Analytic design of optimum holographic optical elements

J. N. Cederquist and J. R. Fienup

Environmental Research Institute of Michigan, P.O. Box 8618, Ann Arbor, Michigan 48107

Received April 21, 1986; accepted December 1, 1986

A method is developed for analytically determining the holographic optical element phase function that is optimum for transforming a set of input wave fronts into a corresponding set of output wave fronts. These sets are allowed to be infinite in the sense that the wave-front phases can be given as functions of continuous parameters. The method can be tolerant of specified wave-front aberrations, with the optimum amount of these aberrations determined as part of the solution process. For many practical design problems, the phase function and its first derivatives will be continuous. The method is applied to the design of a one-dimensional Fourier-transform holographic element with the input wave-front angle of arrival as a continuous parameter and with the optimum distortion of the output plane determined by the solution. The resulting design compares favorably with other work using damped-least-squares optimization.

1. INTRODUCTION

Holographic optical elements (HOE's) have proved to be useful for several applications.[1] Compared with conventional refractive and reflective optics, they can be thinner and more lightweight and have the potential for being very inexpensive in mass production. They can also perform multiple functions simultaneously (for example, focusing, beam splitting, and spectral filtering, all in the same area of the HOE).

A HOE can be defined by the object and reference wave fronts that are used to construct it. If the object and reference wave fronts have phase functions $\phi_{\text{obj}}(\mathbf{x})$ and $\phi_{\text{ref}}(\mathbf{x})$, respectively, where $\mathbf{x}$ is a two-dimensional (2-D) vector in the x-y plane, the HOE phase function $\phi_H(\mathbf{x})$ is

$$\phi_H(\mathbf{x}) = \phi_{\text{obj}}(\mathbf{x}) - \phi_{\text{ref}}(\mathbf{x}). \tag{1}$$

If the HOE is illuminated by a wave front with phase $\phi_{\text{in}}(\mathbf{x})$, the phase $\phi_{\text{out}}(\mathbf{x})$ of the output wave front will be

$$\phi_{\text{out}}(\mathbf{x}) = \phi_{\text{in}}(\mathbf{x}) + \phi_H(\mathbf{x}). \tag{2}$$

In most applications, one can define a set of input wave fronts and a corresponding set of desired output wave fronts. Usually the phase of the input wave front can be written as $\phi_{\text{in}}(\mathbf{x}, \alpha)$, where α is a 2-D vector parameter that may represent different points on an object or different field angles, for example. Different values of α give different members of the set of input wave fronts. If $\phi_{\text{out}}(\mathbf{x}, \alpha)$ is similarly defined, then, for any α, the desired HOE phase function $\phi(\mathbf{x}, \alpha)$ is

$$\phi(\mathbf{x}, \alpha) = \phi_{\text{out}}(\mathbf{x}, \alpha) - \phi_{\text{in}}(\mathbf{x}, \alpha). \tag{3}$$

For a given α, a HOE with a phase function $\phi(\mathbf{x}, \alpha)$ defined by Eq. (3) will perfectly transform $\phi_{\text{in}}(\mathbf{x}, \alpha)$ into $\phi_{\text{out}}(\mathbf{x}, \alpha)$. In general, $\phi(\mathbf{x}, \alpha)$ varies with α, so that a HOE with a phase function $\phi_H(\mathbf{x})$ equal to $\phi(\mathbf{x}, \alpha)$ for one value of α will perform ideally only for that value of α and will have aberrations given by the difference between $\phi_H(\mathbf{x})$ and $\phi(\mathbf{x}, \alpha)$ for other values of α.

A quality criterion can be used to quantify the performance of any given HOE phase function $\phi_H(\mathbf{x})$. For this investigation, the criterion used was a squared-error metric, the square of the phase difference $\phi_H(\mathbf{x}) - \phi(\mathbf{x}, \alpha)$ integrated and possibly weighted over the relevant values of $\mathbf{x}$ and α. This commonly used criterion is closely related to the Strehl ratio[2] and correlates with the root-mean-square (rms) spot size. The objective in HOE design is to find the optimum HOE whose phase function $\phi_H(\mathbf{x})$ minimizes the error metric.

Early work in HOE design was concentrated on defining phase functions $\phi_H(\mathbf{x})$ that would be formed from object and reference beams [see Eq. (1)] that could easily be formed in the laboratory and were simple to describe mathematically (i.e., plane waves and point sources). Since each of the two beams can be specified by three parameters, such a HOE would possess only six degrees of freedom (neglecting such possibilities as changing wavelengths between recording and readout). As the field matured, the number of possible degrees of freedom was greatly enlarged by allowing the object and reference beams to be defined by arbitrary auxiliary optical systems that would produce them.[3] The number of degrees of freedom possible in the HOE is then equal to the number of degrees of freedom allowed in the design of the auxiliary optical systems. Another step forward was taken by allowing the object and reference beams to be described by arbitrary mathematical functions, such as a polynomial series, having a number of free parameters or coefficients.[3] The number of possible degrees of freedom of the HOE is then the number of coefficients that are associated with the mathematical function. In this case, the recording wave-front descriptions are divorced from the means for producing them in the laboratory, and so, after designing the HOE, one must then also design optical systems, which are likely to include computer-generated holograms, for producing the desired recording wave fronts.

In all these ways of describing wave fronts, the design approach is to optimize an error metric over the space of available HOE parameters by using an iterative search method such as damped least squares. Even if the global minimum of the error metric is found for the given set of parameters, one could not be sure that the best possible (optimum) HOE was found, since the optimum HOE may

0740-3232/87/040699-07$02.00

not be describable by the set of parameters used to characterize and optimize the HOE.

More recently, an entirely different approach to HOE design that is capable of finding the optimum HOE was invented.[4] It uses the fact that an analytic solution to the optimum HOE can be found when employing a squared-error metric. That analysis, however, permits only a finite set of input and output wave fronts; i.e., the parameter α takes only a finite number of discrete values. Subsection 2.A of this paper describes a method for improving performance by allowing the parameter α to vary continuously. Subsection 2.B discusses conditions under which the HOE phase $\phi_H(\mathbf{x})$ and its first derivatives are continuous (as is desirable for any HOE that is to be fabricated and to avoid diffraction effects at discontinuities). The earlier method[4] based on a finite set of wave fronts does not guarantee phase continuity. In Subsection 2.C, the method is extended to allow for cases in which either the input phase $\phi_{\text{in}}(\mathbf{x}, \alpha)$ or the output phase $\phi_{\text{out}}(\mathbf{x}, \alpha)$ is not completely specified and some types of aberration can be permitted to exist in order to improve performance in terms of other aberrations that are more important to the application. A numerical example of the method and a comparison with earlier results are given in Section 3. Section 4 comments on another recently published method, and conclusions are stated in Section 5.

2. THEORETICAL DEVELOPMENT

A. Basic Theory

As mentioned in the introduction, the HOE phase function is defined as $\phi_H(\mathbf{x})$, and the desired HOE phase function for any value of the parameter α is defined as $\phi(\mathbf{x}, \alpha)$. Further, define a pupil function $P(\mathbf{x}, \alpha)$. The pupil function can be simply a binary function such that $P(\mathbf{x}, \alpha) = 1$ if the point $\mathbf{x}$ on the HOE is illuminated by the input wave front with parameter α, and $P(\mathbf{x}, \alpha) = 0$ otherwise. It can also be a continuous function to allow for weighting of different parts of an input wave front. In all cases

$$0 \le P(\mathbf{x}, \alpha) \le 1. \tag{4}$$

A second weighting function $W(\alpha)$ can also be defined to vary the relative contributions to the performance criterion of different input wave fronts. Again, in all cases,

$$0 \le W(\alpha) \le 1. \tag{5}$$

Finally, a performance criterion E can be defined as the weighted squared-error metric

$$E = \int\int W(\alpha)P(\mathbf{x}, \alpha)[\phi_H(\mathbf{x}) - \phi(\mathbf{x}, \alpha)]^2 \mathrm{d}\alpha \mathrm{d}\mathbf{x}, \tag{6}$$

where the integration in $\mathbf{x}$ and α may be considered to have infinite limits and the pupil function $P(\mathbf{x}, \alpha)$ can be used to define the actual regions of integration.

A variational method[5] may be used to find the optimum HOE phase $\phi_H(\mathbf{x})$ that minimizes E. Specifically, for a variation $\delta\phi_H(\mathbf{x})$ of the HOE phase function $\phi_H(\mathbf{x})$, the variation δE of the criterion E is

$$\delta E = 2\int\int W(\alpha)P(\mathbf{x}, \alpha)[\phi_H(\mathbf{x}) - \phi(\mathbf{x}, \alpha)]\delta\phi_H(\mathbf{x})\mathrm{d}\alpha \mathrm{d}\mathbf{x}. \tag{7}$$

For E to be a minimum [and therefore for $\phi_H(\mathbf{x})$ to be the optimum HOE phase]

$$\delta E = 0. \tag{8}$$

Since $\delta\phi_H(\mathbf{x})$ is arbitrary, Eqs. (7) and (8) imply that

$$\int W(\alpha)P(\mathbf{x}, \alpha)[\phi_H(\mathbf{x}) - \phi(\mathbf{x}, \alpha)]\mathrm{d}\alpha = 0 \tag{9}$$

for all values of $\mathbf{x}$. The optimum HOE phase is therefore

$$\phi_H(\mathbf{x}) = \frac{\int W(\alpha)P(\mathbf{x}, \alpha)\phi(\mathbf{x}, \alpha)\mathrm{d}\alpha}{\int W(\alpha)P(\mathbf{x}, \alpha)\mathrm{d}\alpha}. \tag{10}$$

Equation (10) is a key result, for it expresses the optimum HOE phase in terms of a desired HOE phase that is a function of a continuous parameter α.

If one uses a computer-generated hologram as the HOE, then $\phi_H(\mathbf{x})$ may be encoded directly. On the other hand, if the HOE is interferometrically recorded, then the choice of either the object beam or the reference beam is arbitrary, as long as Eq. (1) is satisfied; that is, while the difference between $\phi_{\text{obj}}(\mathbf{x})$ and $\phi_{\text{ref}}(\mathbf{x})$ is fixed by Eq. (1), their sum is arbitrary. For volume phase HOE's, this freedom allows one to optimize the diffraction efficiency by manipulating the Bragg angle independently of $\phi_H(\mathbf{x})$.

B. Continuity Considerations

For Eq. (10) to be useful in practical applications, it is desirable that the HOE phase $\phi_H(\mathbf{x})$ and its first derivatives be

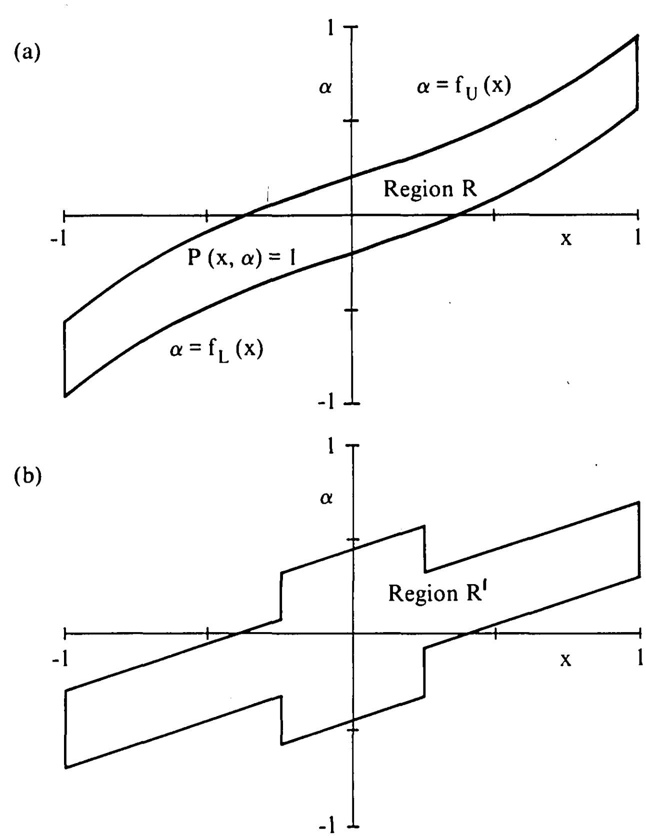

Fig. 1. Pupil functions that give continuous (a) and discontinuous (b) optimum HOE phase functions.

continuous. For most practical cases, $\phi_H(\mathbf{x})$ computed by Eq. (10) will be continuous. The continuity of the case in which $\mathbf{x}$ and α are one-dimensional (1-D) parameters and $P(\mathbf{x}, \alpha)$ is a binary function is analyzed here. Other cases are simple extensions of this analysis.

Assume that $W(\alpha)$ and $\phi(\mathbf{x}, \alpha)$ are continuous in the region R where $P(x, \alpha) = 1$ and that Eq. (5) holds. Assume that the region R is such that its upper boundary is given by the function $\alpha = f_U(x)$ and its lower boundary is given by $\alpha = f_L(x)$, as illustrated in Fig. 1(a). Equation (10) may then be written as

$$\phi_H(x) = \frac{\int_{f_L(x)}^{f_U(x)} W(\alpha)\phi(x, \alpha)\mathrm{d}\alpha}{\int_{f_L(x)}^{f_U(x)} W(\alpha)\mathrm{d}\alpha}. \tag{11}$$

If the numerator and the denominator are continuous functions of x and the denominator is nonzero, then $\phi_H(x)$ will be continuous for values of x in the projection of the region R onto the x axis. The denominator will be a continuous function of x if $f_L(x)$ and $f_U(x)$ are continuous. For example, for $W(\alpha) = 1$, the denominator will be a continuous function of x for the region R of Fig. 1(a) but not for the region R$'$ of Fig. 1(b). A similar argument holds for the numerator assuming that $\phi(x, \alpha)$ is continuous in the region R. Since the functions $f_U(x)$ and $f_L(x)$ will be continuous in many practical cases, $\phi_H(x)$ will be also.

For many, but not all, practical cases, the first derivatives $\partial\phi_H(\mathbf{x})/\partial x$ and $\partial\phi_H(\mathbf{x})/\partial y$ will again be continuous. An analysis is given here for the 1-D case. Differentiating Eq. (11) gives

$$\frac{\mathrm{d}\phi_H(x)}{\mathrm{d}x} = \left[\int_{f_L(x)}^{f_U(x)} W(\alpha)\mathrm{d}\alpha\right]^{-1}\left[\int_{f_L(x)}^{f_U(x)} W(\alpha)\frac{\partial\phi}{\partial x}(x, \alpha)\mathrm{d}\alpha + W[f_U(x)]\phi[x, f_U(x)]\frac{\mathrm{d}f_U(x)}{\mathrm{d}x} - W[f_L(x)]\phi[x, f_L(x)]\frac{\mathrm{d}f_L(x)}{\mathrm{d}x}\right]$$
$$- \frac{\left[\int_{f_L(x)}^{f_U(x)} W(\alpha)\phi(x, \alpha)\mathrm{d}\alpha\right]\left[W[f_U(x)]\frac{\mathrm{d}f_U(x)}{\mathrm{d}x} - W[f_L(x)]\frac{\mathrm{d}f_L(x)}{\mathrm{d}x}\right]}{\left[\int_{f_L(x)}^{f_U(x)} W(\alpha)\mathrm{d}\alpha\right]^2}. \tag{12}$$

Using the same argument as in the previous paragraph, $\mathrm{d}\phi_H(x)/\mathrm{d}x$ will be continuous if $W(\alpha)$, $\phi(x, \alpha)$, $f_L(x)$, $f_U(x)$, $\partial\phi(x, \alpha)/\partial x$, $\mathrm{d}f_L(x)/\mathrm{d}x$, and $\mathrm{d}f_U(x)/\mathrm{d}x$ are continuous functions of x and α.

C. Optimum Aberration

In many cases of practical interest, the desired HOE phase function $\phi(\mathbf{x}, \alpha)$ is not completely known. It is useful to let

$$\phi(\mathbf{x}, \alpha) = \phi_0(\mathbf{x}, \alpha) + \phi_1(\mathbf{x}, \alpha), \tag{13}$$

where $\phi_0(\mathbf{x}, \alpha)$ is an initial estimate of $\phi(\mathbf{x}, \alpha)$ and $\phi_1(\mathbf{x}, \alpha)$ is a function with parameters whose values are as yet unknown. For example, if a particular kind of aberration such as distortion or field curvature can be tolerated, then $\phi_1(\mathbf{x}, \alpha)$ would embody a description of variable amounts of these tolerable aberrations. Since one type of aberration can usually be traded off to some extent for another type of aberration, it is usually possible to obtain lower amounts of the aberrations that one cares about if the tolerable errors are allowed to increase. To take advantage of this fact, parameters describing $\phi_1(\mathbf{x}, \alpha)$ should be allowed to vary in such a way as to optimize the performance criterion E. For example, in one dimension, a possible choice is

$$\phi_1(x, \alpha) = \sum_{i=1}^{M}\sum_{j=0}^{N} c_{ij}\alpha^i x^j, \tag{14}$$

where only those coefficients c_{ij} embodying the tolerable errors are allowed to be nonzero. Equation (10) remains the expression for the optimum HOE phase $\phi_H(x)$ but now contains unknown parameters.

The values of the parameters c_{ij} that minimize E can be found by solving the set of simultaneous linear equations

$$\frac{\partial E}{\partial c_{ij}} = 0, \qquad i = 1, M, \qquad j = 0, N. \tag{15}$$

Inserting Eq. (10) into Eq. (6), differentiating, and combining terms yields the set of $M(N + 1)$ linear equations

$$\sum_{k=1}^{M}\sum_{l=0}^{N} a_{ijkl}c_{kl} = b_{ij}, \qquad i = 1, \ldots, M, \qquad j = 0, \ldots, N, \tag{16}$$

where

$$a_{ijkl} = \int\int W(\alpha)P(x, \alpha)\alpha^{i+k}x^{j+l}\mathrm{d}\alpha\mathrm{d}x - \int \frac{\int W(\alpha_1)P(x, \alpha_1)\alpha_1{}^i x^j\mathrm{d}\alpha_1 \int W(\alpha_2)P(x, \alpha_2)\alpha_2{}^k x^l\mathrm{d}\alpha_2}{\int W(\alpha_3)P(x, \alpha_3)\mathrm{d}\alpha_3}\mathrm{d}x \tag{17}$$

and

$$b_{ij} = -\int\int W(\alpha)P(x, \alpha)\phi_0(x, \alpha)\alpha^i x^j\mathrm{d}\alpha\mathrm{d}x + \int \frac{\int W(\alpha_1)P(x, \alpha_1)\alpha_1{}^i x^j\mathrm{d}\alpha_1 \int W(\alpha_2)P(x, \alpha_2)\phi_0(x, \alpha_2)\mathrm{d}\alpha_2}{\int W(\alpha_3)P(x, \alpha_3)\mathrm{d}\alpha_3}\mathrm{d}x. \tag{18}$$

Using this approach, aberrations of the form of Eq. (14) are introduced in amounts determined by the coefficients c_{ij} such that the performance criterion E is minimized. For example, in Eq. (14), terms that are linear in x will permit distortion, and terms that are quadratic in x permit field curvature. Other choices for $\phi_1(x, \alpha)$ can also be made, for example, a sum of Zernike or Legendre polynominals in x.

After solving for the optimum values of the unknown parameters, the optimum HOE phase $\phi_H(x)$ can be found from Eq. (10). Using the form of Eq. (14),

$$\phi_H(x) = \frac{\int W(\alpha)P(x, \alpha)\left[\phi_0(x, \alpha) + \sum_{i=1}^{M}\sum_{j=0}^{N} c_{ij}\alpha^i x^j\right]\mathrm{d}\alpha}{\int W(\alpha)P(x, \alpha)\mathrm{d}\alpha}. \quad (19)$$

3. NUMERICAL EXAMPLE

The preceding theoretical development was applied to the design of an optimum HOE for a 1-D Fourier-transform system. As shown in Fig. 2, the input is a set of plane wave fronts from an aperture of width d, a distance f from the HOE, and centered at angle θ with respect to the normal to the HOE. Relative to the aperture normal, the normals to the wave fronts will cover the range of angles from $-\alpha_0$ to α_0. The desired output wave front will be spherical waves converging to points $u = u(\alpha)$ in a plane a distance f from the HOE and parallel to it. The function $u(\alpha)$ will determine where an input wave front at angle α will focus. If a ray intercepting the HOE at $x = 0$ and at angle θ is required to be diffracted along the z axis, then the grating spacing s of the HOE at $x = 0$ is

$$s = \frac{\lambda}{\sin\theta}. \quad (20)$$

A ray leaving the input aperture at angle α to the normal will intercept the HOE at an angle $\theta - \alpha$. At $x = 0$, it will leave the HOE at an angle α', where, by the grating equation,

$$s[\sin(\theta - \alpha) + \sin\alpha'] = \lambda. \quad (21)$$

It will intercept the output plane at position $u(\alpha)$, where, from Eqs. (20) and (21) and the system geometry,

$$u(\alpha) = f\tan\{\sin^{-1}[\sin\theta - \sin(\theta - \alpha)]\}. \quad (22)$$

Assuming that this is the desired location of the image point, the desired HOE phase function is therefore, from Eq. (3),

$$\phi_0(x, \alpha) = \frac{2\pi}{\lambda}\left(-\{[x - u(\alpha)]^2 + f^2\}^{1/2} + x\sin(\theta - \alpha)\right), \quad (23)$$

which is the difference between the phase of a spherical wave front converging to point $u(\alpha)$ in the output plane and the phase of a plane wave front incident at an angle $\theta - \alpha$.

For this design, it is desired that distortion be permitted to vary. This will allow the final focus locations to be different from those given by the function $u(\alpha)$. The unknown function $\phi_1(x, \alpha)$ was chosen to be

$$\phi_1(x, \alpha) = \frac{2\pi}{\lambda}\sum_{i=1}^{3}\sum_{j=0}^{1} c_{ij}\alpha^i x^j. \quad (24)$$

The weighting function $W(\alpha)$ was chosen as

$$W(\alpha) = 1, \quad (25)$$

and the pupil function $P(x, \alpha)$ was chosen to be a binary function determined by the regions of the HOE that are illuminated by plane waves leaving the aperture of width d at angles in the range $-\alpha_0$ to α_0.

For numerical computation the values

$$\begin{aligned} d &= 2.5 \text{ cm}, \\ f &= 50 \text{ cm}, \\ \theta &= 20°, \\ \alpha_0 &= 2.4°, \\ \lambda &= 514.5 \text{ nm} \end{aligned} \quad (26)$$

were chosen. The resulting pupil function is shown in Fig. 3. The coefficients a_{ijkl} and b_{ij} given by Eqs. (17) and (18) were computed, and Eq. (16) was solved for the parameters c_{ij}, using IMSL, Inc., subroutine LEQ2S. The optimum values for the c_{ij} were found to be

$$\begin{aligned} c_{10} &= -0.0000040 \text{ cm}, \\ c_{20} &= 22.1 \text{ cm}, \\ c_{30} &= 8.1 \text{ cm}, \\ c_{11} &= -0.00014, \\ c_{21} &= -0.00040, \\ c_{31} &= -0.20. \end{aligned} \quad (27)$$

These values were then used in Eq. (19) to determine the optimum HOE phase $\phi_H(x)$.

It is convenient to compare the HOE phase $\phi_H(x)$ with the phase of a conventional Fourier-transform HOE produced

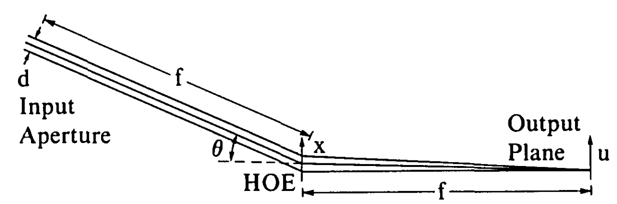

Fig. 2. 1-D Fourier-transform HOE system. Ray trace for input angle $\alpha = 0$ is shown.

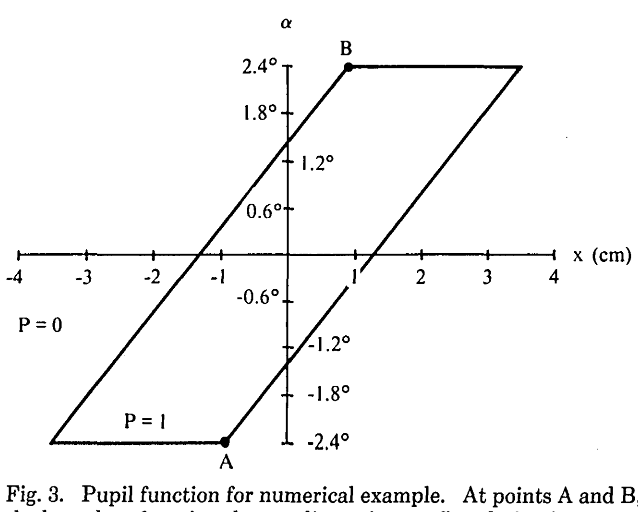

Fig. 3. Pupil function for numerical example. At points A and B, the boundary functions have a discontinuous first derivative.

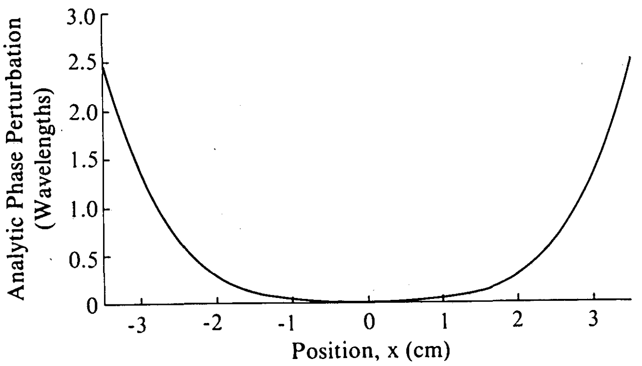

Fig. 4. Difference between optimum and conventional HOE phases.

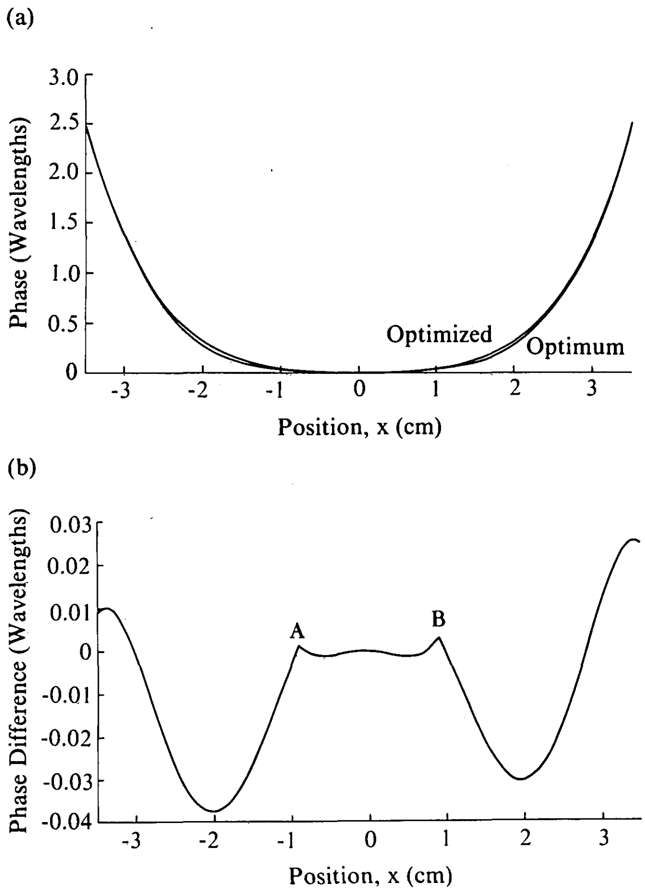

Fig. 5. (a) Optimum and optimized HOE phases (relative to conventional HOE phase). (b) Optimum minus optimized HOE phase. At points A and B, the first derivative of the optimum HOE phase is discontinuous.

by a plane reference wave incident at an angle θ to the HOE normal and a spherical object wave converging to the point $u = 0$ at the center of the output plane. The conventional HOE phase is

$$\phi_0(x, 0) = \frac{2\pi}{\lambda}[x \sin\theta - (x^2 + f^2)^{1/2}]. \tag{28}$$

The difference $\phi_H(x) - \phi_0(x, 0)$ is plotted in Fig. 4, showing that the optimum HOE phase is an aspheric correction to the conventional HOE phase.

This optimum result can be compared with an earlier damped-least-squares-optimized design[3] for a 2-D Fourier-transform HOE. Figure 5(a) shows the two phases (both relative to the conventional HOE phase), and Fig. 5(b) shows their difference. The rms difference is 0.019 wavelength, indicating close agreement between the solutions, which gives confidence in the validity of this approach.

It is interesting to note that, for this example, the conditions for continuity of $\phi_H(x)$ are met and the plot of Fig. 4 shows a continuous $\phi_H(x)$. However, the conditions for continuity of $d\phi_H(x)/dx$ are not met because the functions $f_U(x)$ and $f_L(x)$ have discontinuous derivatives at points A and B in Fig. 3. The discontinuity in $d\phi_H(x)/dx$ is reflected in the corresponding cusps at positions A and B in Fig. 5(b).

For further comparison, the rms wave-front error was also computed and is plotted as a continuous curve in Fig. 6. The rms error for the damped-least-squares-optimized HOE phase as calculated[4] for selected values of α is also shown. (Since these selected values of α are also values at which the design was optimized,[3] it is not surprising that the rms error is lower than that of the optimum design for two of the values. It cannot, of course, be lower when integrated over all values of α.) The rms error is similar for the two designs, with the average value being 0.023 wavelength for the optimum HOE phase and 0.033 wavelength for the damped-least-squares-optimized HOE phase. As expected, the optimum HOE design is better than the damped-least-squares-optimized HOE design. This gives experimental confirmation of the optimum HOE theory.

The distortion $\Delta u(\alpha)$ caused by $\phi_1(x, \alpha)$ will be, to first order,

$$\Delta u(\alpha) = -f \sum_{i=1}^{3} c_{i1}\alpha^i. \tag{29}$$

For plotting purposes it is convenient to show a focus position relative to the focus position $u_0(\alpha)$ of an ideal Fourier-transform system. For the ideal case, position is a linear function of spatial frequency, and so

$$u_0(\alpha) = f \sin\alpha. \tag{30}$$

In Fig. 7(a), $u_0(\alpha)$, $u(\alpha)$, and $u(\alpha) + \Delta u(\alpha)$ are shown. Note that the overall range is 4 cm. In Fig. 7(b), $u(\alpha) - u_0(\alpha)$ and $u(\alpha) + \Delta u(\alpha) - u_0(\alpha)$ are shown. The range is 2.5 mm, and the difference between $u(\alpha)$ and $u(\alpha) + \Delta u(\alpha)$ still cannot be

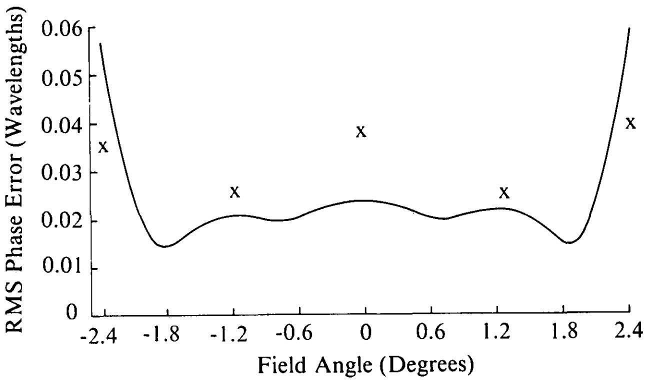

Fig. 6. Rms phase error for optimum HOE phase function. Rms error for optimized HOE phase for selected values of the field angle α are shown by X's for comparison.

seen, as the curves overlap. In Fig. 7(c), $\Delta u(\alpha)$ is plotted. The range is 20 μm. The correction, $\Delta u(\alpha)$, to $u(\alpha)$ is therefore very small. $u(\alpha)$ is a good initial estimate, and including higher-order terms in α in Eq. (24) will have little effect on the optimum HOE phase $\phi_H(x)$.

Note that, as pointed out in Ref. 4, permitting phase

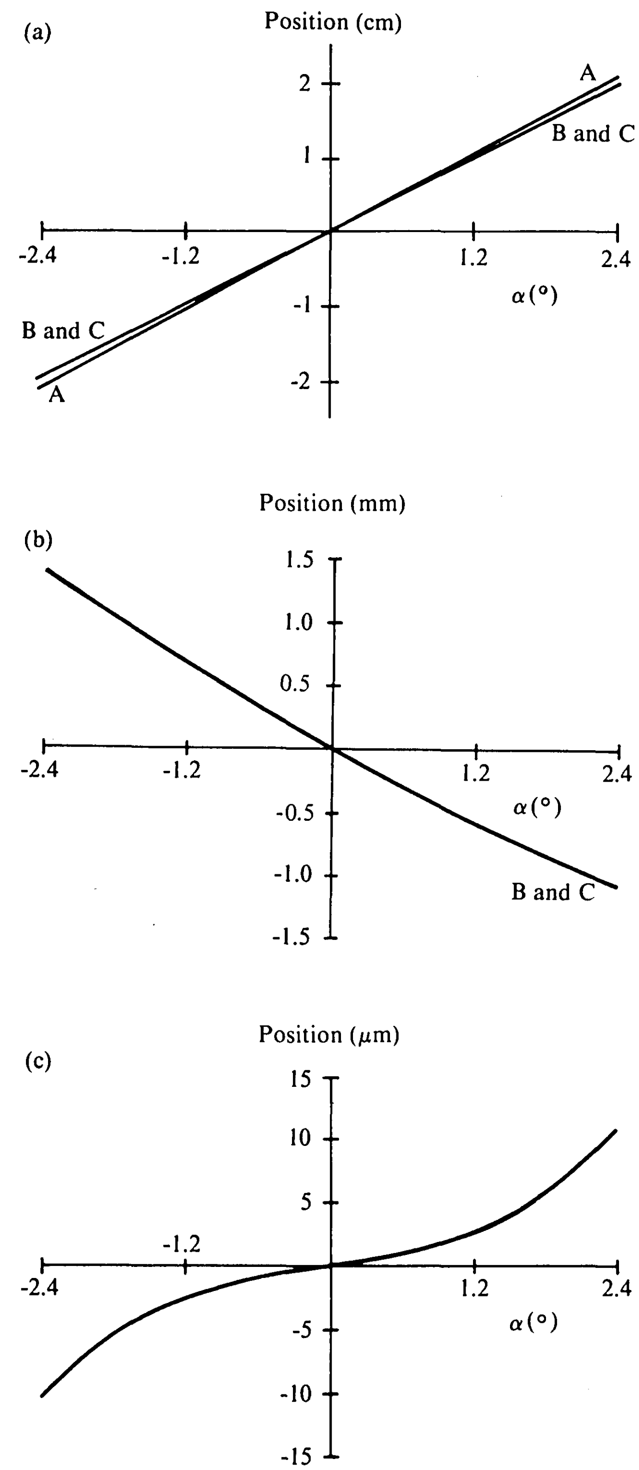

Fig. 7. (a) Ideal (A), estimated (B), and optimum (C) focal positions. (b) Estimated (B) and optimum (C) focal positions relative to ideal ones. The two curves overlap. (c) Difference between optimum and estimated focal positions.

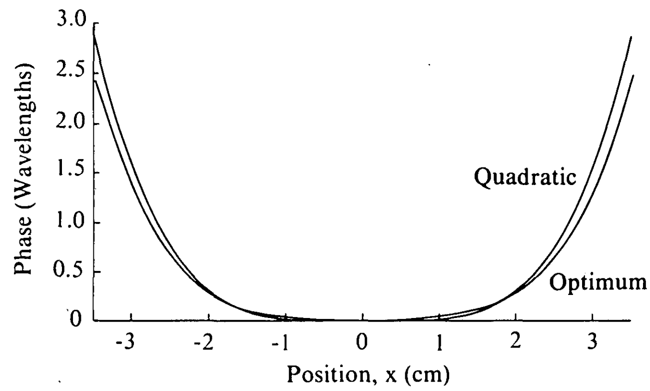

Fig. 8. Optimum and quadratic HOE phases (relative to conventional HOE phase).

variation with angle (the $c_{i0}\,\alpha^i$ terms) and distortion (the $c_{i1}\,\alpha^i x$ terms) greatly affects performance. Had those terms not been included, the performance of the Fourier-transform HOE (Fig. 6) would be many times poorer.

4. COMPARISON WITH THE KEDMI-FRIESEM DESIGN

Kedmi and Friesem also applied the variational method to the design of a Fourier-transform HOE.[6] The resulting design was a quadratic HOE phase function,

$$\phi_H(x) = \frac{2\pi}{\lambda}\left(x \sin\theta - \frac{x^2}{2f}\right). \tag{31}$$

However, in contrast to the development described in Sections 2 and 3 of this paper, they made some simplifications and approximations. For example, in the second paragraph of Sec. 3 of their paper, an approximation is used for the pupil function rather than the exact expression shown in Fig. 3 of this paper. Therefore, contrary to what they claimed, their design cannot be optimum.

Figure 8 compares the optimum HOE phase and the quadratic HOE phase for the design parameters of Section 3. (Both are plotted relative to the conventional HOE phase, as was done in Figs. 4 and 5.) Although it is not optimum, the quadratic HOE phase is significantly better than the conventional HOE phase and may be considered a first approximation to the optimum HOE phase. This interpretation is borne out by the ray traces and experimental results given in Ref. 6, which show reduced spot sizes for the quadratic-phase HOE when compared with a conventional Fourier-transform HOE.

5. CONCLUSION

A mathematical method has been developed that permits an analytic solution for the optimum HOE phase function when the desired HOE phases are defined in terms of continuous parameters. It allows the optimum amount of specific aberrations in the output wave fronts to be determined as part of the solution process. For many practical applications, the optimum HOE phase and its first derivatives will be continuous. In a specific example, the optimum HOE phase was seen to give better performance than a HOE phase determined by damped-least-squares optimization. The advantage of the optimum phase method is that it can consider far more wave fronts (essentially an infinite number) than the previous optimization method[3] without increasing computation time.

The optimum method could be used in conjunction with an optimization method for the design of a complex optical system involving a HOE and more conventional optical components. The optimization method could aid in the design of the conventional part of the system, while the HOE would be updated by the optimum method. This would reduce the number of variables to be optimized and could significantly reduce overall computing time for a system design.

ACKNOWLEDGMENTS

We thank K. Ellis for assistance in computing the numerical example. This research was supported by Kaiser Optical Systems, Inc., under contract 2188.

This paper was presented at the 1985 Annual Meeting of the Optical Society of America in Washington, D.C.[7]

REFERENCES

1. D. H. Close, "Optically recorded holographic optical elements," in *Handbook of Optical Holography*, H. J. Caulfield, ed. (Academic, New York, 1979), pp. 573–585.
2. M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1975), pp. 460–464.
3. R. C. Fairchild and J. R. Fienup, "Computer-originated aspheric holographic optical elements," Opt. Eng. **21**, 133–140 (1982).
4. K. A. Winick and J. R. Fienup, "Optimum holographic elements recorded with nonspherical wave fronts," J. Opt. Soc. Am. **73**, 208–217 (1983).
5. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part 1* (McGraw-Hill, New York, 1953), pp. 275–277.
6. J. Kedmi and A. A. Friesem, "Optimal holographic Fourier-transform lens," Appl. Opt. **23**, 4015–4019 (1984).
7. J. N. Cederquist and J. R. Fienup, "Analytic design of optimum holographic optical elements," J. Opt. Soc. Am. A **2**(13), P68 (1985).