integral U was evaluated at a 100×100 regular grid of nodes by the semi-analytic methods, and at a 512×512 regular grid for the FFT2-approach (this is a realistic size needed to overcome aliasing and obtain accurate results). Moreover, we used $K = K_{GRBF} = 400$ functions in the formula (19) (corresponding to the 20×20 regular grid of centers, as described above) and the cut-off parameter $S = 60$, while for the ENZ method it was observed that a total of $K = K_{ENZ} = 45$ Zernike polynomials (up to the 8th order polynomials) was the optimal. This is actually about twice the number

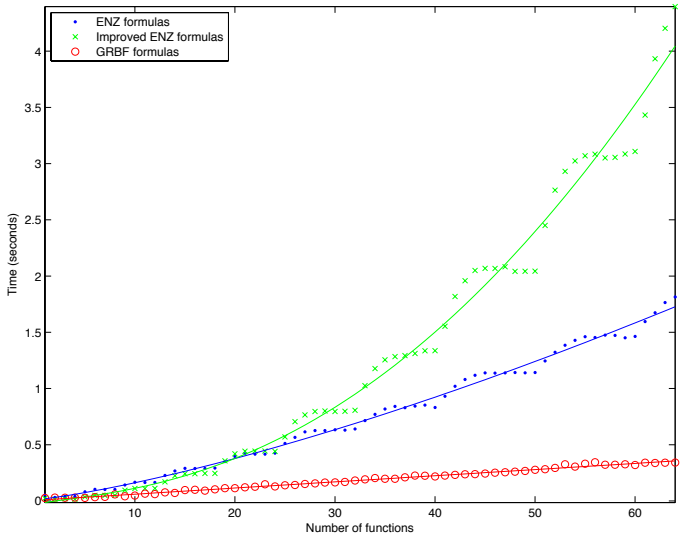

Figure 1: Dependence of the execution time from the number of functions used in the description of the complex pupil function.

of terms recommended by [17], [18]. Higher order polynomials, according to our experiments, did not contribute to higher accuracy, while causing ill conditioning of the computations.

With these settings, we also compared the execution time of the three methods as a function of the length of the vector of defocus parameters $\boldsymbol{f}$. The results appear in Figure 2, along with the regression lines for each scheme. The values of the slopes are approximately 0.24 for FFT2, 0.16 for ENZ and 0.003 for GRBF, all in seconds per value of f. This means that FFT2 is about 75 times slower than GRBF when calculating U for many of values of f simultaneously, while ENZ is also about 50 times slower than GRBF, even though the number of Gaussian functions used was much higher than the number of Zernike polynomials.

5.4. Accuracy

The ENZ-theory, although representing a big step forward, has some limitations that must be taken into account. The obvious one is the use of only even terms in the Zernike expansion of the complex pupil function, which restricts it to the symmetric wavefront errors. Some other, less evident, problems lie in the core of the mathematical properties of the ENZ explicit formula for $U(r,\phi;f)$. This is an infinite series of terms, each of them a finite linear combination of Bessel functions, and each new Zernike term added to the expansion of the pupil function (11) increases the complexity of the terms. The series is slowly convergent, especially for larger values of f, requiring a truncation with a large number of terms depending on f (it is recommended to use $3|f|+5$ terms, according to [17], [18]).

Another issue in evaluating the ENZ expressions is the accuracy. The terms of the infinite series with even and odd orders form sign-changing sequences, which increases the risk of the cancellation errors. This phenomenon can be illustrated by the following experiments: in the case when the wavefront is given only by a

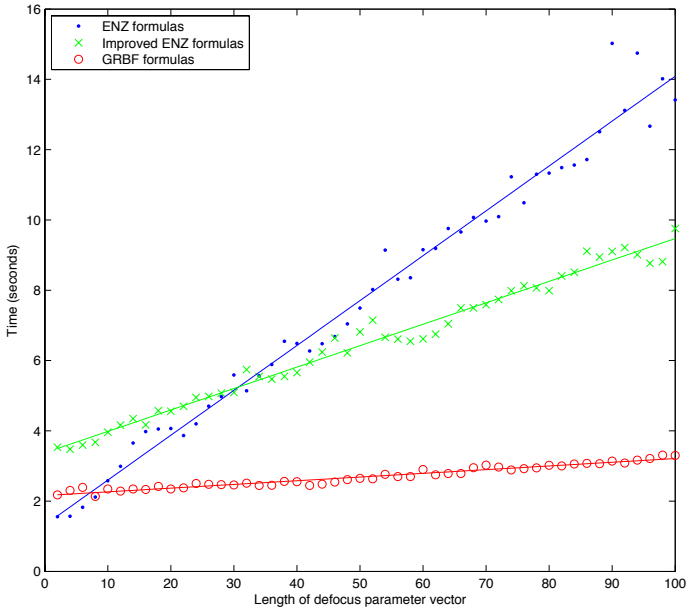

Figure 2: Execution time according to the number of different defocus parameters used in the complex pupil function. A fixed number of functions (400 for GRBF and 45 for ENZ) was used for each value of f.

positive Z_2^2 horizontal astigmatism, the evaluation of, say, the imaginary part of U at a point with radial coordinate $r = 0.9$ consists in adding a finite alternating sequence, with two dominant terms of approximately 0.423, but whose absolute values differ in 3×10^{-5}. This shows that these calculations, if not well organized, can yield a loss of precision in about 5 significant digits. Last but not least, the ENZ formulas contain binomial numbers that must be evaluated with care in order to avoid overflow.

In comparison, the error estimates from Section 3.3 show that the convergence in (15) is very fast, and only a very reasonable number of terms is required for a precise evaluation.

In order to assess accuracy, we carried out a number of numerical experiments where we calculated U. For the ideal wavefront ($\Phi \equiv 0$) and zero defocus ($f = 0$), the results were discussed in [27], where it was shown that the two semi-analytic methods performed similarly, although GRBF calculations were done at a much lower cost.

More informative is to use a synthetic wavefront described by a combination of Zernike polynomial terms and exponential function, seeking a "fair" comparison between the ENZ ("Zernike–oriented") and the GRBF ("exponentially–oriented") approaches. Such class of functions allows also for an easy modeling of low and high-order oscillations. For experiments with a wavefront comprised of basically low-frequency oscillations see [27]. Here we have used the function given by

$$\begin{aligned}\Phi(\rho,\theta) = {} & 0.6\,Z_5^3(\rho,\theta) - 0.4\,Z_4^4(\rho,\theta) - 0.3\,Z_5^5(\rho,\theta) + 0.25\,Z_4^2(\rho,\theta) \\ & + 0.25\,Z_6^4(\rho,\theta) - 0.15\,Z_8^4(\rho,\theta) + 0.4\,g(\rho\cos\theta,\rho\sin\theta;-0.3,0,15) \\ & - 2\,[g(\rho\cos\theta,\rho\sin\theta;0.5,0.3,10) + g(\rho\cos\theta,\rho\sin\theta;0.5,-0.3,10)]\,,\end{aligned} \tag{39}$$

where Z_n^m are the (orthonormal) Zernike polynomials and

$$g(x,y;a,b,\lambda) = \exp\{-\lambda[(x-a)^2+(y-b)^2]\},$$

as well as its "contaminated" version, where we have added a normally distributed random noise of the form `0.5 * randn( )` at a grid of 100×100 equally spaced points (see Figure 3).

For these wavefronts we calculated the diffraction integral (12) for different values of f by quadrature using the scientific software *Mathematica* with extended precision (with the options `PrecisionGoal` set to 8 and `WorkingPrecision` to 16). These values of U (regarded as "exact") were compared with the calculations performed by FFT2 and by two semi-analytic approaches discussed here. For the ENZ we fitted the pupil function using the first 200 Zernike polynomials, while for the GRBF the approximation was performed by the linear combination of 20×20 Gaussian functions, with the parameter $\lambda = 16$. Then the diffraction integral was evaluated by all methods in a grid of 256×256 equally spaced points in the square $[-2,2] \times [-2,2]$.

With the purpose of comparing performance of the computation methods for different values of the defocus parameter f we plot in Figure 4 the values of the normalized PSF, again for the simulated wavefront (39), along the horizontal line ($\phi = 0, \pi$ and $r \in [0,1]$), setting $f = 0$, $f = 2\pi$ and $f = -2\pi$, as in a numerical experiment described in [18]. In the case $f = 0$ the dotted line (computed by the ENZ method) is not observed because it matches exactly the other two curves, found using quadrature and the GRBF algorithm. We see that for larger values of defocus our procedure outperforms the alternative methods.

In our last experiments we analyzed the performance of all methods for a non-circular geometry. Namely, for the same synthetic wavefront as before we set an elliptic pupil, taking in (11) as $A(\rho,\theta)$ the characteristic function of the set (in

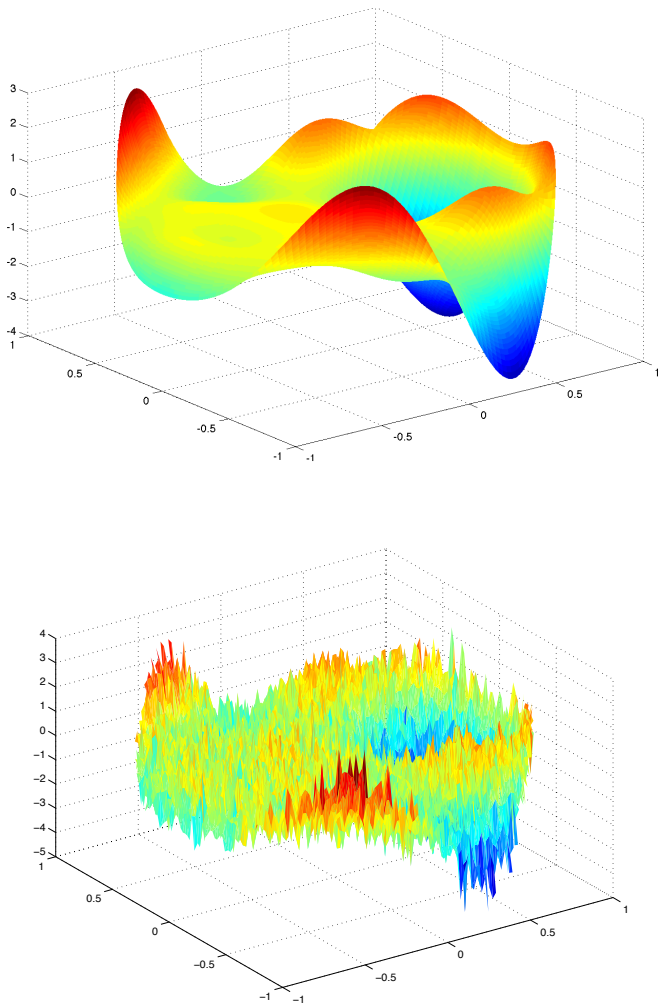

Figure 3: 3D plot of the synthetic wavefront function defined in (39) (upper picture) and of the same wavefront contaminated by a white noise $0.5 \times N(0,1)$ (bottom).

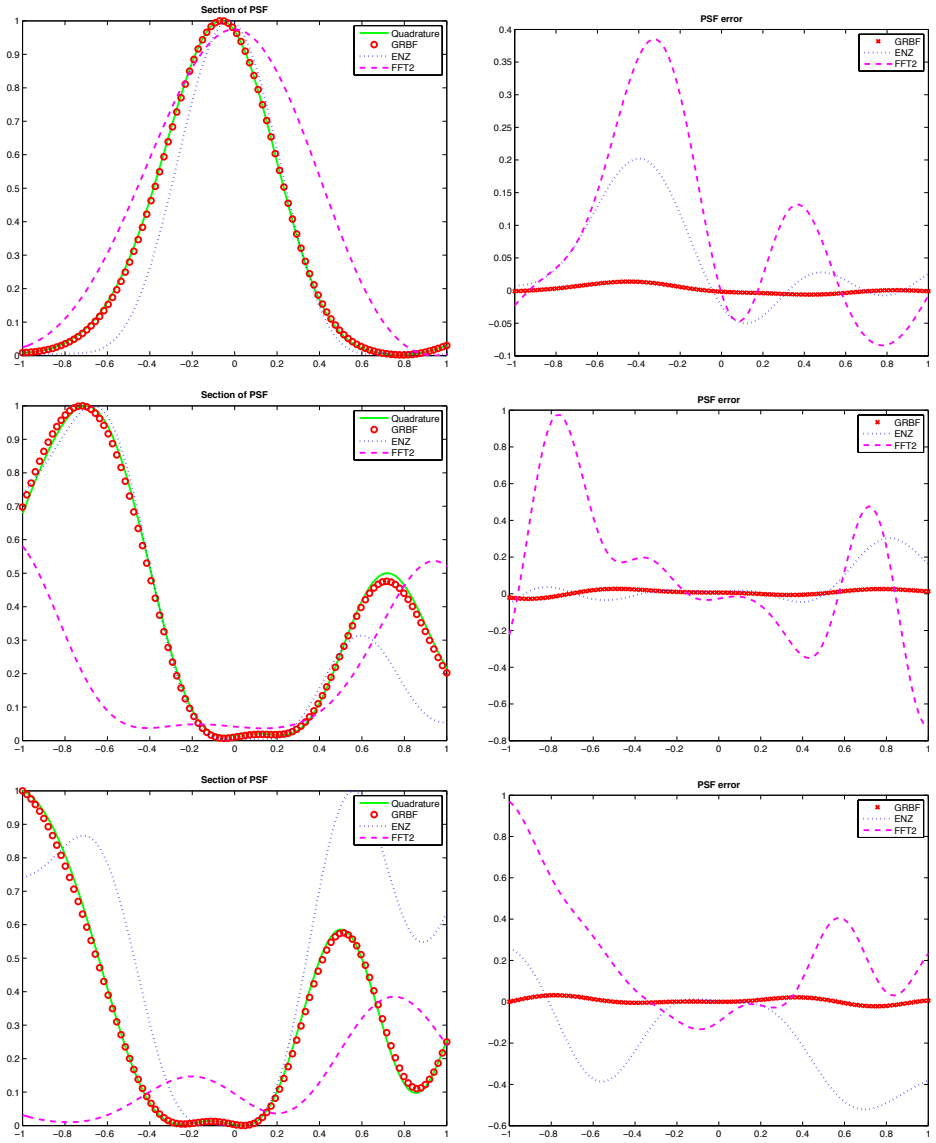

Figure 4: Values of the normalized PSFs and the corresponding errors for the wavefront (39) + noise, along the horizontal diameter of the unit disk, calculated for each method, for $f = 0$ (top row), $f = 2\pi$ (middle) and $f = -2\pi$ (bottom row).

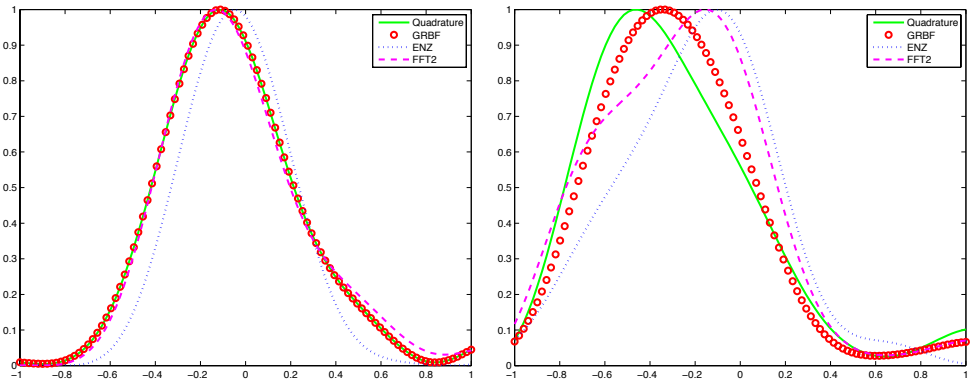

Figure 5: Values of the normalized PSFs for the synthetic wavefront with an elliptic pupil along the horizontal diameter of the unit disk, calculated for each method, for $f = 0$ (left) and $f = \pi$ (right).

cartesian coordinates)

$$x^2 + \frac{y^2}{(0.7)^2} \leq 1.$$

The results are illustrated in Figure 5. The poorer performance of the ENZ method in these examples is due probably to a less accurate fit of the complex wavefront by Zernike polynomials over a non-circular domain.

6. Conclusions

A new procedure for computing 2D Fourier-type integrals, and in particular, the diffraction integrals with variable defocus has been developed. We have discussed some error bounds and developed an efficient scheme for parallel evaluation of these integrals in a grid of points and for a vector of defocus parameters.

It should be noted that when calculating the Fourier-type integrals considered in this paper by replacing the original integrand by its approximant, special care should be put in the quality of approximation. It is easily seen that proximity in an L^2 or even uniform norm is not enough to guarantee small errors (due to a high sensitivity of these integrals to oscillations), and Sobolev-type norms should be

used. The higher accuracy we achieve in the approximation step, the better results will be obtained for the integral calculation. In our case, a linear least squares fit with Tikhonov regularization gave the best results.

The proposed GRBF-based approach has been compared with the two existing procedures, i.e., the 2-dimensional fast Fourier transform and the extended Nijboer-Zernike theory. The results of the comparison show that the new scheme is very competitive, providing higher accuracy and speed. Among other advantages of our method we can mention the following:

- a relative robustness of the computation with respect to the underlying geometry of the pupil function. The GRBF have an almost local character (especially, for higher values of the shape parameter), and the fit that the linear combination of such functions provides is less affected by the shape of the pupil;

- the existence of the shape parameter in the GRBF provides more flexibility to small details, and allows for an easy implementation of a multi-resolution scheme, especially in the case of existence of subdomains with different complexity of the integrand. Since each function used for approximation of the pupil function enters the final expression linearly, one can use two or more layers of GRBF to fit the residual error consecutively using different sets of centers and different shape parameters in order to improve the accuracy of the results;

- in the case of the calculation of diffraction integrals of the form (12), the increase of the computational cost for a vector of values of the defocus parameter is practically negligible, providing a substantial increase in the performance with respect to the other techniques. This is a reliable and efficient

way of obtaining the through-focus characteristics of an optical system at higher resolutions in reasonable time.

In conclusion, the proposed GRBF approach allows calculating through-focus point spread function at a very low computational cost for an arbitrarily selected set of the defocus parameters. This is particularly attractive in those applications in which evaluation of through-focus characteristics of an optical system is required. They include wavefront sensing, phase retrieval, lithography, microscopy, extreme ultraviolet light optics, digital holography and physiological optics. We believe that the GRBF-based method has an even wider scope of application in mathematical imaging and vision.

Acknowledgements

The second and the third authors (AMF and DRL) have been supported in part by the research project MTM2011-28952-C02-01 from the Ministry of Science and Innovation of Spain and the European Regional Development Fund (ERDF), by Junta de Andalucía, Research Group FQM-229 and the Excellence Grant P11-FQM-7276, and by Campus de Excelencia Internacional del Mar (CEIMAR) of the University of Almería. Additionally, DRL was supported by the FPU program of the Ministry of Education of Spain.

This work was completed during a visit of AMF to the Department of Mathematics of the Vanderbilt University. He acknowledges the hospitality of the hosting department, as well as the partial support from the University of Almería, travel grant EST2014/046.

References

References

[1] J. W. Goodman, *Introduction to Fourier Optics*, MaGraw-Hill, 2nd Edition, (1996).

[2] B. H Dean, C. W. Bowers, Diversity selection for phase-diverse phase retrieval, J Opt Soc Am A, **20**(8), 1490–1504 (2003).

[3] B. H. Dean, D. L. Aronstein, J. S. Smith, R. Shiri, D. S. Acton, Phase retrieval algorithm for JWST Flight and Testbed Telescope, Proc. SPIE **6265**, Space Telescopes and Instrumentation I: Optical, Infrared, and Millimeter, 626511 (2006).

[4] S. Thurman, Method of obtaining wavefront slope data from through-focus point spread function measurements, J. Opt. Soc. Am. A **28**, 1–7 (2011).

[5] P. Dirksen, J. J. Braat, A. J. Janssen, A. Leeuwestein, H. Kwinten, D. Van Steenwinckel, Determination of resist parameters using the extended Nijboer-Zernike theory, Proc. SPIE **5377**, Optical Microlithography XVII, 150 (2004).

[6] P. Dirksen, J. J. Braat, A. J. Janssen, A. Leeuwestein, Aberration retrieval for high-NA optical systems using the extended Nijboer-Zernike theory, Proc. SPIE **5754**, Optical Microlithography XVIII, 262 (2005).

[7] A. Artal, S. Marcos, I. Miranda, and M. Ferro, "Through focus image quality of s implanted with monofocal and multifocal intraocular lenses", Opt. Eng. **34**, 772–779 (1995).

[8] S. Marcos, S. Barbero, and I. Jimenez-Alfaro, "Optical quality and depth-of-field of eyes implanted with spherical and aspheric intraocular lenses," J. Refract. Surg. **21**, 1–13 (2005).

[9] P. A. Piers, H. A. Weeber, P. Artal, and S. Norrby, "Theoretical comparison of aberration-correcting customized and aspheric intraocular lenses," J. Refract. Surg. **23**, 374–384 (2007).

[10] D. A. Atchison, "Depth of focus of the human eye", in *Presbyopia: Origins, effects and treatment*, I. Pallikaris, S. Plainis, and W. N. Charman, eds. (Slack Incorporated, 2012).

[11] F. Yi, D. R. Iskander, and M. J. Collins, "Estimation of the depth of focus from wavefront measurements", J. Vis. **10**, 4:3 (2010).

[12] W. J. Donnelly, III A. Roorda, Optimal pupil size in the human eye for axial resolution J Opt Soc Am A, **20**(11), 2010–2015 (2003).

[13] U. Schnars, W. Jüptner. *Digital Holography*. Springer Verlag, Berlin, 2005.

[14] J. T. Sheridan, "Optical signal processing: holography, speckle and algorithms". In: Signal Recovery and Synthesis, OSA Technical Digest (CD) (Optical Society of America, 2011).

[15] B. Hennelly, D. Kelly, N. Pandey and D. Monaghan, "Zooming algorithms for Digital Holography". Journal of Physics: Conference Series **206** (2010), 012027.

[16] Ch. Liu, D. Wang, J. J. Healy, B. M. Hennelly, J. T. Sheridan, and M. K. Kim, "Digital computation of the complex linear canonical transform". J Opt Soc Am A **28** (7), 1379–1386 (2011).

[17] A. J. E. M. Janssen, "Extended Nijboer–Zernike approach for the computation of optical point-spread functions," J. Opt. Soc. Am. A **19**, 849–857 (2002).

[18] J. J. M. Braat, P. Dirksen, and A. J. E. M. Janssen, "Assessment of an extended Nijboer–Zernike approach for the computation of optical point-spread functions", J. Opt. Soc. Am. A **19**, 858–870 (2002).

[19] A. J. E. M. Janssen, J. J. M. Braat, and P. Dirksen, "On the computation of the Nijboer–Zernike aberration integral at arbitrary defocus", J. Modern Optics **51**(5), 687–703 (2004).

[20] M. Born and E. Wolf. *Principles of Optics* (4th rev. ed., Pergamon Press, 1970).

[21] G. E. Fasshauer. *Meshfree Approximation Methods with MATLAB* (World Scientific, 2007).

[22] M. Montoya-Hernández, M. Servín, D. Malacara-Hernández, and G. Paez. Wavefront fitting using Gaussian functions. Opt Commun. **163**, 259–269 (1999).

[23] S. Rippa, "An algorithm for selecting a good value for the parameter c in radial basis function interpolation", Adv. Comput. Math. **11**, 193–210 (1999).

[24] G. E. Fasshauer and J. G. Zhang, "On choosing 'optimal' shape parameters for RBF approximation", Numer. Algor. **45**, 345–368 (2007).

[25] A. Martínez-Finkelshtein, A.M. Delgado, G.M. Castro-Luna, A. Zarzo, and J.L. Alió. Comparative analysis of some modal reconstruction methods of the shape of the cornea from corneal elevation data. Invest. Ophtalmol. Vis. Sci. **50**, 5639–5645 (2009).

[26] A. Martínez-Finkelshtein, D. Ramos-López, G.M. Castro-Luna, and J.L. Alió. Adaptive corneal modeling from keratometric data. Invest. Ophtalmol. Vis. Sci. **52**, 4963-4970 (2011).

[27] D. Ramos-López, A. Martínez-Finkelshtein, and D. R. Iskander. Computational aspects of the through-focus characteristics of the human eye. J. Opt. Soc. Am. A **31** (7), 1408-1415 (2014).

[28] *NIST handbook of mathematical functions*. Edited by F. W. J. Olver, D. W. Lozier, R. F. Boisvert and Ch. W. Clark. National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge, 2010.

[29] C. Hansen. *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion* (SIAM, 1998).

[30] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari. "Tikhonov regularization and the *L*-curve for large discrete ill-posed problems", J. Comput. Appl. Math. **123**, 423–446 (2000).

[31] F. Andersson, M. Carlsson, M. V. de Hoop, "Nonlinear approximation of functions in two dimensions by sums of exponential functions", Appl. Comput. Harmon. Anal. **29**, 156–181 (2010).

[32] G. A. Baker, and P. Graves-Morris. *Padé approximants* (2nd ed., Encyclopedia of Mathematics and its Applications, 59. Cambridge University Press, Cambridge, 1996).

[33] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover Publications, N.Y., 1972).

[34] D. H. Bailey, and P. N. Swartztrauber. “A fast method for the numerical evaluation of continuous Fourier and Laplace transforms”, SIAM J. Sci. Comp. **15**, Vol. 5, 1105–1110 (1994).

[35] E. Candès, L. Demanet, and L. Ying. “A fast butterfly algorithm for the computation of Fourier integral operators”, Multiscale Model. Simul. **7**, Vol. 4, 1727–1750 (2009).

[36] M. Gai, and R. Cancelliere. “An efficient point spread function construction method”, Mon. Not. R. Astron. Soc. **377**, 1337–1342 (2007).