Iterative method applied to image reconstruction and to computer-generated holograms

J. R. Fienup

Radar and Optics Division
Environmental Research Institute
of Michigan
P.O. Box 8618
Ann Arbor, Michigan 48107

Abstract. This paper discusses an iterative computer method that can be used to solve a number of problems in optics. This method can be applied to two types of problems: (1) synthesis of a Fourier transform pair having desirable properties in both domains, and (2) reconstruction of an object when only partial information is available in any one domain. Illustrating the first type of problem, the method is applied to spectrum shaping for computer-generated holograms to reduce quantization noise. A problem of the second type is the reconstruction of astronomical objects from stellar speckle interferometer data. The solution of the latter problem will allow a great increase in resolution over what is ordinarily obtainable through a large telescope limited by atmospheric turbulence. Experimental results are shown. Other applications are mentioned briefly.

Keywords: image reconstruction, holograms, image processing, computers, iterative techniques.

Optical Engineering 19(3), 297-305 (May/June 1980)

INTRODUCTION

There exist a number of mathematical problems in optics that, because of their enormous complexity, do not yield to analytical solutions. When analytical methods fail, it is often possible to solve a problem by an iterative method, of which there are many. In this paper we discuss an iterative method for solving a large class of such problems. The problems fall into two general categories: (1) synthesize a Fourier transform pair having desirable properties in both domains, and (2) reconstruct an object when only partial information is available in each of two domains. A synthesis problem typically arises when one wants the Fourier transform of an object (or a signal, aperture, antenna array, etc.) to have certain desirable properties (such as uniform spectrum, low sidelobes, etc.) while the object itself must satisfy certain constraints or have certain desirable properties. There may not exist a Fourier transform pair that is completely desirable and satisfies all the constraints. Nevertheless, one seeks a Fourier transform pair that comes as close as possible to having the desirable properties and satisfying the constraints in both domains. A reconstruction problem arises when only partial information is measured in one domain, and in the other domain either partial information is measured or certain constraints are known *a priori*. The information available in any one domain is insufficient to reconstruct the object or its complex Fourier transform. Both the synthesis and the reconstruction problems can be expressed as follows:

> Given a set of constraints placed on an object and another set of constraints placed on its Fourier transform, find a Fourier transform pair (i.e., an object and its Fourier transform) that satisfies both sets of constraints.

Once a solution is found to such a problem, the question remains: is the solution unique? For synthesis problems, the uniqueness is usually unimportant—one is satisfied with any solution that satisfies all the constraints; often a more important problem is whether there exists *any* solution that satisfies what may be arbitrary and conflicting constraints. For reconstruction problems, the uniqueness properties of the solution are of central importance. If many different objects satisfying the constaints could give rise to the same measured data, then a solution that is found could not be guaranteed to be the correct solution. Fortunately, as will be described later, the uniqueness of the solution is often not a problem.

Another useful way to classify such problems is according to the type of information available. For one set of problems, the modulus (magnitude or amplitude) of the Fourier transform is measured (or is given) and the object function is known to be real and nonnegative. These include the phase problems of x-ray crystallography, Fourier transform spectroscopy, and imaging through atmospheric turbulence using interferometer data.

For another set of problems, the modulus of a complex-valued object and the modulus of its Fourier transform are measured (or are given), and one wishes to know the phase in both domains. These include the phase retrieval problem in electron microscopy, the design optimization of radar signals and antenna arrays having desirable properties, and phase coding and spectrum shaping problems for computer-generated holograms.

In this paper, we describe the iterative method and show results of computer experiments applying it to two different problems: phase coding for spectrum-shaping and reduction of quantization noise in computer-generated holograms, and reconstruction of space objects from interferometer data. The former is an example of a synthesis problem, and the latter is a reconstruction problem. The extension of the method to solve other problems is reasonably straightforward.

The iterative method is shown to be very effective in solving these problems. The results obtained by the iterative method could not have been achieved by any other practical method. The results of the reconstruction problem are particularly significant: they indicate the possibility of obtaining images of space objects with resolution many times finer than what is ordinarily allowed by the turbulent atmosphere. The

Original manuscript OD-205 received July 15, 1979.
Revised manuscript received December 5, 1979.
Accepted for publication December 20, 1979.
This paper is a revision of a paper presented at the SPIE seminar on Applications of Digital Image Processing III, August 27-29, 1979, San Diego, California, which appears in SPIE proceedings Vol. 207.

iterative method should prove to become an important tool in a number of areas of optics and related fields.

THE ITERATIVE METHOD

The iterative method is not limited to a single fixed algorithm—a number of useful variations exist. The basic Gerchberg-Saxton[1] algorithm, which was originally used to solve a problem in electron microscopy, can be applied to the more general class of problems; we refer to this generalization of the Gerchberg-Saxton algorithm as the "error-reduction" approach. In an attempt to speed up the convergence of the Gerchberg-Saxton algorithm, we arrived at a more powerful approach, which we call the "input-output" approach. In the following, both the error reduction approach and the input-output approach will be described.

Error-reduction approach

The first published account of the error-reduction approach was its use by Gerchberg and Saxton[1] to solve the electron microscopy problem in which both the modulus of a complex-valued image and the modulus of its Fourier transform are measured, and the goal is to reconstruct the phase in both domains. Apparently unknown to them, the error reduction approach was invented somewhat earlier by Hirsch, Jordan, and Lesem[2] to solve a synthesis problem for computer-generated holograms which has a similar set of constraints. (This will be described later in more detail.) The method was again reinvented for a similar problem in computer holography by Gallagher and Liu.[3] The error reduction approach was also used by Gerchberg[4] for a problem in which the complex Fourier transform is measured out to a maximum frequency, and the object is known to have a certain width; the goal is to achieve super-resolution of the object by analytic continuation of its Fourier transform to frequencies beyond the maximum measured frequency. By far, the most concentrated use of the error-reduction approach has been for the electron microscopy problem.[1,5]

For a reconstruction problem, suppose that the object is given by the function f(x) and its Fourier transform by

$$F(u) = |F(u)|\, e^{i\theta(u)} = \mathcal{F}[f(x)] = \int_{-\infty}^{\infty} f(x)\, e^{i2\pi u\cdot x} dx \quad (1)$$

where the vector x may represent spatial, angular, or other coordinates, and the vector u represents spatial, angular, or other frequencies. The coordinates may be one-, two-, or three-dimensional, depending on the problem. For a reconstruction problem, only partial information is available in each domain. Given limited information (or constraints) in each domain, the problem is to reconstruct f(x) and/or F(u). For example, in the Fourier domain, only $|F(u)|$ may be measured and $\theta(u)$ is unknown. For a synthesis problem, one sees an f(x) and F(u) having certain desirable properties (or satisfying certain constraints). For example, in the Fourier domain, it may be desirable to obtain a specified value of $|F(u)|$ while simultaneously having a specified value of $|f(x)|$ in the object domain. Notice that for digital image processing, only a sampled version of the object and Fourier functions are available. We use continuous function notation only as a matter of convenience.

The problem of finding a Fourier transform pair satisfying the constraints in both domains can often be solved by the error-reduction approach, a block diagram of which is shown in Figure 1. One iteration (the k^{th} iteration) of the error-reduction approach proceeds as follows. A trial solution for the object (or an estimate of the object) $g_k(x)$ is Fourier transformed, yielding $G_k(u) = |G_k(u)| \exp[i\phi_k(u)] = \mathcal{F}[g_k(x)]$. $G_k(u)$ is then made to satisfy the Fourier domain constraints. That is, a new Fourier-domain function $G_k'(u)$ is formed from $G_k(u)$ by making the smallest possible changes in $G_k(u)$ that allow it to satisfy the constraints. For example, if the Fourier domain constraint is that the Fourier transform has a modulus equal to $|F(u)|$, then $G_k'(u)$ is given by

$$G_k'(u) = |F(u)|\, e^{i\phi_k(u)} \quad (2)$$

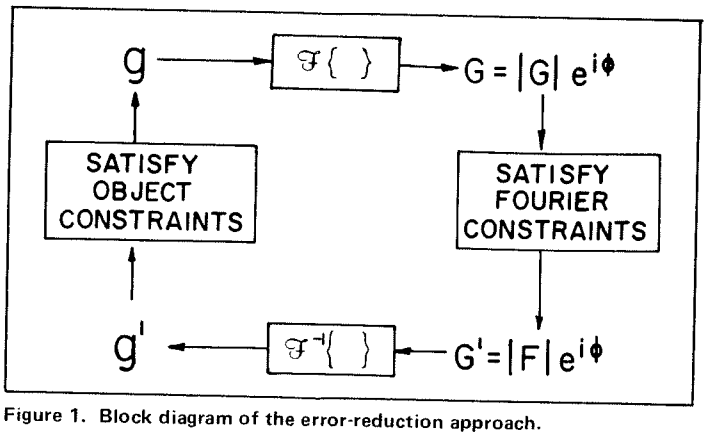

Figure 1. Block diagram of the error-reduction approach.

That is, the given (or measured) modulus $|F(u)|$ is substituted for the modulus of $G_k(u)$, and the phase of $G_k(u)$ is left unchanged. The resulting $G_k'(u)$, which satisfies the Fourier-domain constraints, is inverse Fourier transformed yielding the object-domain function, $g_k'(x)$. Then the iteration is completed by forming a new function, $g_{k+1}(x)$ by making $g_k'(x)$ satisfy the object-domain constraints. In summary, one transforms back and forth between the two domains, forcing the function to satisfy the constraints in each domain. The first iteration can be started in a number of ways, for example, by setting $g_1(x)$ or $\phi_1(u)$ equal to an array of random numbers. The iterations continue until a Fourier transform pair is found that satisfies all the constraints in both domains (or until the money runs out).

A measure of the progress of the iterations, and a criterion by which one can determine when a solution has been found, is the mean-squared error, which is defined in the Fourier domain by

$$E_F^2 = \frac{\int_{-\infty}^{\infty} |G_k(u) - G_k'(u)|^2\, du}{\int_{-\infty}^{\infty} |G_k'(u)|^2\, du} \quad (3)$$

or in the object domain by

$$E_o^2 = \frac{\int_{-\infty}^{\infty} |g_{k+1}(x) - g_k'(x)|^2\, dx}{\int_{-\infty}^{\infty} |g_k'(x)|^2\, dx} . \quad (4)$$

When the mean-squared error is zero, then the object and its Fourier transform satisfy all the constraints, and a solution has been found.

It has been shown for a particular problem[1,6] (and it is perhaps true in the general case) that the mean-squared error can only decrease after each iteration. This fact gives rise to the name error-reduction approach.

Typically, the error is reduced very rapidly for the first few iterations of the error-reduction approach, but more slowly for later iterations. For some applications, the error-reduction approach has been very successful in finding solutions using a reasonable number of iterations. However, for some other applications, the mean-squared error decreases very slowly with each iteration, requiring an impractically large number of iterations for convergence.

Input-output approach

Resulting from an investigation into the problem of the slow convergence of the error-reduction approach, a new and faster converging approach was developed, the input-output approach.[7-9] The input-output approach differs from the error-reduction approach only in the object-domain operation. The first three operations—Fourier transforming g(x), satisfying Fourier domain constraints, and inverse Fourier transforming the result—are the same for both approaches. Those three operations, if grouped together as shown in Figure 2, can be considered as a nonlinear system with an input g(x) and an output g'(x). A property of this system is that its output is always a function having a Fourier

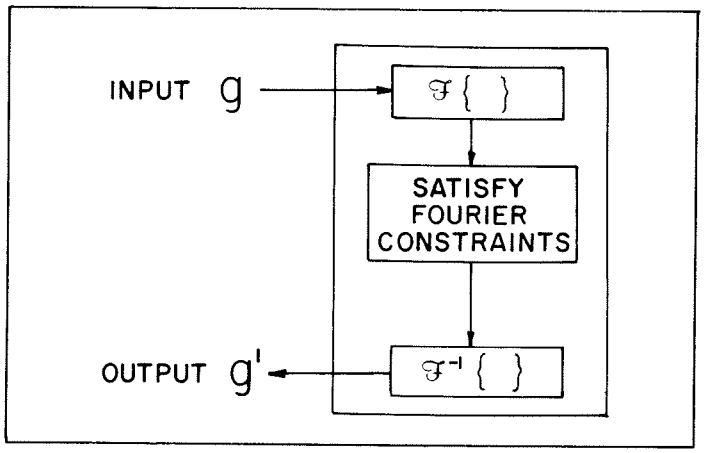

Figure 2. Block diagram of the system for the input-output concept.

transform that satisfies the Fourier-domain constraints. Therefore, if the output also satisfies the object-domain constraints, then all the constraints are satisfied and it is a solution to the problem. Then it is necessary to determine how to manipulate the input in such a way as to force the output to satisfy the object-domain constraints.

For the error-reduction approach, the next input g(x) is chosen to be the current best estimate of the object, satisfying the object-domain constraints. However, for the input-output approach, the input is not necessarily an estimate of the object or a modification of the output, nor does it have to satisfy the constraints; instead, it is the driving function for the next output.

How the input should be changed in order to drive the ouput to satisfy the constraints depends on the particular problem at hand. Specific examples will be shown in the sections that follow. The analysis given in the Appendix for a specific application can be generalized as follows. Consider what happens when an arbitrary change is made in the input. Suppose that at the k^{th} iteration, the input $g_k(x)$ results in the output $g_k'(x)$. Further, suppose that the input is then changed by adding $\Delta g(x)$:

$$g_{k+1}(x) = g_k(x) + \Delta g(x) \quad . \tag{5}$$

Then we would expect the new output resulting from $g_{k+1}(x)$ to be of the form

$$g'_{k+1}(x) = g'_k(x) + \alpha \Delta g(x) + \text{additional noise} \quad . \tag{6}$$

That is, the expected (or statistical mean) value of the change of the output, due to the change $\Delta g(x)$ of input, is $\alpha \Delta g(x)$, a constant times the change of the input. The system shown in Figure 2 is not linear; nevertheless, changes of the input tend to result in similar changes of the output. The expected value of the change of the output can be predicted, but its actual value cannot be predicted since it has a nonzero variance. In Eq. (6), this lack of predictability is indicated by the "additional noise" term. The constant α depends on the statistics of $G_k(u)$ and F(u) and on the Fourier-domain constraints.

If the output $g_k'(x)$ does not satisfy the object-domain constraints and if $g_k(x) + \Delta g_d(x)$ does, then we might try to drive the output to satisfy the constraints by changing the input in such a way as to cause the output to change by $\Delta g_d(x)$. According to Eqs. (5) and (6), the change of the input that will, on the average, cause a change $\Delta g_d(x)$ of the output is

$$\Delta g(x) = \alpha^{-1} \Delta g_d(x) \quad . \tag{7}$$

Then a logical choice for the new input is

$$g_{k+1}(x) = g_k(x) + \beta \Delta g_d(x) \tag{8}$$

where β is a constant ideally equal to α^{-1}, and where $\Delta g_d(x)$ is a function such that $g_k(x) + \Delta g_d(x)$ satisfies the object-domain constraints. If α is unknown, then a value of β only approximately equal to α^{-1} will usually work nearly as well. The use of too small a value of β in Eq. (8) will only cause the algorithm to converge more slowly. The noise-like terms in Eq. (6) are kept to a minimum by minimizing $|\Delta g(x)|$.

As mentioned earlier, for the input-output approach $g_k(x)$ is not necessarily an estimate of the object; it is instead the driving function for the next output. Therefore, it does not matter whether its Fourier transform, $G_k(u)$, satisfies the Fourier domain constraints. Consequently, for the input-output approach the mean-squared error, E_F^2, is unimportant; E_o^2 is the meaningful quality criterion.

Another interesting property of the system shown in Figure 2 is that if an output g'(x) is used as an input, then its output will be itself. Since the Fourier transform of g'(x) already satisfies the Fourier-domain constraints, g'(x) is unaffected as it goes through the system. Therefore, no matter what input actually resulted in the output g'(x), the output g'(x) can always be considered to have resulted from itself as an input. From this point of view, another logical choice of the next input is

$$g_{k+1}(x) = g'_k(x) + \beta \Delta g_d(x) \quad . \tag{9}$$

Note that if $\beta = 1$ in Eq. (9), then the input-output approach reduces to the error-reduction approach. Since the optimum value of β is usually not unity, the error-reduction approach can be looked on as a suboptimal subset of one version of the more general input-output approach. Depending on the problem being solved, other variations on Eqs. (8) and (9) may be successful ways for choosing the next input.

COMPUTER HOLOGRAPHY PROBLEMS

Reduction of quantization noise

The objective of computer holography[10] is to synthesize a transparency that can modulate a wavefront according to a calculated wavefront, usually corresponding to Fourier coefficients (or samples of the Fourier transform of an image) computed from the discrete Fourier transform. Let $F = \mathcal{F}[f]$ be the desired wavefront modulation and f be the complex-valued function describing the desired image. Due to the limitations of the display devices and materials used to synthesize computer holograms, it is often not possible to exactly represent any arbitrary complex Fourier coefficient. For example as illustrated in Figure 3(a), Lohmann's binary detour-phase hologram[11] can represent only a discrete set of complex values, depending upon the number of resolution elements of the display device used to form one cell to represent a Fourier coefficient. The kinoform[12] allows nearly continuous phase control, but quantizes all the magnitudes to a single level, as illustrated in Figure 3(b). (If a gray-level display device, or recorder, used to synthesize a kinoform has a finite number of gray levels, then the phase is quantized as well.) In either case, the desired complex coefficient F is only approximated by a quantized value Q = F + N, where N is

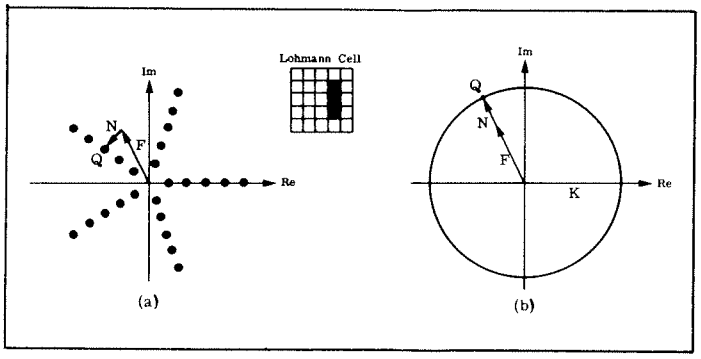

Figure 3. (a) Lohmann binary detour-phase hologram. Magnitude and phase are determined by the area and position, respectively, of an aperture within the cell. For 5×5 subcells per Fourier coefficient, only 26 points in the complex plane can be addressed. (b) Kinoform. Phase is determined by the thickness of the film, and the magnitude is quantized to a single level.

the Fourier domain (hologram plane) quantization noise. The resulting image is $f' = \mathcal{F}^{-1}[Q] = \mathcal{F}^{-1}[F] + \mathcal{F}^{-1}[N] = f + n$, where n is the image-domain quantization noise. Quantization noise is described in more detail in Ref. 13 and 14.

Since only the squared magnitude (the intensity) of the image is observed, we are free to choose the phase of the object (phase code the object) in such a way as to reduce the variance (dynamic range) of $|F|$, which reduces the quantization noise in kinoforms and, to a lesser extent, in the Lohmann hologram. Random phase and various deterministic phase codes[15] cause a considerable reduction in the variance of $|F|$, but substantial errors remain.

This problem of phase coding to reduce quantization noise fits the problem statement in the introduction: it is a synthesis problem in which the Fourier domain constraint is that the values of the Fourier transform F fall on a set of prescribed quantized values, and the magnitudes of the image f' equal that of the desired image at each point. In fact, the nonlinear system shown in Figure 2 can represent a system for making quantized computer-generated holograms, where the input g is the digital description of the ideal image, the operation of satisfying Fourier-domain constraints is the fabrication of a quantized hologram, and the output g' is the image produced by the quantized hologram. The error-reduction approach was used for kinoforms by Hirsch et al.[2] and by Gallagher and Liu[3] in order to reduce quantization noise to much lower levels than what is ordinarily achieved by random or deterministic phase codes. It was for this problem that the input-output approach was first developed[7] in order to improve on the convergence properties of the error-reduction approach.

To gain a better understanding of how the input-output concept applies to this problem and how we arrived at Eq. (6), in the Appendix we consider the kinoform case in some detail. A more rigorous proof of Eq. (6) is available elsewhere.[8]

For the computer holography quantization problems, the Fourier-domain constraint is that the values of the Fourier transform fall on allowed quantized values; the object-domain constraint is that the magnitude of the image equal a desired magnitude, $|f(x)|$. Since any image phase is allowable, there is an infinity of changes of the output that would cause it to satisfy the constraint. A logical choice of the desired change of the ouput is the smallest change $\Delta g_d(x)$ such that $g'(x) + \Delta g_d(x) = |f(x)|$. That would be $\Delta g_d(x) = |f(x)|g'(x)/|g'(x)| - g'(x)$. We also noticed that the phase difference between $g'(x)$ and $g(x)$ tends to have the same sign as the change of phase of $g'(x)$ on successive iterations. Therefore, it is desirable to choose a $\Delta g_d(x)$ that tends to rotate the phase angle of the new input toward that of the last output. For these reasons, a good choice for the desired change of the output is

$$\Delta g_d(x) = \left[|f(x)| \frac{g'(x)}{|g'(x)|} - g'(x) \right] + \left[|f(x)| \frac{g'(x)}{|g'(x)|} - |f(x)| \frac{g(x)}{|g(x)|} \right] \quad (10)$$

in which the first component boosts (or shrinks) the magnitude of the output to the desired level, and the second component rotates the phase angle toward the angle of the output. The next input is then given by inserting Eq. (10) into Eq. (8) or (9). Other algorithms for choosing $\Delta g_d(x)$ were also found to be successful.[8]

The iterative method was tested using a binary (= 0 or 1) image of the block letters SU. The first example is for a hologram with four magnitude and four phase quantization levels, plus the zero level, as would be the case for a Lohmann hologram (see Figure 3(a)) using only 4 x 4 subcells to represent a Fourier coefficient. The object was random phase coded and Fourier transformed. The Fourier transform was quantized, and the inverse transform was computed, resulting in the sampled image shown in Figure 4(a). After 13 iterations of the input-output approach using Eqs. (10) and (8) with $\beta = 1$, the greatly improved image shown in Figure 4(b) was obtained. (Grayscales shown below the images are for calibration purposes.)

A quality criterion pertinent to the optical memory application is the ratio of the intensity of the weakest "one bit" (image sample with ideal intensity equal to unity) to the strongest "zero bit" (image sample with ideal intensity equal to zero). Figure 5 shows a plot of the range of output image intensities $|g'(x)|^2$ for one bits and for zero bits (that is, the maximum and minimum one bit and the maximum zero bit) as a function of the number of iterations. Initially there was very little difference between the weakest one bit and the strongest zero bit, indicating a relatively high error rate; however, after a few iterations, there is a comfortable gap between the weakest one bit and the strongest zero bit, despite the severe quantization involved in the hologram.

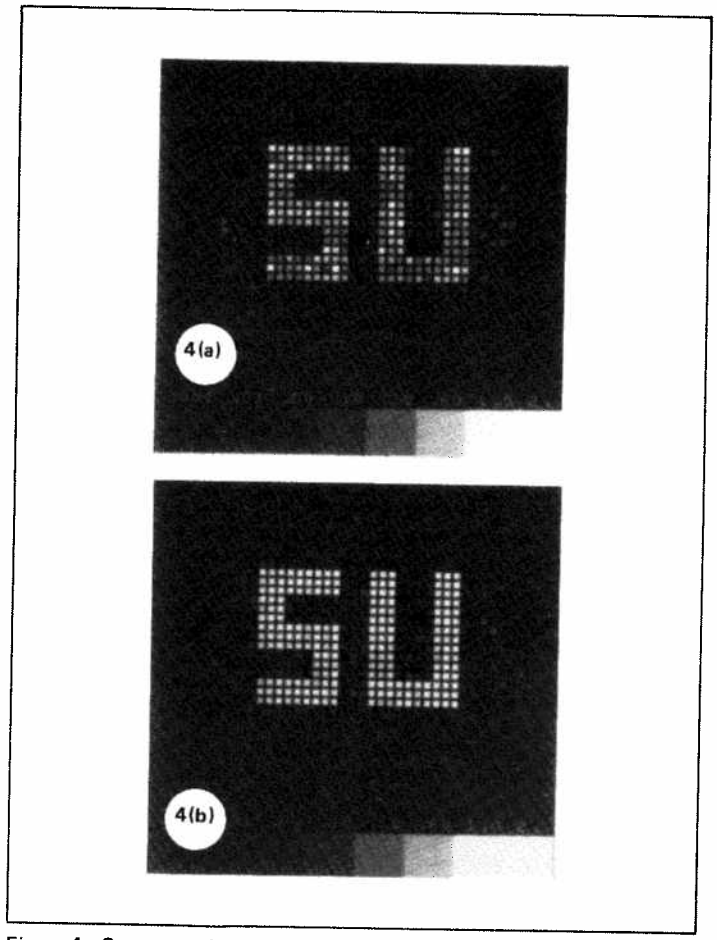

Figure 4. Computer-simulated images from hologram with 4 magnitude and 4 phase quantized levels. (a) object random phase coded; (b) after 13 iterations of the iterative method.

The second example is for a kinoform, which has continuously controlled phase but only one magnitude level. Figure 6(a) shows the resulting output image when the input image was random phase coded. Figure 6(b) shows the output image after 8 iterations of the error-reduction approach, using $g_{k+1}(x) = |f(x)|g_k(x)/|g_k(x)|$; and Figure 6(c) shows the output image after 8 iterations of the input-output approach using Eqs. (10) and (8) with $\beta = 1$. Despite the severe magnitude quantization, the image is greatly improved by both approaches. Figure 7 shows the range of output intensities for the one bits and zero bits for both cases; the greatest error of the intensity of the one bits is considerably less when the input-output approach is used. When judged by the mean-squared error, the results in the two cases were comparable.

Spectrum shaping

Spectrum shaping is a synthesis problem that can be stated as follows: given the magnitude $|f(x)|$ of a complex-valued object, $g(x) = |f(x)| \exp[i\theta(x)]$, find a phase function $\theta(x)$ such that $|\mathcal{F}[g(x)]|$ is equal to a given spectrum $|F(u)|$. The problem of reducing quantization noise for kinoforms, discussed in the previous section, is a special case of spectrum

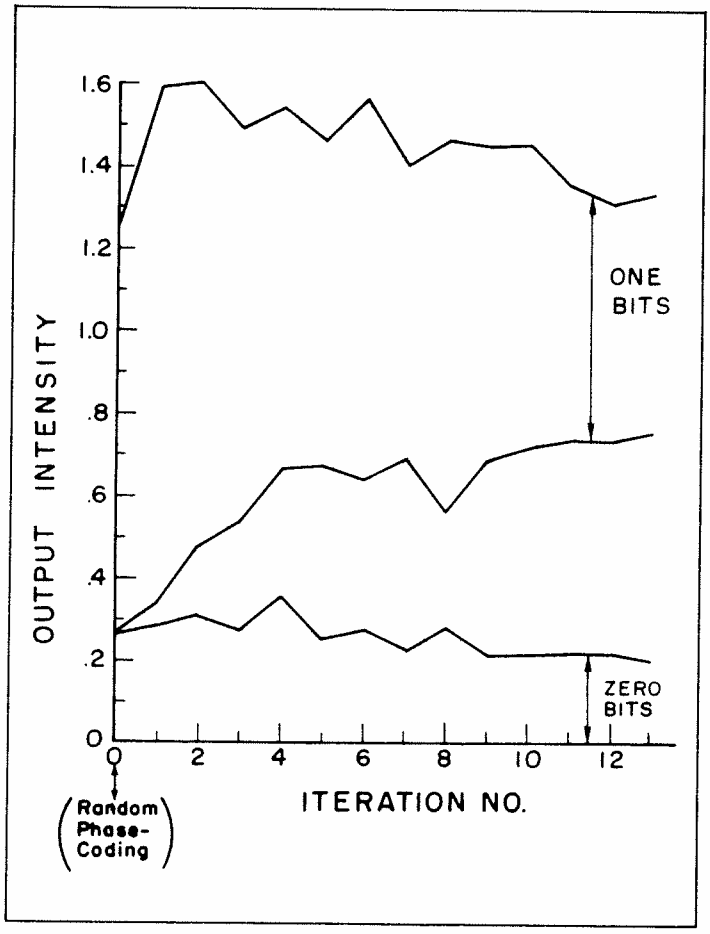

Figure 5. Range of output intensities vs number of iterations, for hologram with 4 magnitude and 4 phase quantized levels.

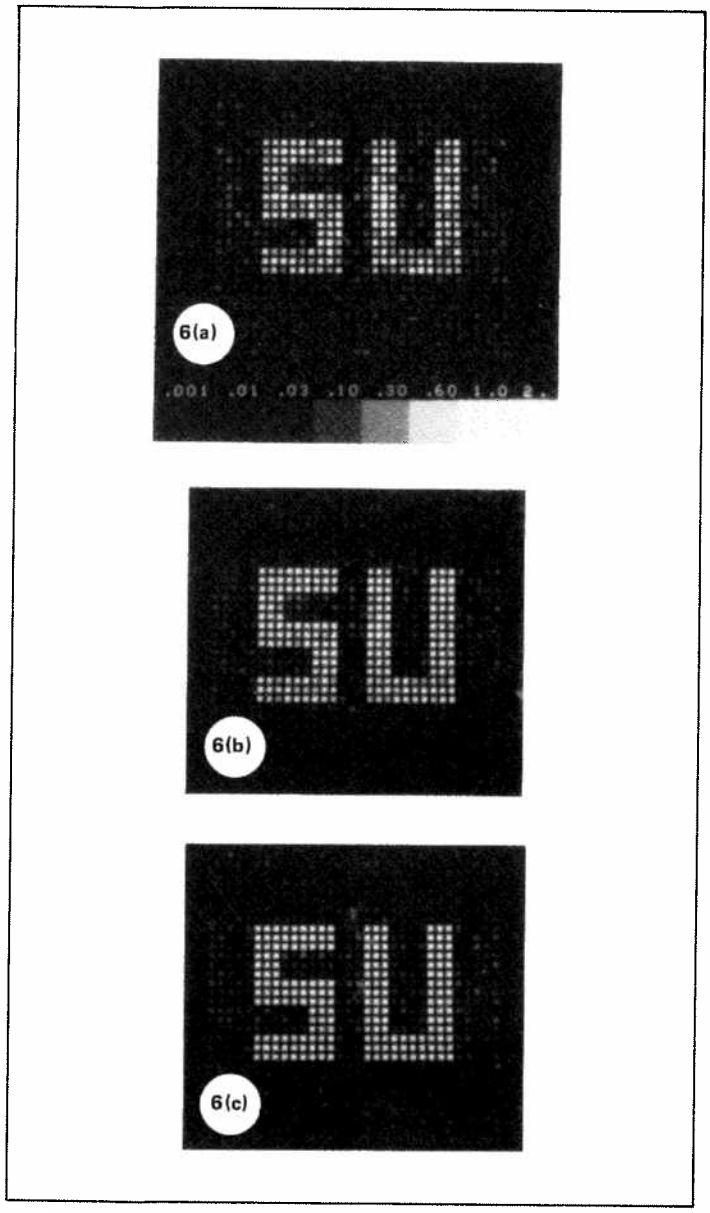

Figure 6. Computer-simulated images from kinoform. (a) object random phase coded; (b) after 8 iterations of the error-reduction approach; (c) after 8 iterations of the input-output approach.

shaping for which $|F(u)|$ is a constant. A more complex problem is one suggested by the Escher engraving shown in Figure 8, in which a bird transforms into a fish. We wish to find a function with magnitude being a picture of a fish, which has a Fourier transform of magnitude being a picture of a bird. Or, in terms of computer holography, find a phase function to assign to the image of a fish so that the hologram will look like an image of a bird. Figure 9(a) shows the actual "bird" and "fish" binary patterns used for our experiment. For the first iteration, the fish object was random phase coded, Fourier transformed, and the magnitude of the Fourier transform was replaced with the magnitude of the bird pattern shown in Figure 9(a). The result was inverse Fourier transformed, yielding the very noisy output image shown in Figure 9(b). The input-output approach was then used for seven iterations, resulting in the improved image shown in Figure 9(c). For this as well as for the examples shown earlier, increasing the number of iterations resulted in a further improvement of the quality of the image.

IMAGE RECONSTRUCTION FROM INTERFEROMETER DATA

For telescopes operating at optical wavelengths, atmospheric turbulence limits the resolution of astronomical objects to one second of arc or worse, although the theoretical diffraction limit is fifty times as fine for the largest telescopes. Despite atmospheric turbulence, it is possible to measure the modulus of the Fourier transform of a space object out to the diffraction limit of the telescope using interferometric techniques.[16–19] The autocorrelation of the object can be computed from the Fourier modulus, allowing the diameter of the object to be determined. However, unless the Fourier transform phase is also measured, it has not been possible to determine the object itself, except for some special cases. Previous attempts to solve this problem[20] have not proven to be practical for complicated two-dimensional objects.

The problem of reconstructing an object from interferometer data can be sovled by the iterative method. The Fourier-domain constraint is that the Fourier modulus equal the Fourier modulus measured by an interferometer, and the object-domain constraint is that the object function be real-valued and nonnegative. Where the output image satisfies the constraints, $\Delta g_d(x) = 0$. Where it violates the constraints (where it is negative or where its extent exceeds the diameter of the object as determined from its autocorrelation), it can be made to satisfy the constraints by having it become equal to zero, and so $\Delta g_d(x) = -g'(x)$. For the error-reduction approach, the next input would be given by $g_{k+1}(x) = g'_k(x) + \Delta g_d(x)$. For the input-output approach, the next input would be given by Eq. (8) or Eq. (9). We found that a particularly successful method of choosing the next input is to use Eq. (9) where the output satisfies the constraints and Eq. (8) where it violates the constaints. In experiments using computer-simulated data, we found that the error-reduction approach decreased the mean-squared error rapidly for the first few iterations, but extremely slowly for later iterations. Much faster convergence was obtained using the input-output approach or by alternating between the two approaches every few iterations.

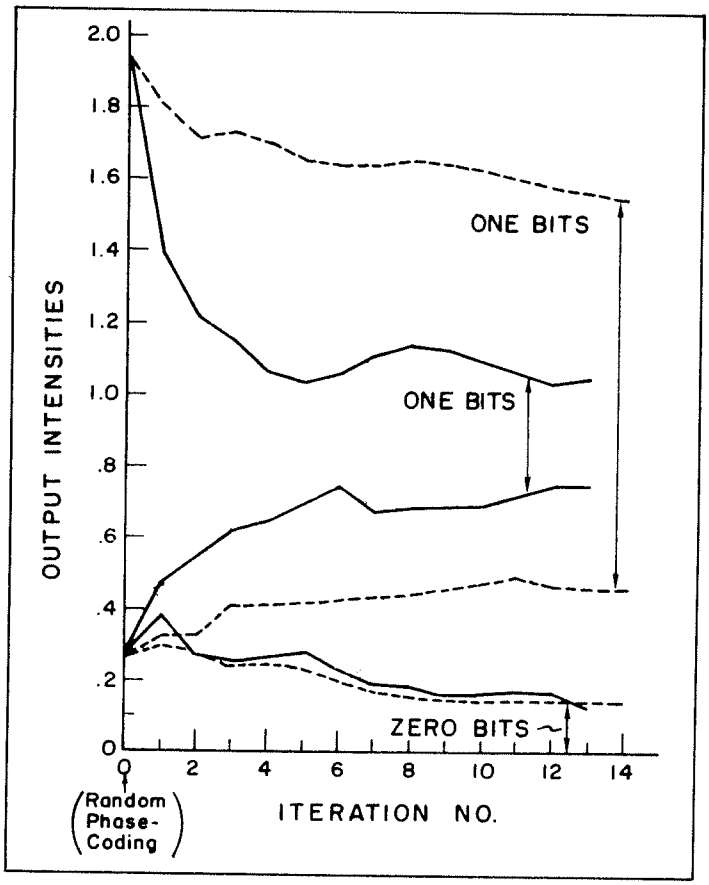

Figure 7. Range of output intensities vs number of iterations for kinoform: dashed lines for the error-reduction approach, solid lines for the input-output approach.

Figure 10(a) shows a computer-synthesized object used for the experiment[9]—a sun-like disk having "solar flares" and bright and dark "sunspots." The modulus of its Fourier transform is shown in Figure 10(b). Figure 10(c) shows a square of random numbers used as the initial input for the iterative method. Figures 10(d), (e) and (f) show the reconstruction results after 20, 230, and 600 iterations, respectively. Figure 10(g) shows the initial input for a second trial, and the reconstruction results after 2 and 215 iterations are shown in Figures 10(h) and (i), respectively. Comparing Figures 10(f) and 10(i) with the original object in Figure 10(a), we see that for both trials, the reconstructed images match the original object very closely. Note that inverted solutions such as Figure 10(f) are permitted for this problem since $|F(u)| = |\mathcal{F}[f(x)]| = |\mathcal{F}[f(-x)]|$. Other examples of reconstruction experiments, including blind tests using data simulated to have the noise that would be present in an actual speckle interferometer, are described in Ref. 21. These results with computer-simulated data indicate that, using the iterative method with speckle interferometer data, it should be possible to reconstruct images having resolution many times finer than what is ordinarily allowed by the turbulent atmosphere.

CONCLUSION

In this paper, we have described an iterative method for solving a number of diverse problems in optics and related fields. Experimental results were shown for synthesis problems and for a reconstruction problem, each having a different set of constraints on the solution. The method can also be applied to a number of other problems. One version of the iterative method, the error-reduction approach, which is a simple modification of the Gerchberg-Saxton algorithm, was found to be successful for some applications but not for others. The more powerful input-output approach was found to converge faster and make the iterative method practical for a wider range of problems. The iterative method promises to be a very valuable tool for the field of optics.

APPENDIX

Suppose that the input g(x) to a kinoform system (Figure 3(b)) results in the output g′(x). The kinoform has a transmittance $Q(u) = K \exp[i\phi(u)]$ where $\phi(u)$ is the phase of $G(u) = |G(u)| \exp[i\phi(u)] = \mathcal{F}[g(x)]$, and K is a constant. The resulting image is $g'(x) = \mathcal{F}^{-1}[Q(u)]$. Now consider what happens when a change $\Delta g(x)$ is made in the input. As illustrated in the phasor diagrams in Figure 11, the change $\Delta g(x)$ in the input causes a change $\Delta G(u)$ in its Fourier transform, which causes a change $\Delta Q(u)$ in the kinoform and a corresponding change $\Delta g'(x) = \mathcal{F}^{-1}[\Delta Q(u)]$ in the output image. Our goal here is to determine the relationship between the change $\Delta g'(x)$ of the output and the change $\Delta g(x)$ of the input. Figure 12 shows the relationship between $\Delta Q(u)$ and two orthogonal components of $\Delta G(u)$. By similar triangles, we have, for $|\Delta G| \ll |G|$,

$$\Delta Q(u) \simeq \Delta G^t(u)\left[\frac{K}{|G(u)|}\right] \tag{A1}$$

where the two orthogonal components of $\Delta G(u)$ are

$$\Delta G^r(u) = |\Delta G(u)| \cos\beta(u)\, e^{i\phi(u)} \tag{A2}$$

parallel to G(u), and

$$\Delta G^t(u) = |\Delta G(u)| \sin\beta(u)\, e^{i[\phi(u)+\pi/2]}; \tag{A3}$$

orthogonal to G(u); and

$$\Delta G(u) = \Delta G^r(u) + \Delta G^t(u) = |\Delta G(u)|\, e^{i[\phi(u)+\beta(u)]} \tag{A4}$$

where $\beta(u)$ is the angle between $\Delta G(u)$ and $G(u)$. Only one of the two orthogonal components of $\Delta G(u)$, namely $\Delta G^t(u)$, contributes to $\Delta Q(u)$.

In order to compute the expected change of the output, $E[\Delta g'(x)]$, we treat the phase angles $\beta(u)$ and the magnitudes $|G(u)|$ as random variables. Inserting $|\Delta G(u)|$ from Eq. (A4) into Eq. (A3), we have

$$\begin{aligned}\Delta G^t(u) &= \Delta G(u)\, e^{-i[\phi(u)+\beta(u)]} \cos\beta(u)\, e^{i\phi(u)} \\ &= \Delta G(u)[\sin^2\beta(u) + i \sin\beta(u)\cos\beta(u)] \quad .\end{aligned} \tag{A5}$$

For $\beta(u)$ uniformly distributed over $(0, 2\pi)$,[14] the expected value of $\Delta G^t(u)$ is

$$E[\Delta G^t(u)] = \Delta G(u)\left[\frac{1}{2} + i\cdot 0\right] = \frac{1}{2}\Delta G(u) \quad . \tag{A6}$$

Therefore, the expected value of the change of the output is, using Eqs. (A1) and (A6) and assuming that the magnitudes $|G(u)|$ are identically distributed random variables[14] independent of $\beta(u)$,

$$\begin{aligned}E[\Delta g'(x)] &= E[\mathcal{F}(\Delta Q)] = \mathcal{F}[E(\Delta Q)] \\ &\simeq \mathcal{F}\left[E(\Delta G^t)\cdot E\left(\frac{K}{|G|}\right)\right] \\ &\simeq \mathcal{F}\left[\frac{1}{2}\Delta G(u)\right] \cdot E\left(\frac{K}{|G|}\right) \\ &= \frac{1}{2}\Delta g(x) E\left(\frac{K}{|G|}\right) \quad .\end{aligned} \tag{A7}$$

That is, the expected change of the output is α times the change of the input giving us the second term of Eq. (6), where $\alpha = (\frac{1}{2})E(K/|G|)$. After a few iterations, $|G(u)|$ will not differ greatly from K; then $\alpha \simeq \frac{1}{2}$.

Similarly, the variance of the change of the output can be shown to be[8]

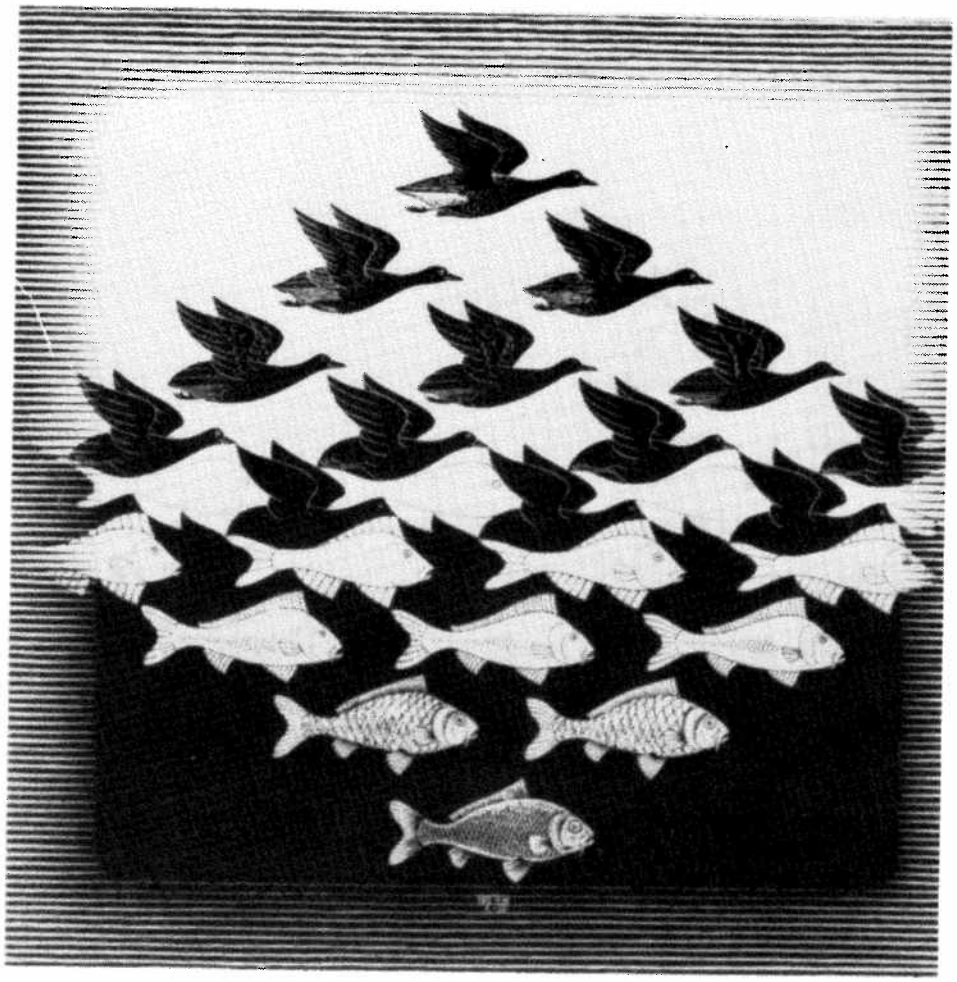

Figure 8. Bird transforms into fish ("Sky and Water" by M. C. Escher). This reproduction was authorized by the M. C. Escher Foundation, The Hague, Holland/G. W. Breughel.

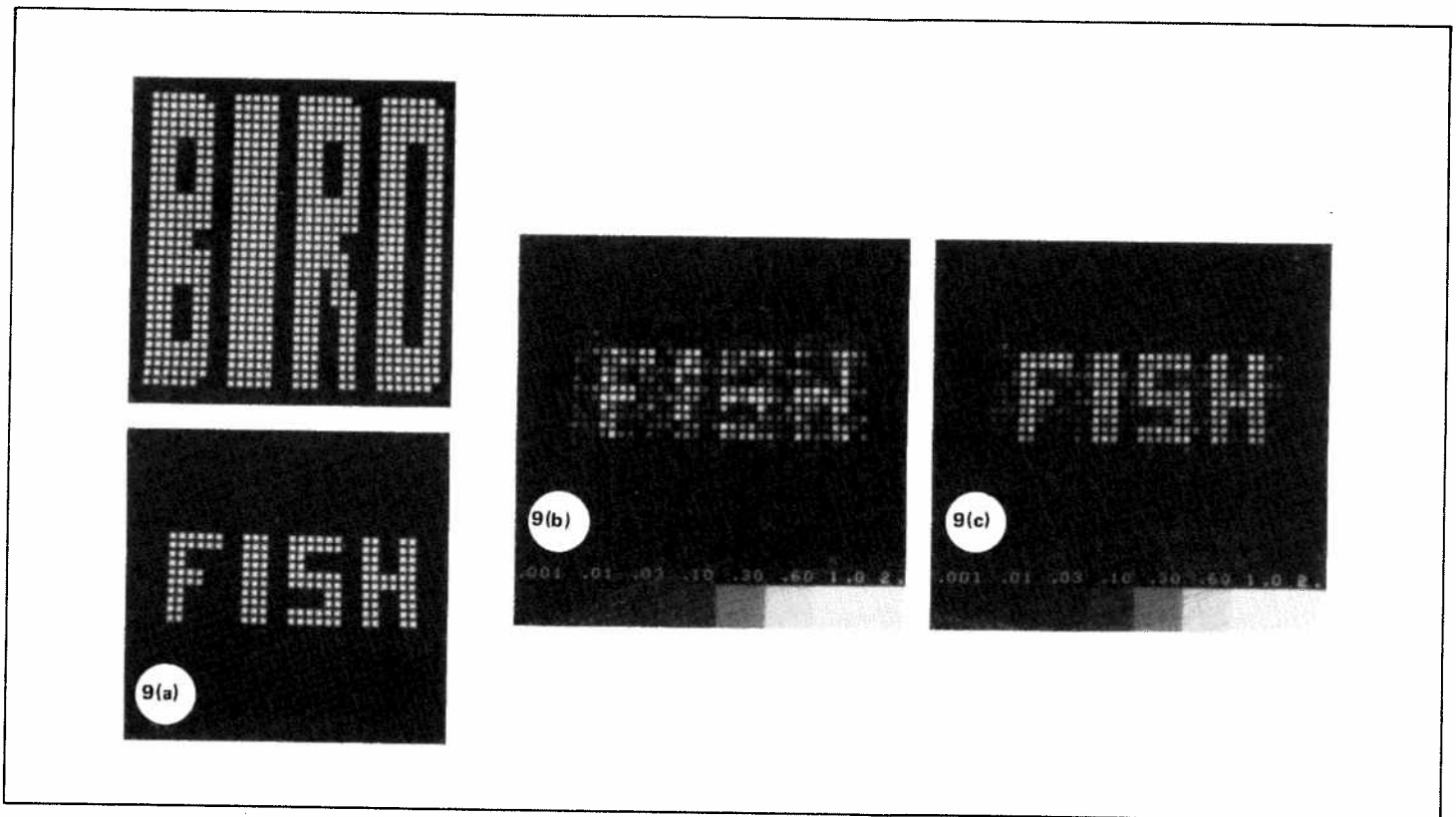

Figure 9. (a) Bird hologram and desired fish image; (b) fish output image after random phase coding of input; (c) output image after seven iterations.

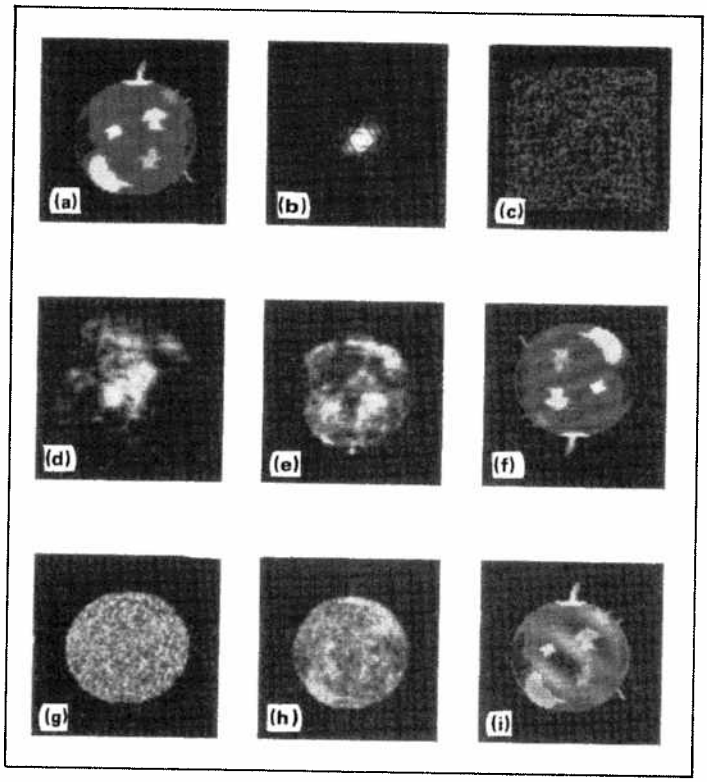

Figure 10. (a) Test object; (b) modulus of its Fourier transform; (c) initial estimate of the object (first test); (d)-(f) reconstruction results—number of iterations: (d) 20, (e) 230, (f) 600; (g) initial estimate of the object (second test); (h)-(i) reconstruction results—number of iterations: (h) 2, (i) 215.

$$E[|\Delta g'(x)|^2] - |E[\Delta g'(x)]|^2$$

$$\simeq \frac{1}{4}\left\{ 2E\left(\frac{K^2}{|G|^2}\right) - \left[E\left(\frac{K}{|G|}\right)\right]^2 \right\} \cdot \frac{1}{A}\int_{-\infty}^{\infty} |\Delta g(x)|^2 \, dx \qquad \text{(A8)}$$

where A is the area of the image. That is, the variance of the change of the output $\Delta g'(x)$ at any given x is proportional to the integrated squared change of the entire input. The predictability of $\Delta g'(x)$, and the degree of control with which we can manipulate it, decreases as we make larger changes in the input. The difference between the actual change of the output and the expected change of the output given by Eq. (A7) is what is meant by the additional noise term in Eq. (6). If after a few iterations, $|G(u)| \simeq K$, then in Eq. (A8), the factor $(1/4)\{2E(K^2/|G|^2) - [E(K/|G|)]^2\} \simeq 1/4$.

Equations (A7) and (A8) are a justification for the input-output concept; small changes of the input result in similar changes of the

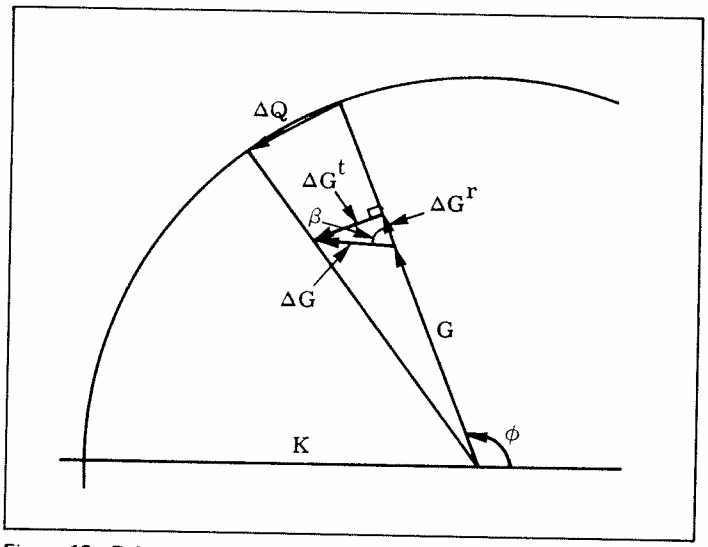

Figure 12. Relationship between ΔQ, the change of the kinoform, and two components of ΔG, the Fourier transform of the change of the input.

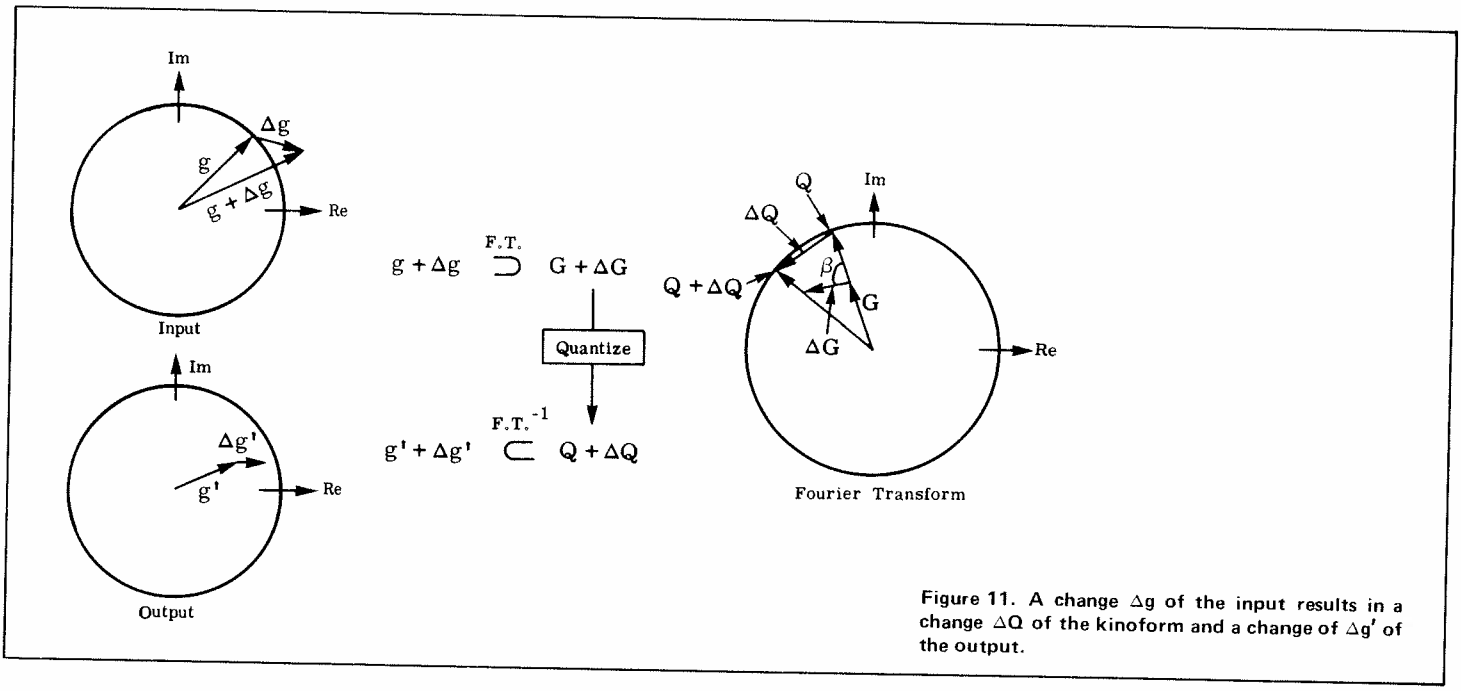

Figure 11. A change Δg of the input results in a change ΔQ of the kinoform and a change of Δg' of the output.

output, and so the output can be driven to satisfy the constraints by appropriate changes of the input, as in Eq. (8) or (9).

REFERENCES

1. Gerchberg, R. W. and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," Optik **35,** 237-46 (1972).
2. Hirsch, P. M., J. A. Jordan, Jr., and L. B. Lesem, "Method of Making an Object-Dependent Diffuser," U.S. Patent No. 3,619,022 (Nov. 9, 1971; filed Sept. 17, 1970).
3. Gallagher, N. C. and B. Liu, "Method for Computing Kinoforms that Reduces Image Reconstruction Error," Appl. Opt. **12,** 2328-35 (1973).
4. Gerchberg, R. W., "Super-Resolution Through Error Energy Reduction," Optica Acta **21,** 709-20 (1974); A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," IEEE Trans. Circuits and Systems CAS-**22,** 735-42 (1975).
5. Saxton, W. O., *Computer Techniques for Image Processing in Electron Microscopy* (Academic Press, 1978); D. L. Misell, "A Method for the Solution of the Phase Problem in Electron Miscroscopy," J. Phys. D: Appl. Phys. **6,** L6-L9 (1973); A. M. Huiser, A. J. J. Drenth, and H. A. Ferwerda, "On Phase Retrieval in Electron Microscopy from Image and Diffraction Pattern," Optik **45,** 303-16 (1976); A. M. J. Huiser, P. Van Toorn, and H. A. Ferwerda, "On the Problem of Phase Retrieval in Electron Microscopy from Image and Diffraction Pattern I-IV," Optik **47,** 123-34 (1977); R. A. Gonsalves, "Phase Retrieval from Modulus Data," J. Opt. Soc. Am. **66,** 961-64 (1976).
6. Liu, B. and N. C. Gallagher, "Convergence of a Spectrum Shaping Algorithm," Appl. Opt. **13,** 2470-71 (1974).
7. Fienup, J. R., "Reduction of Quantization Noise in Kinoforms and Computer-Generated Holograms," J. Opt. Soc. Am. **64,** 1395 (1974) (Abstract).
8. Fienup, J. R., "Improved Synthesis and Computational Methods for Computer-Generated Holograms," Ph.D. thesis, Stanford University, May 1975 (University Microfilms No. 75-25523), Chapter 5.
9. Fienup, J. R., "Reconstruction of an Object from the Modulus of Its Fourier Transform," Opt. Lett. **3,** 27-29 (1978).
10. For reviews of computer holography, see T. S. Huang, "Digital Holography," Proc. IEEE **59,** 1335-46 (1971); R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography* (Academic Press, New York, 1971), pp. 542-563; D. C. Chu and J. R. Fienup, "Recent Approaches to Computer-Generated Holograms," Opt. Eng. **13,** 189-95 (1974); W.-H. Lee, "Computer-Generated Holograms: Techniques and Applications," in E. Wolf, ed., *Progress in Optics,* Vol. **16** (North-Holland, 1978), pp 121-232.
11. Brown, B. R. and A. W. Lohmann, "Computer-Generated Binary Holograms," IBJ J. Res. Develop. **13,** 160-68 (1969); A. W. Lohmann and D. P. Paris, "Binary Fraunhofer Holograms, Generated by Computer," Appl. Opt. **6,** 1739-48 (1967).
12. Lesem, L. B., P. M. Hirsch, and J. A. Jordan, Jr., "The Kinoform; A New Wavefront Reconstruction Device," IBM J. Res. Develop. **13,** 150-55 (1969).
13. Goodman, J. W. and A. M. Silvestri, "Some Effects of Fourier-Domain Phase Quantization," IBM J. Res. Develop. **14,** 478-84 (1970); R. A. Gabel and B. Liu, "Minimization of Reconstruction Errors with Computer-Generated Binary Holograms," Appl. Opt. **9,** 1180-90 (1970).
14. Powers, R. S. and J. W. Goodman, "Error Rates in Computer-Generated Holographic Memories," Appl. Opt. **14,** 1690-1701 (1975).
15. Akahori, H., "Comparison of Deterministic Phase Coding with Random Phase Coding in Terms of Dynamic Range," Appl. Opt. **12,** 2336-43 (1973).
16. Michelson, A. A. and F. G. Pease, "Measurement of the Diameter of Alpha Orionis with the Interferometer," Astrophys. J. **53,** 249-59 (1921).
17. Hanbury Brown, R. and R. Q. Twiss, "Correlation Between Photons in Two Coherent Beams of Light," Nature **177,** 27-29 (1956).
18. Labeyrie, A., "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images," Astron. & Astrophys. **6,** 85-87 (1970); Gezari, D. Y., A. Labeyrie, and R. V. Stachnik, "Speckle Interferometry: Diffraction-Limited Measurements of Nine Stars with the 200-inch Telescope," Astrophys. J. Lett. **173,** L1-L5 (1972).
19. Currie, D. G., S. L. Knapp, and K. M. Liewer, "Four Stellar-Diameter Measurements by a New Technique: Amplitude Interferometry," Astrophys J. **187,** 131-44 (1974).
20. See, for example, A. Walther, "The Question of Phase Retrieval in Optics," Optica Acta **10,** 41-49 (1963); D. Kohler and L. Mandel, "Source Reconstruction from the Modulus of the Correlation Function: A Practical Approach to the Phase Problem of Optical Coherence Theory," J. Opt. Soc. Am. **63,** 126-134 (1973); B. R. Frieden and D. G. Currie, "On Unfolding the Autocorrelation Function," J. Opt. Soc. Am. **66,** 1111 (1976) (Abstract).
21. Fienup, J. R., "Space Object Imaging through the Turbulent Atmosphere," Opt. Eng. **18,** 529-34 (Sept.-Oct., 1979).