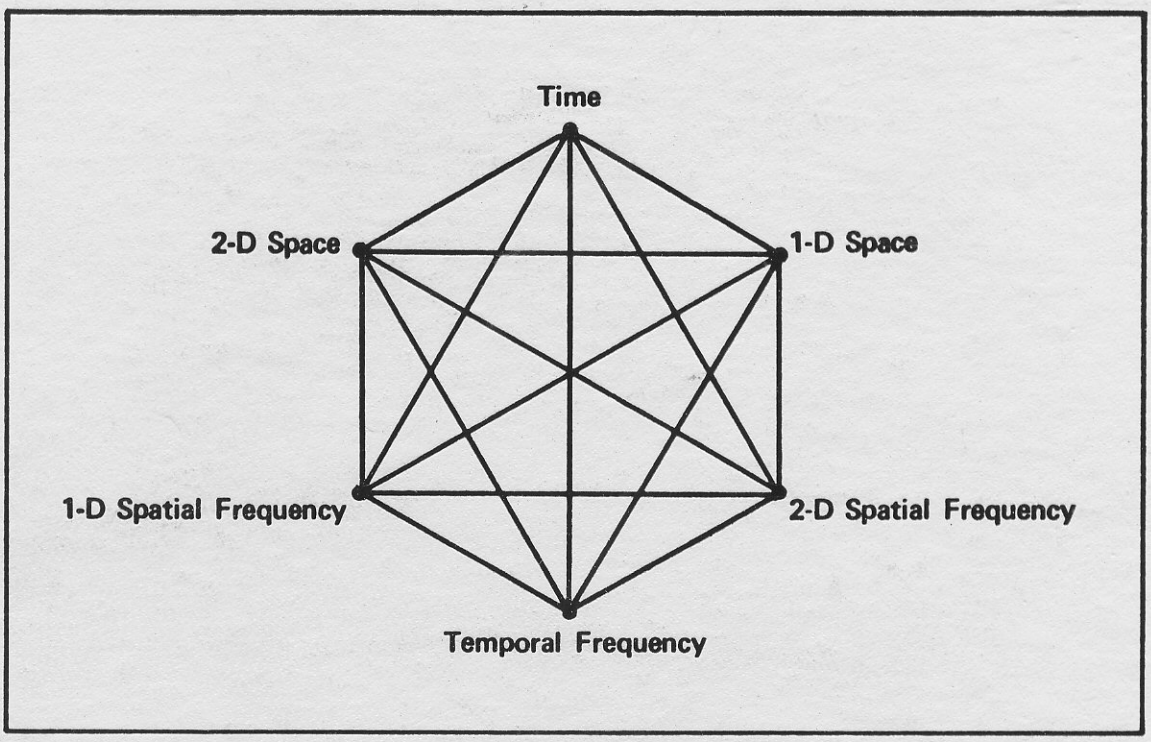
Proceedings of the SPIE Advanced Institute on

Transformations in Optical Signal Processing

SPIE Volume 373



Editors:
William T. Rhodes
James R. Fienup
Bahaa E. A. Saleh

23–25 February 1981
Battelle Northwest Conference Facilities
Seattle, Washington USA

Gerchberg-type linear deconvolution and extrapolation algorithms

Robert J. Marks, II
David K. Smith
University of Washington
Department of Electrical Engineering
Seattle, Washington 98195

Abstract. Gerchberg's iterative extrapolation algorithm is generalized to two dimensions in two distinct ways. The first generalization is implemented on a coherent optical processor. Fundamental limitations are discussed. A second generalization is reformulated discretely and placed in closed form. A number of digital implementations are presented. A generalized methodology is then developed for a certain class of deconvolution problems. Gerchberg's algorithm and other deconvolution algorithms are shown to be special cases. Algorithm convergence and stability (posedness) are discussed and exemplified. Last, methods of incorporating further object information into the iterative algorithms are explored.

1. INTRODUCTION

With knowledge of an image in a given function class only over a portion of its domain, the extrapolation problem is to determine or estimate the image over the rest of its domain. An example of such a function class is that of analytic signals. With knowledge of the signal over an arbitrarily small interval one can, in principle, generate the signal everywhere by a Taylor series or some other method of analytic continuation.

Much attention has been focused on extrapolation of finite energy bandlimited images. Indeed, all bandlimited functions are analytic everywhere.[1,2] As is reviewed in Appendix A, Slepian and Pollak[3] formulated a classic extrapolation algorithm involving image expansion in terms of prolate spheroidal wave functions.[4-9] Clever statistical approaches to the same problem were developed by Frieden.[10-13] A least squares approach has recently been proposed by Howard.[14-15] A comparison of some extrapolation algorithms is given by Rushford and Frost.[16]

Herein, we will deal with an extrapolation algorithm developed by Gerchberg.[17-20] Gerchberg's algorithm, in iterative form, involves only the elementary operations of Fourier transformation and truncation. The algorithm is generalized to two dimensions in two distinct ways. Implementation of one of the extensions on a coherent optical processor with passive feedback is discussed. The second extension, reformulated discretely, is placed in closed form. The algorithms can also be easily adapted to perform interpolations.

Extrapolation in the frequency domain is referred to as super resolution. Super resolution is a special case of deconvolution. We demonstrate that Gerchberg's algorithm is simply a special case of a large class of deconvolution algorithms. Other special cases include Cadzow's extrapolation algorithm[21,22] and Van Cittert's deconvolution algorithm.[23-25]

2. GERCHBERG'S EXTRAPOLATION ALGORITHM

2.1. Background

A finite energy signal f(x) is herein defined to be bandlimited with bandwidth interval Ω if it conforms to the following three conditions:

$$f(x) = \int_{\Omega} F(u)\exp(j2\pi ux)du \; ; \tag{1a}$$

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |F(u)|^2 du < \infty \; ; \tag{1b}$$

$$|\Omega| \equiv \int_{-\infty}^{\infty} W_{\Omega}(u)du < \infty \; ; \tag{1c}$$

where F(u) is the signal's Fourier transform,

$$F(u) = \mathscr{F}[f(x)] = \int_{-\infty}^{\infty} f(x)\exp(-j2\pi ux)dx \; ,$$

and $W_{\Omega}(u)$ is the gate window corresponding to the bandwidth in-

terval:

$$W_\Omega(u) = \begin{cases} 1 \; ; \; u \in \Omega \\ 0 \; ; \; u \notin \Omega \; . \end{cases}$$

Let T correspond to an interval on x, and let

$$g_T(x) = \begin{cases} f(x) \; ; \; x \in T \\ 0 \quad ; \; x \notin T \; . \end{cases}$$

The extrapolation problem is to determine f(x) with knowledge of $g_T(x)$ and Ω. If T corresponds to the complement of a finite interval, we have an interpolation problem. The development to follow is applicable for any T.

Gerchberg's algorithm, illustrated in Fig. 1, is a specific application of the iterative imposition of both frequency and spatial domain constraints, as discussed by Fienup in the preceding chapter. Beginning with $g_T(x)$, we first perform a Fourier transform. The frequency constraint is satisfied by keeping only those frequency components within the Ω interval in step 2. Step 3 is a simple inverse Fourier transformation. To satisfy spatial constraints, we replace this function in the T interval by $g_T(x)$ in steps 4 and 5. This first estimate is Fourier transformed and the cycle is repeated. In the absence of noise, convergence of the N^{th} estimate of f(x), $f_N(x)$, as $N \to \infty$ has been proven in three distinct ways.[17-20]

If we define the bandlimiting operator

$$\mathcal{B} = \mathcal{F}^{-1} W_\Omega(u) \mathcal{F} \; ,$$

then the iterative algorithm can be written as

$$f_N(x) = g_T(x) + [1 - r_T(x)] \mathcal{B}_\Omega f_{N-1}(x) \; , \qquad (2)$$

where the spatial rectangular window is

$$r_T(x) = \begin{cases} 1 \; ; \; x \in T \\ 0 \; ; \; x \notin T \; . \end{cases}$$

Motivated by Eq. (2), a condensed illustration of the algorithm is shown in Fig. 2. Note that $g_T(x)$ is block orthogonal to the N^{th} tail estimate $[1-r_T(x)]f_N(x)$ (i.e., the product of the two functions is identically zero). This observation allows an enlightening geometric view of the iterative algorithm in a Hilbert space setting.[19] In Fig. 3, we illustrate three subspaces as three lines in a planar L_2 function space. The horizontal line contains all L_2 functions identically zero outside the interval $x \in T$. The vertical line contains all L_2 functions identically zero *within* the interval $x \in T$. These two function classes are block orthogonal and are thus drawn perpendicular. The third function class, B_Ω, consists of all L_2 functions with bandwidth interval Ω. The bandlimited function we desire as an extrapolation (interpolation) result—henceforth referred to as the target function f(x)—lies somewhere in this space. The information we have, $g_T(x)$, is the projection of f(x) onto the space of T interval functions.

To regain f(t) from $g_T(x)$ we first project $g_T(x)$ onto B_Ω. With reference to Eq. (2), this, in turn, is projected onto the space of images identically zero within T. This vector is added to the $g_T(x)$ vector to arrive at the first estimate, $f_1(x)$, of the target function. The process is then repeated and the estimate iteratively converges to f(x).

2.2. An interative optical implementation

A coherent processor capable of executing Gerchberg's algorithm in two dimensions is pictured in Fig. 4.[26,27] In plane P1, a 2-D truncated signal $g_T(x,y)$ is input into the system. The input is assumed zero outside of a specified area T. Outside this aperture is a mirror whose purpose will be explained shortly.

Lens L1 performs a Fourier transform on the input corresponding to step 1 in Fig. 1. Thus, $\mathcal{F}[g_T(x,y)]$ is incident on plane P2 where a pupil of dimension Ω is placed. The dimension of this aper-

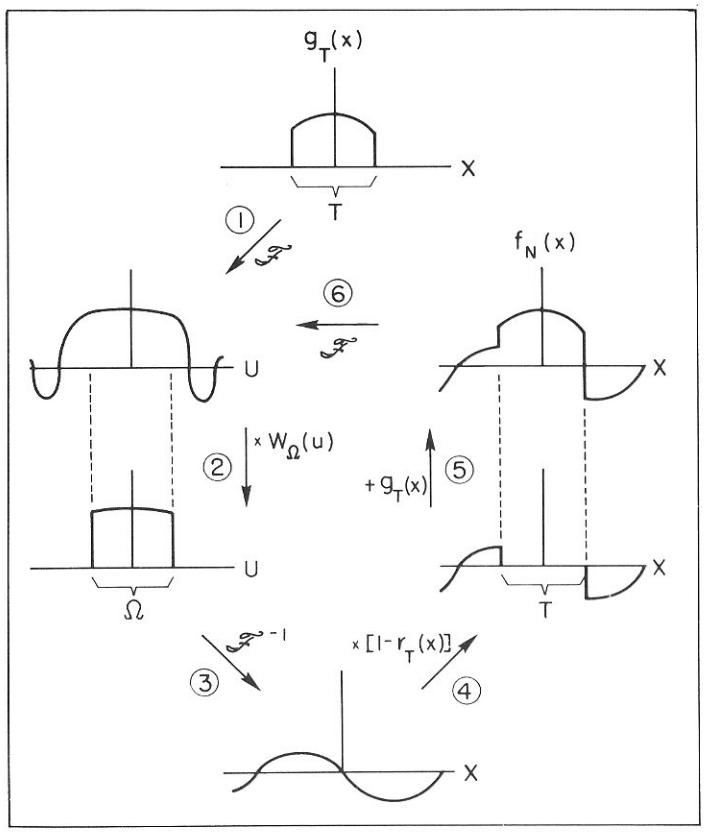


Fig. 1. Illustration of Gerchberg's iterative extrapolation algorithm. In the limit, $f_N(x) \to f(x)$.

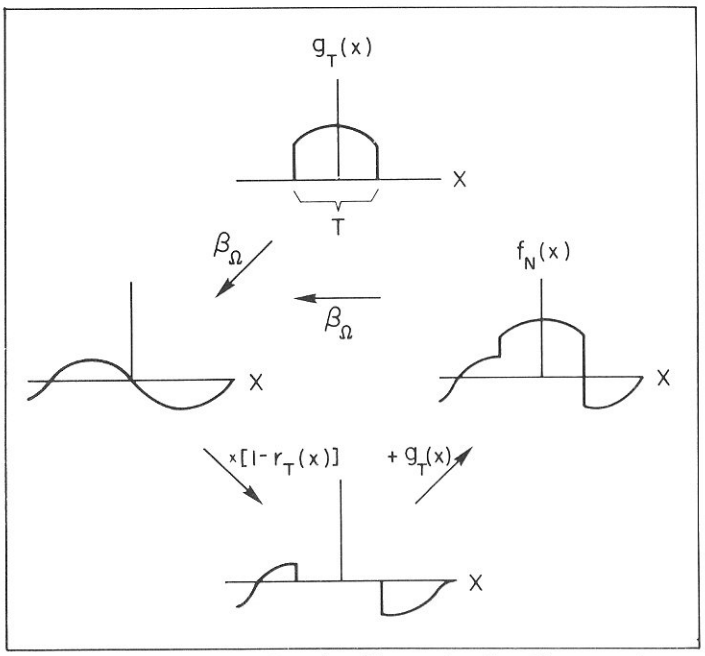


Fig. 2. Gerchberg's iterative algorithm using the $\mathcal{B}_\Omega$ operator.

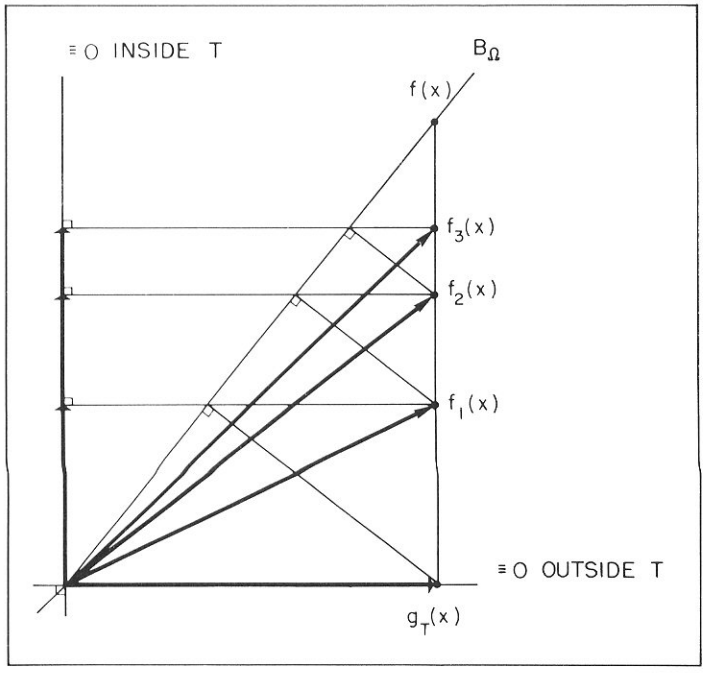


Fig. 3. Youla's illustration of the convergence of Gerchberg's algorithm in Fig. 2 in a Hilbert space.

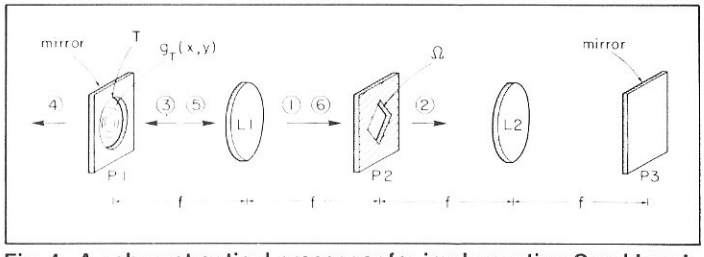


Fig. 4. A coherent optical processor for implementing Gerchberg's extrapolation algorithm in two dimensions.

ture is determined by the known bandwidth area Ω of the target object f(x,y). The field amplitude immediately to the right of plane P2 is thus the truncated version of the input's Fourier transform. This corresponds to step 2 in Fig. 1.

The next step is to perform an inverse Fourier transform—equivalent to performing three Fourier transforms. The first is done by lens L2. The transform is reflected by the right-hand mirror and is again transformed by Lens L2. If f(x,y) is real, its spectrum will be Hermetian, and Ω will therefore by symmetric about the origin. The light propagating from right to left will thus pass through the Ω pupil unaltered. The third Fourier transform is then performed by lens L1.

Steps 4 and 5 take place simultaneously. The unwanted center portion of the signal exits through the mirror's aperture and is lost to the system. The remainder of the signal, corresponding to the tails in the function shown between steps 4 and 5 in Fig. 1, is reflected back into the processor. The truncated image $g_T(x,y)$ is, of course, still being input into the system. The net result is that the field amplitude exiting plane P1 is the first estimate of f(x,y) in Gerchberg's algorithm. This is put into the system, and step 6 is performed. After a number of iterations, the extrapolated signal should appear on plane P1 (and also on plane P3).

Note, however, that there is no method yet to detect the processor output. This can be done by placing a highly transmitting pellicle at an angle in the feedback path to sample the output.

There are a number of degrative factors associated with this implementation. Extrapolation algorithms are highly sensitive to inut noise and perturbations. Both of these degrative factors are

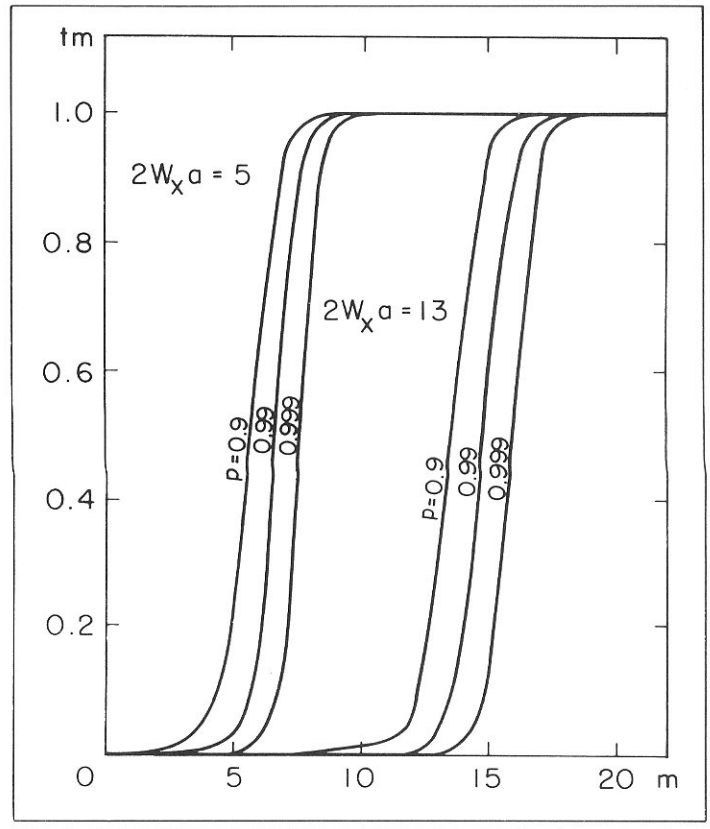


Fig. 5. A plot of the degrading coefficient t_m in the absorptive loss error in Eq. (4). The ideal $t_m \equiv 0$ situation occurs when there is no absorptive loss.

characteristic of coherent optical processors.

To illustrate the sensitivity to a small algorithm perturbation, consider the loss in each feedback cycle due to the reflectance of the detection pellicle and absorptance of the optical elements. Instead of Eq. (2), we are implementing

$$\overline{f}_N(x) = g_T(x) + p[1-r_T(x)]\,\mathscr{B}_\Omega f_{N-1}(x)\,, \tag{3}$$

where $p < 1$ denotes the transmittance of one feedback cycle. We will consider the effect of absorptive loss for the case where T extends from $-a/2$ to $a/2$, and Ω is the interval defined by $|u| \le W$. As is shown in Appendix B, the rms error ϵ_A generated from absorptive losses is given by

$$\epsilon_A = \sqrt{\sum_{m=0}^{\infty} t_m |\hat{f}_m|^2}\,, \tag{4}$$

where t_m, derived in Appendix B, is plotted in Fig. 5 for two values of space-bandwidth product, and $\hat{f}_m$ is the inner product of f(x) with an m^{th} order prolate spheroidal wave function (see Appendix A). For $p = 1$, $t_m \equiv 0$.

Making the very *rough* approximation

$$t_m = \begin{cases} 0\,; & 0 \le m < 2Wa \\ 1\,; & m \ge 2Wa\,, \end{cases} \tag{5}$$

we obtain

terval:

$$W_\Omega(u) = \begin{cases} 1 \; ; \; u \in \Omega \\ 0 \; ; \; u \notin \Omega \; . \end{cases}$$

Let T correspond to an interval on x, and let

$$g_T(x) = \begin{cases} f(x) \; ; \; x \in T \\ 0 \quad ; \; x \notin T \; . \end{cases}$$

The extrapolation problem is to determine f(x) with knowledge of $g_T(x)$ and Ω. If T corresponds to the complement of a finite interval, we have an interpolation problem. The development to follow is applicable for any T.

Gerchberg's algorithm, illustrated in Fig. 1, is a specific application of the iterative imposition of both frequency and spatial domain constraints, as discussed by Fienup in the preceding chapter. Beginning with $g_T(x)$, we first perform a Fourier transform. The frequency constraint is satisfied by keeping only those frequency components within the Ω interval in step 2. Step 3 is a simple inverse Fourier transformation. To satisfy spatial constraints, we replace this function in the T interval by $g_T(x)$ in steps 4 and 5. This first estimate is Fourier transformed and the cycle is repeated. In the absence of noise, convergence of the N^{th} estimate of f(x), $f_N(x)$, as $N \to \infty$ has been proven in three distinct ways.[17-20]

If we define the bandlimiting operator

$$\mathcal{B} = \mathcal{F}^{-1} W_\Omega(u) \mathcal{F} \; ,$$

then the iterative algorithm can be written as

$$f_N(x) = g_T(x) + [1 - r_T(x)] \mathcal{B}_\Omega f_{N-1}(x) \; , \qquad (2)$$

where the spatial rectangular window is

$$r_T(x) = \begin{cases} 1 \; ; \; x \in T \\ 0 \; ; \; x \notin T \; . \end{cases}$$

Motivated by Eq. (2), a condensed illustration of the algorithm is shown in Fig. 2. Note that $g_T(x)$ is block orthogonal to the N^{th} tail estimate $[1-r_T(x)]f_N(x)$ (i.e., the product of the two functions is identically zero). This observation allows an enlightening geometric view of the iterative algorithm in a Hilbert space setting.[19] In Fig. 3, we illustrate three subspaces as three lines in a planar L_2 function space. The horizontal line contains all L_2 functions identically zero outside the interval $x \in T$. The vertical line contains all L_2 functions identically zero *within* the interval $x \in T$. These two function classes are block orthogonal and are thus drawn perpendicular. The third function class, B_Ω, consists of all L_2 functions with bandwidth interval Ω. The bandlimited function we desire as an extrapolation (interpolation) result—henceforth referred to as the target function f(x)—lies somewhere in this space. The information we have, $g_T(x)$, is the projection of f(x) onto the space of T interval functions.

To regain f(t) from $g_T(x)$ we first project $g_T(x)$ onto B_Ω. With reference to Eq. (2), this, in turn, is projected onto the space of images identically zero within T. This vector is added to the $g_T(x)$ vector to arrive at the first estimate, $f_1(x)$, of the target function. The process is then repeated and the estimate iteratively converges to f(x).

2.2. An interative optical implementation

A coherent processor capable of executing Gerchberg's algorithm in two dimensions is pictured in Fig. 4.[26,27] In plane P1, a 2-D truncated signal $g_T(x,y)$ is input into the system. The input is assumed zero outside of a specified area T. Outside this aperture is a mirror whose purpose will be explained shortly.

Lens L1 performs a Fourier transform on the input corresponding to step 1 in Fig. 1. Thus, $\mathcal{F}[g_T(x,y)]$ is incident on plane P2 where a pupil of dimension Ω is placed. The dimension of this aper-

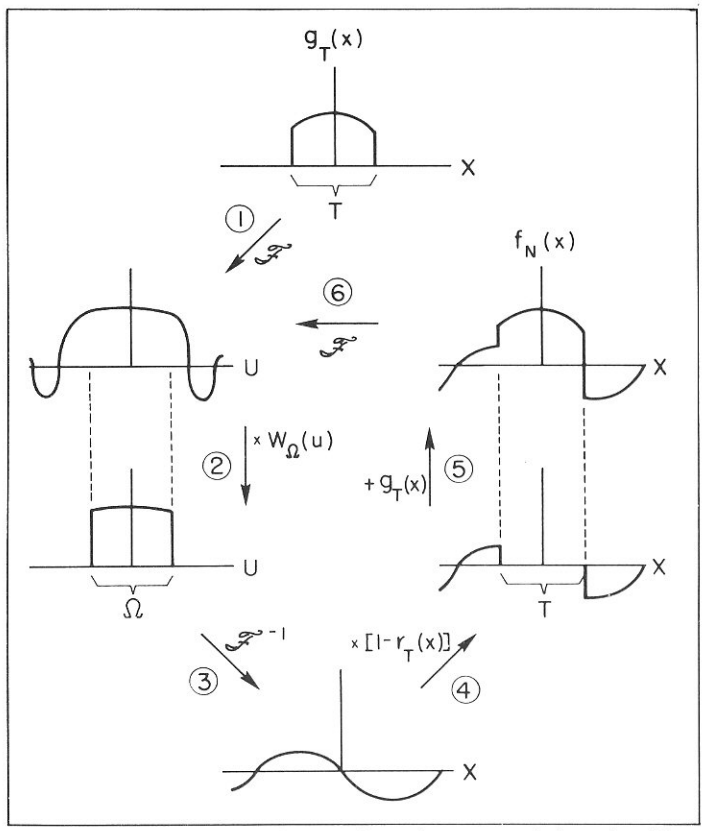


Fig. 1. Illustration of Gerchberg's iterative extrapolation algorithm. In the limit, $f_N(x) \to f(x)$.

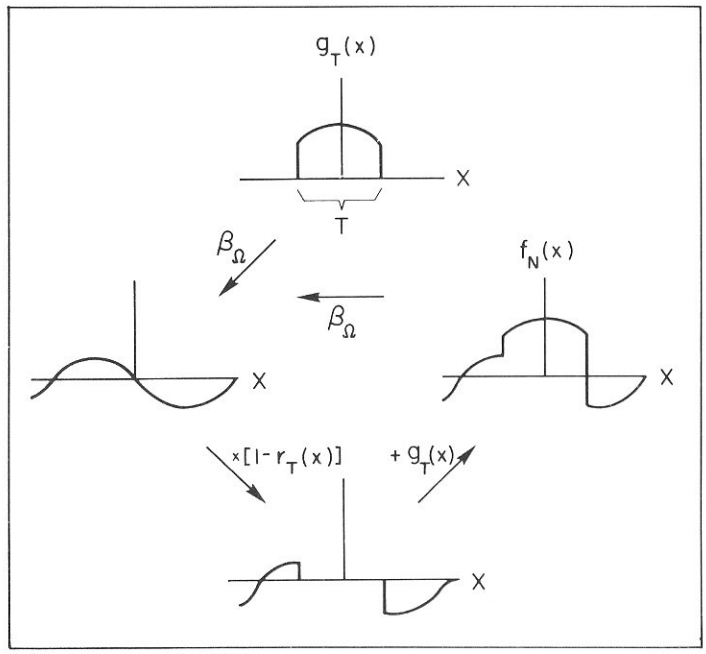


Fig. 2. Gerchberg's iterative algorithm using the $\mathcal{B}_\Omega$ operator.

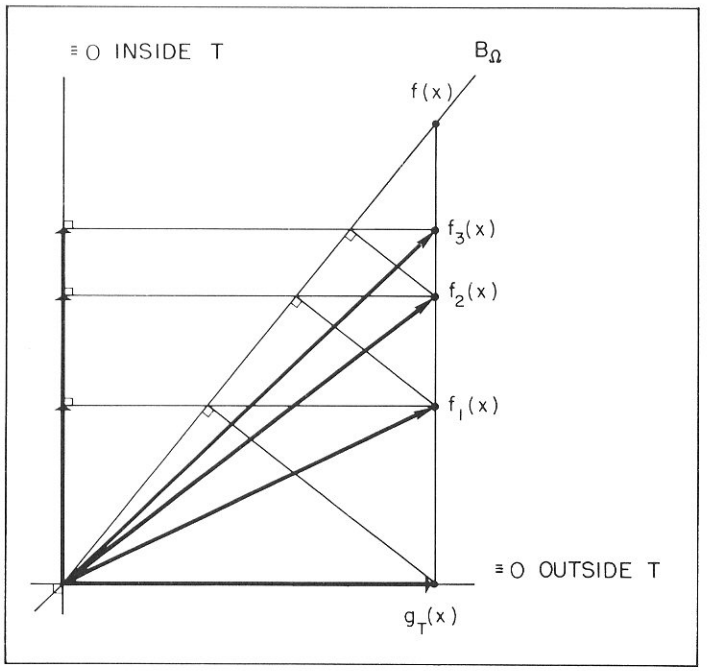


Fig. 3. Youla's illustration of the convergence of Gerchberg's algorithm in Fig. 2 in a Hilbert space.

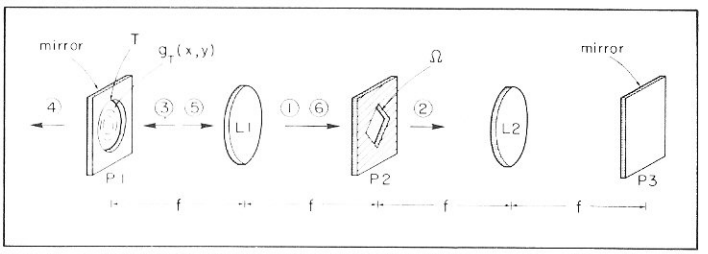


Fig. 4. A coherent optical processor for implementing Gerchberg's extrapolation algorithm in two dimensions.

ture is determined by the known bandwidth area Ω of the target object f(x,y). The field amplitude immediately to the right of plane P2 is thus the truncated version of the input's Fourier transform. This corresponds to step 2 in Fig. 1.

The next step is to perform an inverse Fourier transform—equivalent to performing three Fourier transforms. The first is done by lens L2. The transform is reflected by the right-hand mirror and is again transformed by Lens L2. If f(x,y) is real, its spectrum will be Hermetian, and Ω will therefore by symmetric about the origin. The light propagating from right to left will thus pass through the Ω pupil unaltered. The third Fourier transform is then performed by lens L1.

Steps 4 and 5 take place simultaneously. The unwanted center portion of the signal exits through the mirror's aperture and is lost to the system. The remainder of the signal, corresponding to the tails in the function shown between steps 4 and 5 in Fig. 1, is reflected back into the processor. The truncated image $g_T(x,y)$ is, of course, still being input into the system. The net result is that the field amplitude exiting plane P1 is the first estimate of f(x,y) in Gerchberg's algorithm. This is put into the system, and step 6 is performed. After a number of iterations, the extrapolated signal should appear on plane P1 (and also on plane P3).

Note, however, that there is no method yet to detect the processor output. This can be done by placing a highly transmitting pellicle at an angle in the feedback path to sample the output.

There are a number of degrative factors associated with this implementation. Extrapolation algorithms are highly sensitive to inut noise and perturbations. Both of these degrative factors are

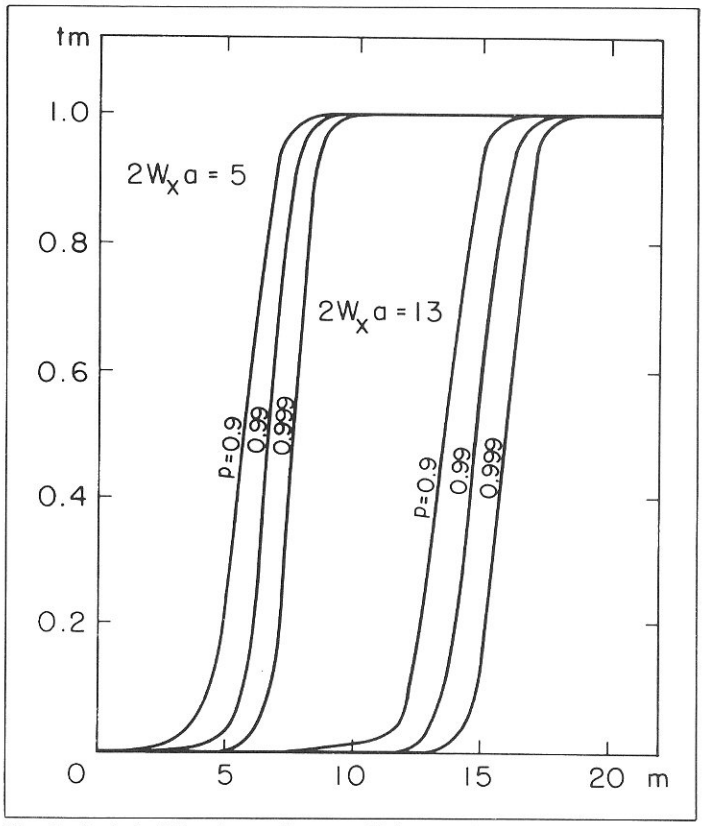


Fig. 5. A plot of the degrading coefficient t_m in the absorptive loss error in Eq. (4). The ideal $t_m \equiv 0$ situation occurs when there is no absorptive loss.

characteristic of coherent optical processors.

To illustrate the sensitivity to a small algorithm perturbation, consider the loss in each feedback cycle due to the reflectance of the detection pellicle and absorptance of the optical elements. Instead of Eq. (2), we are implementing

$$\overline{f}_N(x) = g_T(x) + p[1-r_T(x)]\,\mathcal{B}_\Omega f_{N-1}(x)\,, \tag{3}$$

where $p < 1$ denotes the transmittance of one feedback cycle. We will consider the effect of absorptive loss for the case where T extends from $-a/2$ to $a/2$, and Ω is the interval defined by $|u| \leq W$. As is shown in Appendix B, the rms error ϵ_A generated from absorptive losses is given by

$$\epsilon_A = \sqrt{\sum_{m=0}^{\infty} t_m |\hat{f}_m|^2}\,, \tag{4}$$

where t_m, derived in Appendix B, is plotted in Fig. 5 for two values of space-bandwidth product, and $\hat{f}_m$ is the inner product of f(x) with an m^{th} order prolate spheroidal wave function (see Appendix A). For $p = 1$, $t_m \equiv 0$.

Making the very *rough* approximation

$$t_m = \begin{cases} 0\,; & 0 \leq m < 2Wa \\ 1\,; & m \geq 2Wa\,, \end{cases} \tag{5}$$

we obtain

$$\epsilon_A^2 \cong \sum_{m=2Wa}^{\infty} |\hat{f}_m|^2 ,$$

which is roughly equivalent to the amount of energy outside of $|x| \leq a/2$ (see Appendix A):

$$\epsilon_0^2 = \int_{|x| \geq a/2} |f(x)|^2 dx$$

$$= \sum_{m=0}^{\infty} (1-\lambda_m) |\hat{f}_m|^2$$

$$\cong \sum_{m=2Wa}^{\infty} |\hat{f}_m|^2 ,$$

where we have made the approximation[9]

$$\lambda_m \simeq \begin{cases} 1 ; 0 \leq m < 2Wa \\ 0; \quad m \geq 2Wa . \end{cases} \tag{6}$$

Thus, for $p < 1$, the error ϵ_A will be of the same order of magnitude as ϵ_0.

To empirically illustrate the effects of absorptive losses, extrapolation of a sin x/x function was chosen. This function has been shown to extrapolate well, thus minimizing effects due to input noise and other algorithm perturbations. The extrapolation results for two values of p are shown in Fig. 6. For $p = 1$, the extrapolation is indistinguishable from the target function. The ratio $\epsilon_A^2/\epsilon_0^2$ is shown versus p in Fig. 7. A typical value for p is 0.9.[28]

In Fig. 8, we show some experimental results from a coherent extrapolation. Figure 8(a) shows a bandlimited image with Ω being a circle. Note the starlike structure immediately to the left of center. Figure 8(b) shows the extrapolation result. The portion of the image to the left of the vertical line was passed, and the right side is the extrapolation result. Although seemingly little of the structure far from the border has been faithfully reconstructed, note that three of the limbs of the star appear to be regenerated. As is the case with any extrapolation algorithm, we can, at best, only hope to extract a few more degrees of freedom from a truncated image. Other experimental results are given by Marks and Smith.[27,29]

Cederquist and Lee[28] have suggested use of a similar coherent processor—based on a confocal Fabry-Perot interferometer—to iteratively implement Gerchberg's algorithm. A noniterative *gedanken* coherent processor for extrapolation has been proposed by Frieden.[30]

2.3. The extrapolation matrix

Sabri and Steenaart[31] have placed Gerchberg's iterative algorithm into closed form. In this section, we review Sabri and Steenaart's technique, extend it to two dimensions, and illustrate its performance through various digital implementations.

One can easily show by induction that Eq. (2) can be written as

$$f_N(x) = \sum_{n=0}^{N} [\{1-r_T(x)\} \mathcal{B}_\Omega]^n g_T(x) , \tag{7}$$

or, in the limit as $N \to \infty$,

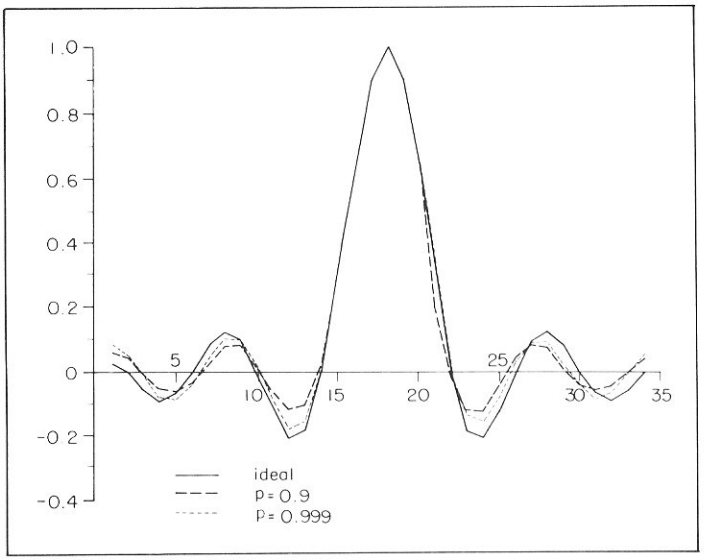


Fig. 6. Effects of two values of absorptive loss on extrapolation of a sin x/x function.

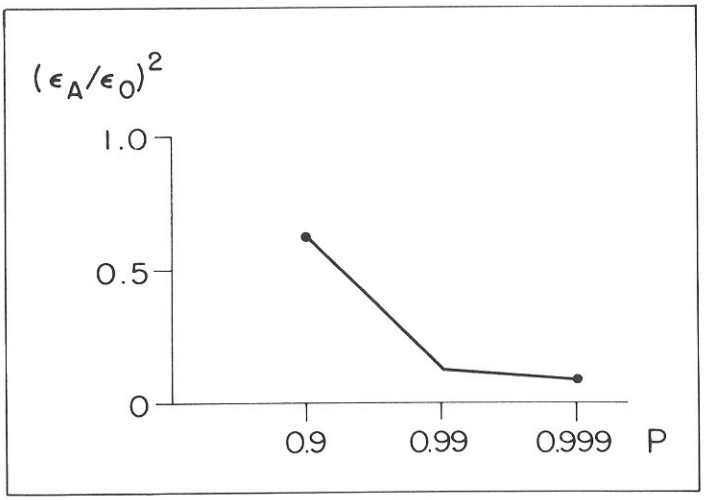


Fig. 7. A plot of the ratio of extrapolation error to exterior target energy versus single cycle transmission p.

$$f(x) = \sum_{n=0}^{\infty} [\{1-r_T(x)\} \mathcal{B}_\Omega]^n g_T(x)$$

$$= [1-\{1-r_T(x)\} \mathcal{B}_\Omega]^{-1} g_T(x) , \tag{8}$$

where, in the second step, we have used a generalized geometric series.

Let us digitize the operators. $\mathcal{B}_\Omega$ becomes a low-pass filter matrix, $\underline{B}_\Omega$, $r_T(x)$ becomes a matrix $\underline{r}_T$ with ones appropriately placed along the diagonal, and 1 becomes the identity matrix $\underline{I}$. Let $\vec{f}$ and $\vec{g}_T$ denote the vectors containing sample values of f(x) and $g_T(x)$, respectively. Then

$$\vec{f} = \underline{E}\, \vec{g}_T ,$$

where

$$\underline{E} = [\underline{I}-\{\underline{I}-\underline{r}_T\}\underline{B}_\Omega]^{-1} \tag{9}$$

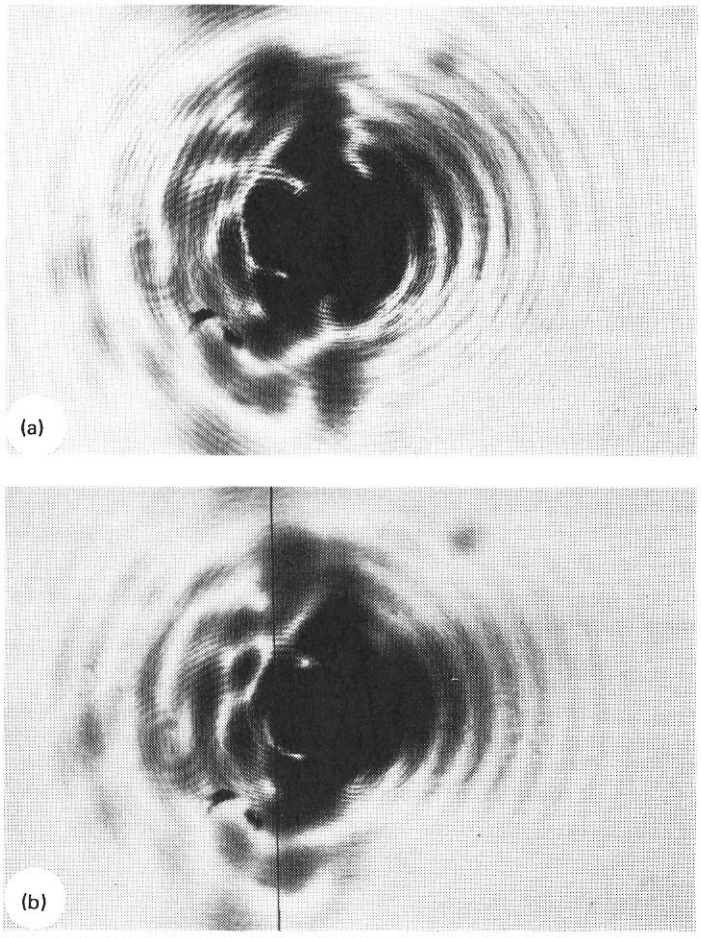


Fig. 8. Results of coherent optical extrapolation: (a) target image; (b) extrapolation.

is one form of the Sabri-Steenaart extrapolation matrix. If, instead, the N^{th} iteration estimate $\vec{f}_N$ is desired, then, from Eq. (7), we have

$$\vec{f}_N = \underline{E}_N \, \vec{g}_T , \tag{10}$$

where

$$\underline{E}_N = \sum_{n=0}^{N} [(\underline{I}-\underline{r}_T)\underline{B}_\Omega]^n . \tag{11}$$

Note that, in either case, the extrapolation matrices $\underline{E}$ and $\underline{E}_N$ are parametrized only by T and Ω.

2.4. A second 2-D iterative extrapolation algorithm

In Sec. 2.2, we generalized Gerchberg's algorithm to two dimensions in a manner that required knowledge of the entire spectral pupil Ω. Now we develop a second generalization to two dimensions —first in iterative and then in closed form—that requires knowledge only of a vertical and horizontal projection of Ω.[32]

Consider the 2-D case where T consists of one or more disjoint "islands," as pictured in Fig. 9(a). Consider the 1-D function corresponding to the horizontal slice of $g_T(x,y)$ at $y = y_o$. The "duration" of this function is dictated by T. In order to extrapolate the slice, however, we must also know its corresponding 1-D bandwidth interval. To determine this bandwidth interval, consider the spectrum in Fig. 9(d) and its inverse transform in y in Fig. 9(b). View the inverse transform from Figs. 9(d) to 9(b) as being performed along vertical slices. If the slice intersects Ω, we are inverse

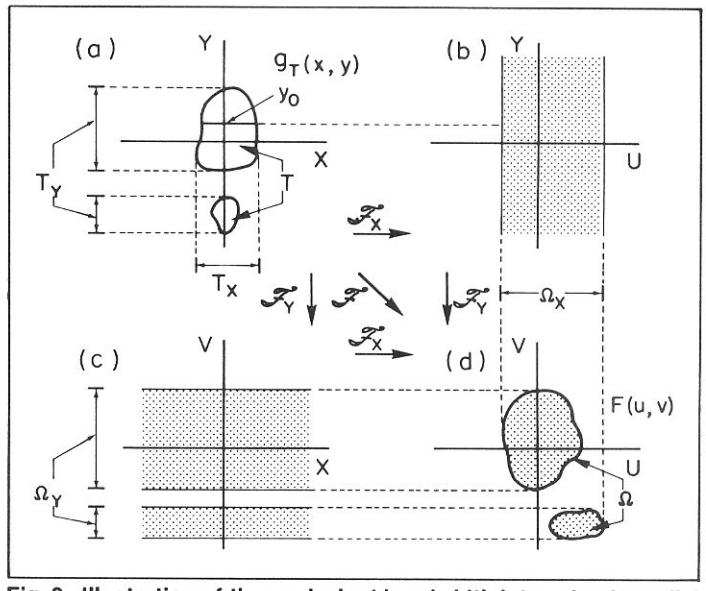


Fig. 9. Illustration of the equivalent bandwidth intervals of parallel slices of bandlimited images.

transforming a function with compact support. From the uncertainty principle of Fourier analysis, the result is a function which is bandlimited (in the 1-D sense) and is thus not identically zero over any finite region. If the slice does not intersect Ω, the inverse transform is, of course, zero. We thus conclude that the function in Fig. 9(b) is nonzero only within the shaded strip defined by the interval Ω_x. The bandwidth interval of the horizontal slice in Fig. 9(a) is therefore Ω_x irrespective of our choice of y_o. Generalizing, we conclude that *two 1-D functions corresponding to two parallel slices of a bandlimited image have identical bandwidth intervals.* Note, as shown in Fig. 9(c) for the vertical case, that this interval can be disjoint.

With knowledge of the duration and bandwidth intervals of each horizontal slice, we can apply Gerchberg's algorithm in one dimension to each horizontal slice in Fig. 9(a) and generate f(x,y) for the strip defined by all $y \in T_y$. Then, using the bandwidth interval Ω_y, this result can be vertically extrapolated to yield f(x,y) over the entire (x,y) plane.

Mathematically, we can write the horizontal extrapolation as

$$f(x,y) r_{T_y}(y) = \sum_{m=0}^{\infty} \mathcal{H}^m_{T,\Omega_x} g_T(x,y) ,$$

where

$$\mathcal{H}_{T,\Omega_x} \equiv [1 - r_T(x,y)] \, \mathcal{B}_{\Omega_x} ,$$

and $r_T(x,y)$ defines the 2-D truncating pupil:

$$r_T(x,y) = \begin{cases} 1 \; ; \; (x,y) \in T \\ 0 \; ; \; (x,y) \notin T . \end{cases}$$

Vertical extrapolation follows as

$$f(x,y) = \sum_{n=0}^{\infty} \mathcal{H}^n_{T_y,\Omega_y} f(x,y) r_{T_y}(y)$$

$$= \sum_{n=0}^{\infty} \mathcal{H}^n_{T_y,\Omega_y} \sum_{m=0}^{\infty} \mathcal{H}^m_{T,\Omega_x} g_T(x,y) , \tag{12}$$

where

$$\mathcal{H}_{T_y,\Omega_y} = [1-r_{T_y}(y)]\,\mathcal{B}_{\Omega_y} .$$

Note that we can rewrite Eq. (12) as

$$f(x,y) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \mathcal{H}^{n-m}_{T_y,\Omega_y} \mathcal{H}^{m}_{T,\Omega_x} g_T(x,y)$$

and, in the spirit of iteration, define

$$f_N(x,y) = \sum_{n=0}^{N} \sum_{m=0}^{n} \mathcal{H}^{n-m}_{T_y,\Omega_y} \mathcal{H}^{m}_{T,\Omega_x} g_T(x,y) . \qquad (13)$$

Note that

$$f_N(x,y) = f_{N-1}(x,y) + a_N(x,y) , \qquad (14)$$

where

$$a_N(x,y) = \sum_{m=0}^{N} \mathcal{H}^{N-m}_{T_y,\Omega_y} \mathcal{H}^{m}_{T,\Omega_x} g_T(x,y) .$$

Furthermore,

$$a_N(x,y) = b_N(x,y) + \mathcal{H}_{T_y,\Omega_y} a_{N-1}(x,y) , \qquad (15)$$

where

$$b_N(x,y) = \mathcal{H}^{N}_{T,\Omega_x} g_T(x,y) .$$

Obviously,

$$b_N(x,y) = \mathcal{H}_{T,\Omega_x} b_{N-1}(x,y) . \qquad (16)$$

Equations (14), (15), and (16) define an iterative form of Eq. (13) with initializations

$$f_0(x,y) = a_0(x,y) = b_0(x,y) = g_T(x,y) .$$

Some proofs of the convergence of $f_N(x,y)$ to $f(x,y)$ for some specific truncation apertures are given by Marks.[32]

A contrast between the above algorithm (#2) and that implemented on the coherent processor (#1) are in order. Algorithm 1 requires knowledge of the entire spectral pupil region Ω. Algorithm 2 requires only knowledge of two projections of Ω: Ω_x and Ω_y. We are thus utilizing less information in this case and, as might be expected, will in some sense diminish algorithm effectiveness.

Consider Fig. 10, in which we wish to extrapolate $g_T(x,y)$. Using algorithm 2, the value of the extrapolation at point P_1, which lies within T_y, is determined solely from information gained from the intersection of $g_T(x,y)$ with line L_1. Point P_2 lies in the area where we have extrapolated the horizontal extrapolation. Point P_3 is a hybrid case, formed both from information from $g_T(x,y)$ and the horizontal extrapolation. Thus, we conclude that algorithm 2 extrapolates to a point using only 1-D slices of the original signal and/or previous extrapolations. Every point exterior to T, however, is related to every point within T. This observation is made clear upon inspection of point P_4 in Fig. 10. The extrapolated value at P_4 can, in principle, be determined from the intersection of any line through P_4 that intersects T. Algorithm 1, on the other hand, clearly relates each interior point to each exterior point with the price that the entire spectral region Ω must be known.

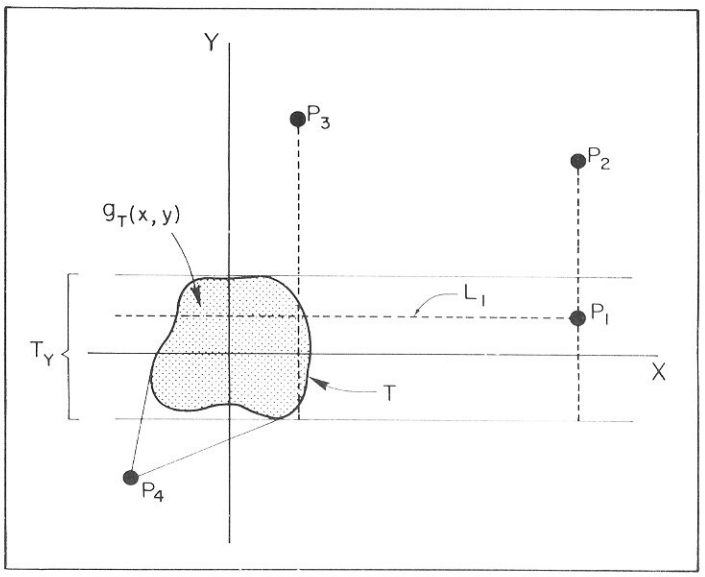


Fig. 10. Illustration of the contribution of the known portion of the image to extrapolated values at various points.

2.5. Closed form 2-D extrapolation

In this section, algorithm 2 is placed in closed form. We limit ourselves initially to separable truncating pupils:

$$r_T(x,y) = r_{T_x}(x) r_{T_y}(y) .$$

For this case, the duration *and* bandwidth interval of each horizontal slice is equivalent. Thus, when digitized, the same extrapolation matrix is used on each horizontal slice of the truncated image. Once extrapolated into a horizontal strip, each vertical slice can also be extrapolated by a single extrapolation matrix.

Let $\underline{g}_T$ denote the matrix of sample values from $g_T(x,y)$ and $\underline{E}_x$ ($\underline{E}_y$) be the extrapolation matrix in either Eq. (9) or Eq. (11) parametrized by T_x and Ω_x (T_y and Ω_y). Then,

$$\underline{f} = \underline{E}_y \underline{g}_T \underline{E}'_x , \qquad (17)$$

where $\underline{f}$ is the matrix of sample values from the extrapolated image, and the prime denotes matrix transposition.

A straightforward generalization holds when $r_T(x,y)$ has a finite T_x and T_y but is not separable. For a given y within T_y, we can extrapolate each horizontal slice of $g_T(x,y)$ using an extrapolation matrix parametrized by the same bandwidth interval Ω_x and the interval corresponding to the intersection of the horizontal line at our chosen y with $r_T(x,y)$. Once $g_T(x,y)$ is extrapolated into a horizontal strip within T_y, vertical extrapolation can be performed with a single extrapolation matrix parametrized by T_y and Ω_y. The algorithm, in fact, is applicable to all T such that a single vertical or horizontal extrapolation does not fill the entire plane. Consider, for example, Fig. 11 and let $T = T_4$. The horizontal "extrapolation" would fill T_3. The vertical "extrapolation" would then fill T_1 and T_2.

The algorithm is not applicable to the case where T is chosen such that a single vertical or horizontal extrapolation fills the plane. Such a case is when $T = T_2 + T_4$ in Fig. 11. An alternate approach is thus necessary. One method is simply to perform a single 1-D extrapolation:

$$f(x,y) = \sum_{n=0}^{\infty} \mathcal{H}^{n}_{T,\Omega_x} g_T(x,y) . \qquad (18)$$

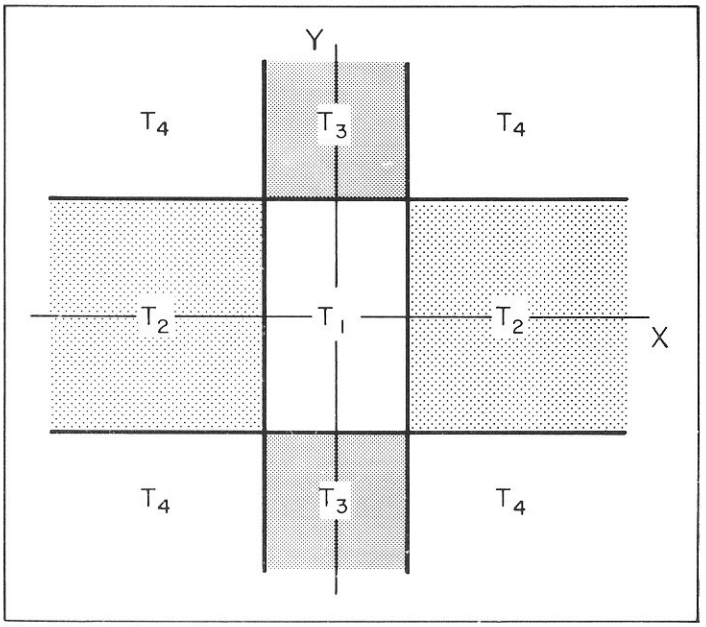


Fig. 11. Regions of convergence.

Note, however, for $T = T_2 + T_3$ in Fig. 11, the portion of $g_T(x,y)$ in T_3 is not used to determine $f(x,y)$ in region T_1. A better method, in this case, equivalent to averaging, is

$$f(x,y) = \frac{1}{2} \sum_{n=0}^{\infty} [\mathcal{H}^n_{T,\Omega_x} + \mathcal{H}^n_{T,\Omega_y}] g_T(x,y) \ . \tag{19}$$

For $T = T_2 + T_3 + T_4$, even this method does not use information in T_4 for the result. One might conclude that better extrapolation can be gained by averaging the contributions of a larger and larger number of radial strips. Radial extrapolation could similarly be applied to algorithm 2, as presented at the beginning of this section. Each radial strip, however, requires knowedge of another projection of Ω. Thus, in the limit, we would require complete knowledge of Ω.

Note, lastly, that the algorithms in Eqs. (18) and (19) can easily be placed in either iterative or closed form.

2.6. Some closed form extrapolation results

In this section, we present empirical results of the closed form 2-D extrapolation algorithm.[33] A number of bandlimited target images were created and a square ($N \times N$), $N = 34$, matrix of sample values $\underline{f}$ was formed. A smaller ($m \times m$) square matrix $\underline{g}_T$ was then used as the truncated image. The extrapolation matrices were formed by the matrix inversion technique. A Hilbert transform low-pass matrix[34] was utilized for B_Ω. Each target function was chosen to explore a particular aspect of the algorithm performance.

It is desirable to have a figure of merit to quantify the goodness of the extrapolation results. Let $\hat{\underline{f}}$ denote the extrapolation result. One obvious merit comparison is

$$\phi = \frac{(\underline{f}|\hat{\underline{f}})}{\|\underline{f}\| \ \|\hat{\underline{f}}\|} ,$$

where, for real-valued matrices,

$$(\underline{f}|\hat{\underline{f}}) = \sum_{i,j=1}^{N} f_{ij}\hat{f}_{ij} ,$$

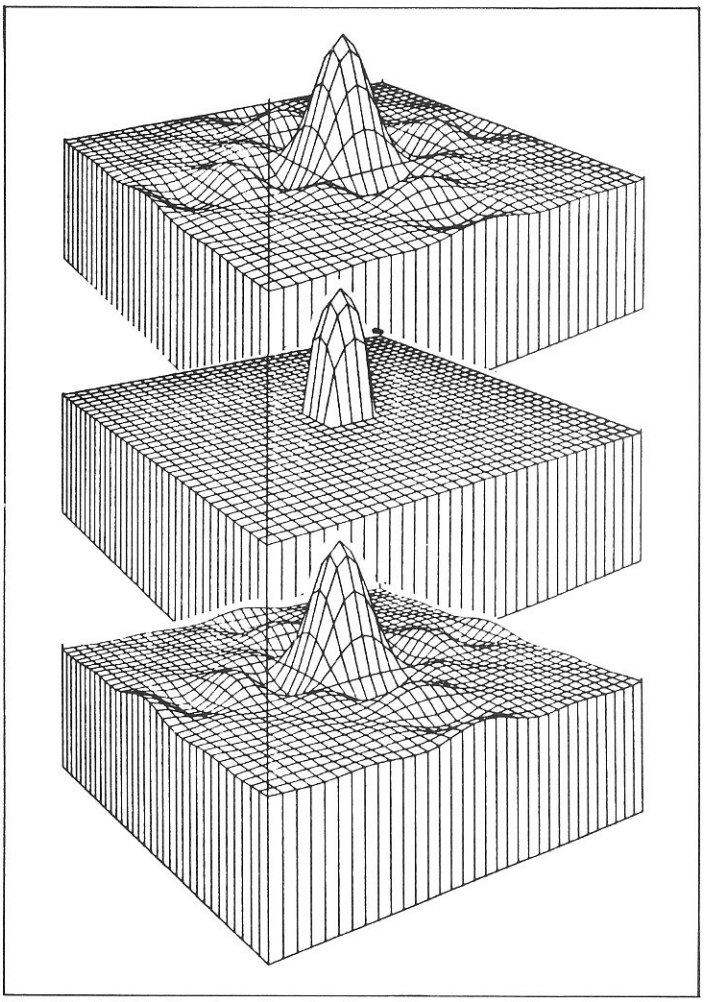


Fig. 12. A 2-D sinc function (bottom figure) and its extrapolation (top figure) generated from the truncated image (center figure).

and $|k|^2 \equiv (k\ k)$. From the Cauchy-Schwarz inequality, $|\phi| \leq 1$. This figure of merit, however, is insensitive to the goodness of the extrapolation "near" the truncated image. An obvious alteration is to run the inner product sum from the center of the matrix to a centered $d \times d$ square. The matrices $\underline{f}$ and $\hat{\underline{f}}$, however, are both equivalent to $\underline{g}_T$ within the $m \times m$ centered square. We remove this bias and write our final figure of merit as

$$\phi(d) = \frac{(\underline{f}|\hat{\underline{f}})_{m,d}}{\|\underline{f}\|_{m,d} \ \|\hat{\underline{f}}\|_{m,d}} ,$$

where

$$(\underline{f}|\underline{f})_{m,d} \equiv \sum_{\text{outside } m \times m \text{ square}}^{d \times d \text{ square}} f_{ij}\hat{f}_{ij} \ .$$

A value of $\phi(d)$ near unity then dictates a good result. Under the assumption that the extrapolation is better "near" to where the image is known, $\phi(d)$ should be a monotonically decreasing function of d. The example implementations to follow are presented in pseudo 3-D plots. The top figure in each case corresponds to the target function. The center plot shows the truncated image, and the bottom plot is the extrapolation result.

Example #1. In Fig. 12, the target function is a 2-D sinc function:

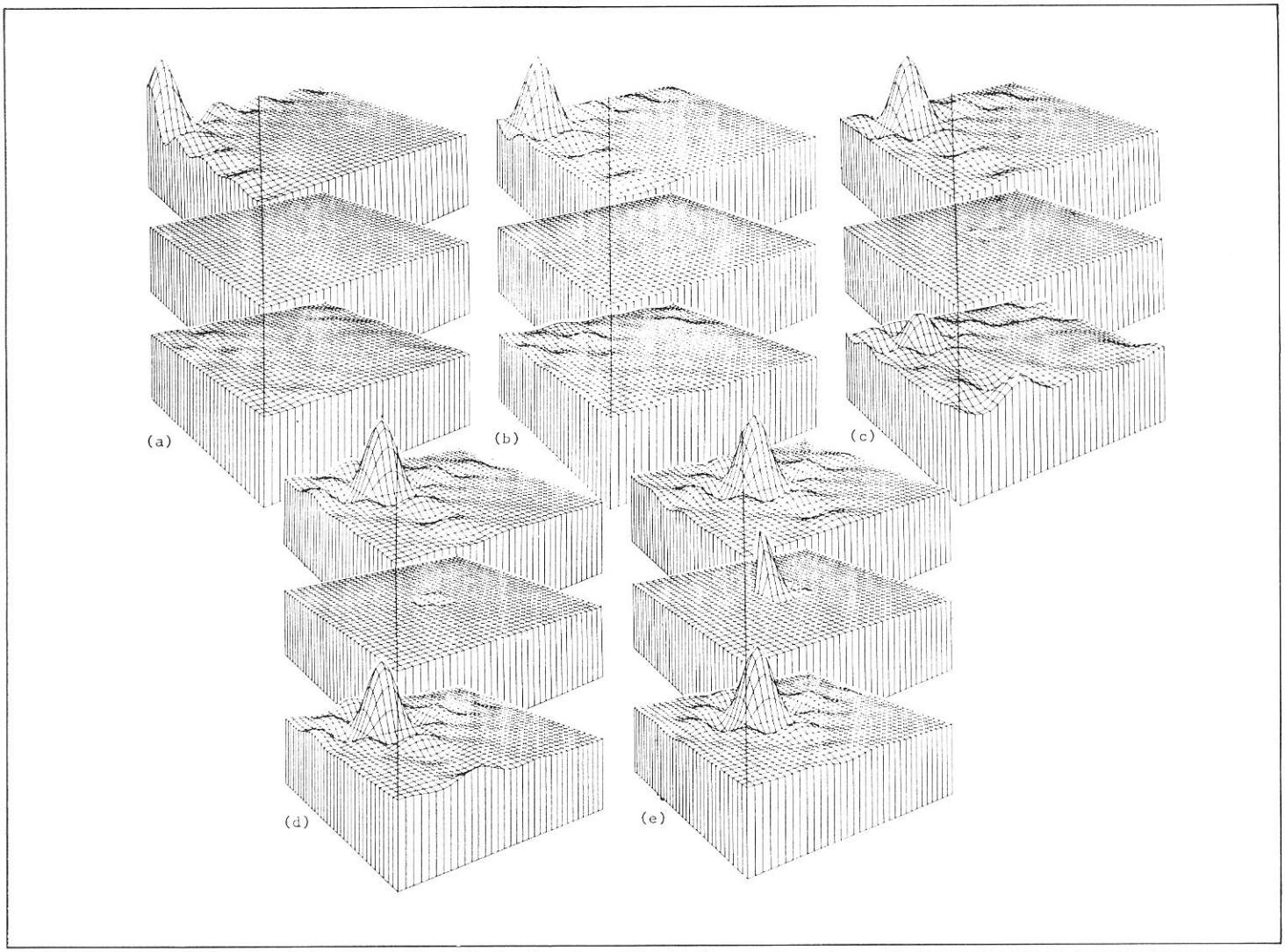


Fig.13(a-e). Extrapolation results for five shifted 2-D sinc functions. The top figure—in each case is the target image. The center figure is the truncated image, and the bottom figure is the extrapolation result.

$$g(x,y) = \{\text{sinc}[2(0.125)(x-17)]\}\ \{\text{sinc}[2(0.125)(y-17)]\}\ .$$

The sinc function was found to produce the most accurate reconstruction of any of the functions that were tried. The figure of merit was found to slowly decrease monotonically with respect to d with a minimum value of 0.975.

Example #2. This group of figures depicts the extrapolation results for five sequentially shifted off-axis sinc functions. Figure 13(a) shows the extrapolation result when the center of the target sinc function was located at point (2,32) in the x,y plane. The input is extremely low in energy. In Fig. 13(b), the target function has been moved closer to the center of the matrix along the diagonal. For this shift and all subsequent ones, the target function was moved three units in the positive x direction and three units in the negative y direction. Figure 13(c) shows the result of extrapolating the target sinc, which has been moved yet closer to the center along the diagonal. There is still very little energy available as input. A pronounced peak is forming in the corner. In Fig. 13(d), the target function has been shifted again. The remarkable thing about this particular extrapolation is its accuracy in view of the small amount of energy that was passed as input. The final extrapolation in this group is depicted in Figure 13(e). The extrapolation here is very good. Figure 13(f) is the figure of merit graph for the five preceding cases. The lines corresponding to Figs. 13(a) and 13(b) show better

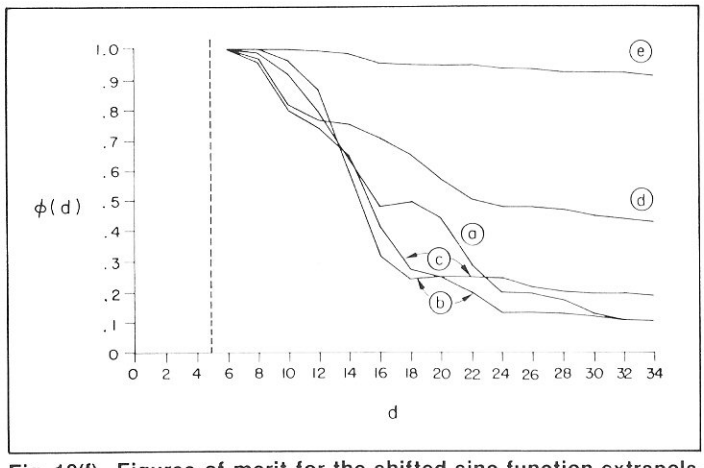


Fig. 13(f). Figures of merit for the shifted sinc function extrapolation results.

extrapolation close to the truncation aperture than do those corresponding to Figs. 13(c) and 13(d). Thus, even though a low percentage of the target function was available, extrapolation near to the truncation is still good. Note that, in each case, ϕ(d)

decreases roughly monotonically.

Example #3. In Fig. 14, we illustrated that extrapolation should be better on the vertical and horizontal strips than at the corners. The centered sinc in example #1 is a very nice signal to extrapolate because most of its structure lies along centered horizontal and vertical strips. The sinc function does not give a very good indication of the performance of the algorithm within the corner regions of the matrix however. Therefore, a rotated sinc was used as an input in order to study the extrapolation in these regions. The rotated sinc has most of its structure within the corner regions, and it is given by

$$g(x,y) = [\text{sinc2}(0.125)(x-y)][\text{sinc2}(0.125)(x+y-34)] .$$

As expected, the results in Fig. 14(a) are inferior to the unrotated case in Fig. 12(a). The corresponding figure of merit, shown in Fig. 14(b), decreases monotonically.

Example #4. So far, all of the target functions discussed have been single sinc functions. The results for the sinc functions were quite good, and it was hoped that a sum of sincs would also produce similarly good results. A target function was created that consisted of a sum of four sinc functions arranged to look like a stylized face. Two sinc functions with equal widths in the x and y dimensions were positioned as "eyes," a sinc function that was elongated in the y direction was centered and became the "nose" (which partially obscures the left eye), and a sinc function that was broad in the x direction was positioned below center for the "mouth." In Fig. 15, the results of the extrapolation are shown. It was hoped that by passing the nose, we could reconstruct the entire face. The mouth came through okay, but the eyeballs were a bit attenuated.

Example #5. In this example, we attempt to extrapolate a sinc function on a constant bias. A value of -0.6 was added to each sampled value of the sinc function, and the resulting extrapolation is pictured in Fig. 16. It appears that the algorithm was unable to predict the presence of the bias term. The finite (x,y) aperture does not convey low frequency information well since only a fraction of the "periods" of these terms can be generated.

Example #6. The final set of figures illustrates the performance of the extrapolation algorithm when the input is perturbed by zero-mean, additive, white Gaussian noise. The signal-to-noise (SNR) ratio was changed by specifying different variances for the Gaussian distribution.

In Fig. 17(a), the output for a sinc perturbed by a Gaussian noise is illustrated. The density function of this noise is given by

$$P(\alpha) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp[-\alpha^2/2\sigma^2] ,$$

where σ^2 is the variance. The signal-to-noise ratio is 10.0 (SNR = $1/\sigma$), where the signal strength in each case is taken to be the maximum value of the target signal, which is roughly unity. The input is obviously very distorted, with only the main lobe of the sinc function being even partially recognizable. The extrapolated output is extremely poor. Note the scale change caused by the high amplitude peaks in the corners of the output matrix. The horizontal and vertical strips corresponding to the first-order (direct) extrapolation have relatively low amplitude and could possibly be accurate, but the corners, corresponding to the second-order (indirect) extrapolations, are wildly varying. This same effect is characteristic when the bandwidth has been specified incorrectly. The value of σ reduces to 0.01 in Fig. 17(b), 0.001 in Fig. 17(c), 0.0001 in Fig. 17(d) and 0.00001 in Fig. 17(e). This last result corresponds to an incredible SNR of 10^5, dramatically illustrating extrapolation's extreme sensitivity to noise. The figures of merit for the five extrapolations are graphed in Fig. 17(f). As expected, the more noise, the worse the extrapolation.

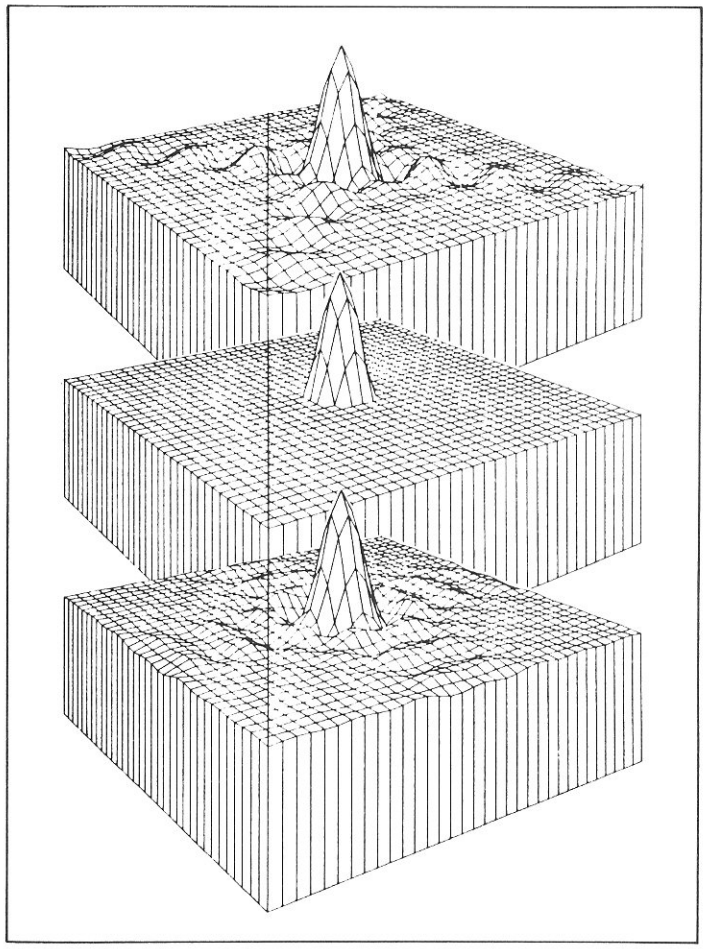


Fig. 14(a). Extrapolation results (bottom figure) for a rotated sinc function (top figure) from the truncated image (center figure).

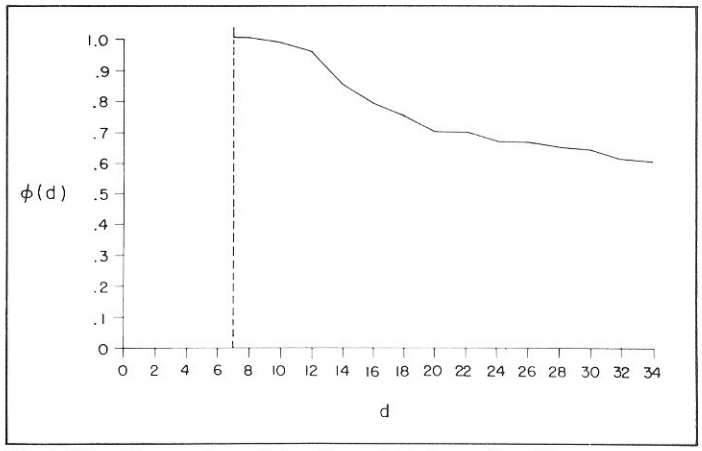


Fig. 14(b). Figure of merit for the rotated sinc function.

The sensitivity of the results in this example result directly from the ill-posedness of the particular extrapolation problem (discussed in detail later in this chapter). Such ill-posed problems many times manifest themselves digitally as ill-conditioned matrices.[35] Such matrices are characterized by an eigenvalue range of many orders of magnitude and are extremely sensitive to small input data perturbations.[36]

For an elementary, yet dramatic, example of an ill-conditioned matrix, consider the operation

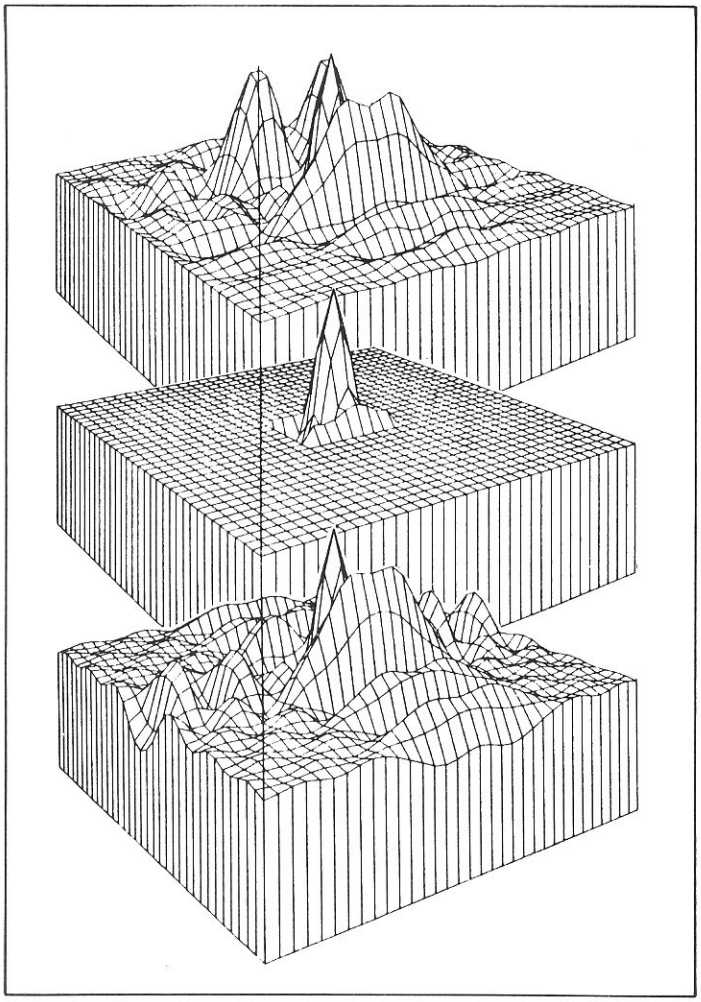


Fig. 15. Extrapolation results (bottom figure) of a stylized face (top figure) from the nose (center figure).

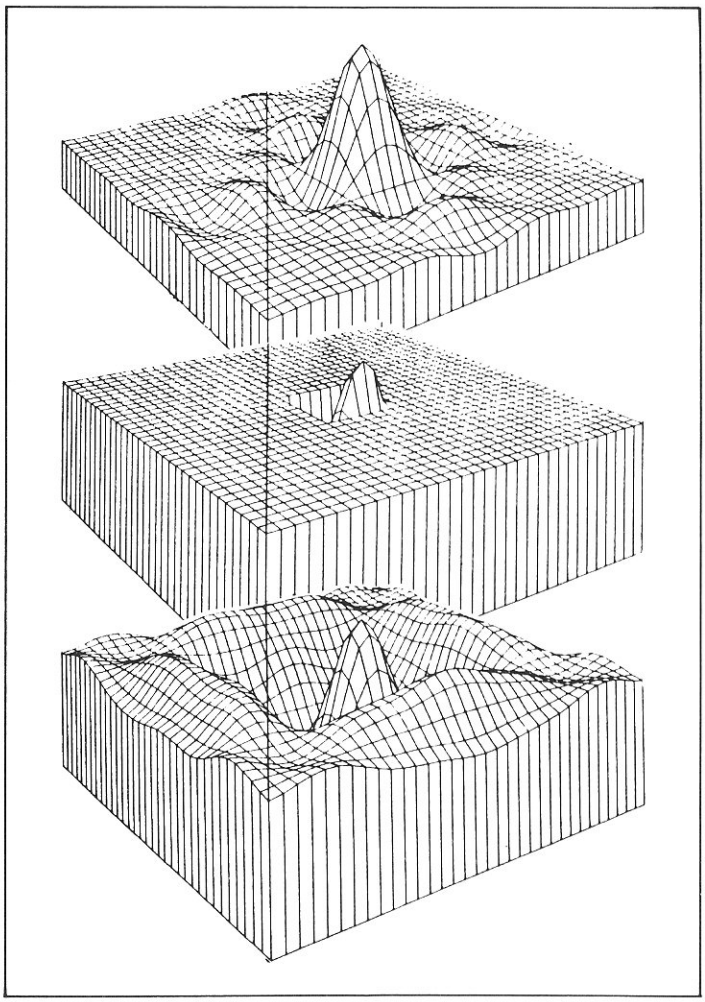


Fig. 16. Extrapolation results (bottom figure) of a 2-D sinc function with a bias (top figure) generated from the truncated image (center figure).

$$\begin{bmatrix} 39.604 & 32.258 \\ 31.000 & 25.250 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 7.346 \\ 5.750 \end{bmatrix} .$$

The ratio of maximum to minimum matrix eigenvalue is of the order of 10^6. Conventional inversion, as expected, gives

$$\begin{bmatrix} 8416.667 & -10{,}752.667 \\ -10{,}333.333 & 13{,}201.333 \end{bmatrix} \begin{bmatrix} 7.346 \\ 5.750 \end{bmatrix} = \begin{bmatrix} 1.00 \\ -1.00 \end{bmatrix} .$$

If, however, we perturb the data point 7.346 to 7.347, we obtain

$$\begin{bmatrix} 8416.667 & -10{,}752.667 \\ -10{,}333.333 & 13{,}201.333 \end{bmatrix} \begin{bmatrix} 7.347 \\ 5.750 \end{bmatrix} = \begin{bmatrix} 15.17 \\ -11.33 \end{bmatrix} .$$

The solution changes by an order of magnitude. This same phenomenon is that encountered in the results of Fig. 17.

3. A CLASS OF DECONVOLUTION ALGORITHMS

3.1. Background

In this section, we will show that Gerchberg's extrapolation algorithm is one of a number of deconvolution algorithms that can be deduced from a general methodology. Let f(x) be bandlimited in the sense of Eq. (1). Let k(x) be an as of yet arbitrary function, and define the degraded image g(x) by

$$g(x) = f(x)k(x) .$$

Given g(x), k(x), and Ω, the problem at hand is to determine f(x). Note that if we switch the roles of the spatial and frequency domain, this amounts to a deconvolution problem. If $k(x) = r_T(x)$, we can have either an extrapolation problem for finite T or, if T is the complement of a finite interval, an interpolation problem. In the frequency domain, the corresponding deconvolution problem is referred to as super resolution.

Define an operator $\mathscr{C}$ such that

$$\mathscr{C} f(x) = g(x) . \tag{20}$$

Since f(x) is already bandlimited,

$$\mathscr{B}_\Omega f(x) = f(x) . \tag{21}$$

Admissible $\mathscr{C}$'s thus include

$$\mathscr{C} = k(x) ; \tag{22a}$$

$$\mathscr{C} = 1-[1-k(x)]\ \mathscr{B}_\Omega ; \tag{22b}$$

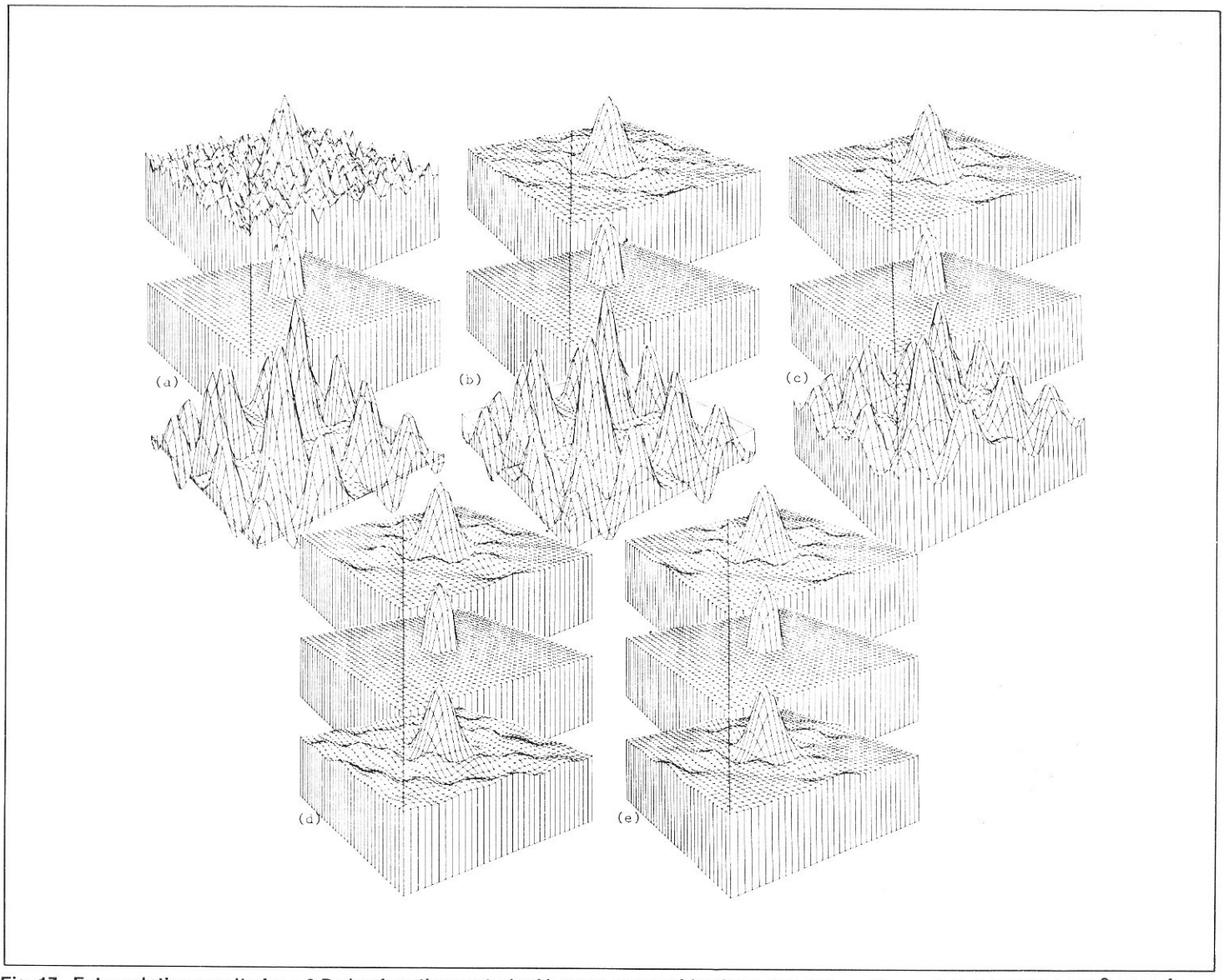


Fig. 17. Extrapolation results for a 2-D sinc function perturbed by zero mean white Gaussian noise with σ = (a)0.1, (b)0.01, (c)10^{-3}, (d)10^{-4}, and (e)10^{-5}. In each case, the top figure is the desired image, the center figure is the truncated image, and the bottom figure is the extrapolation result.

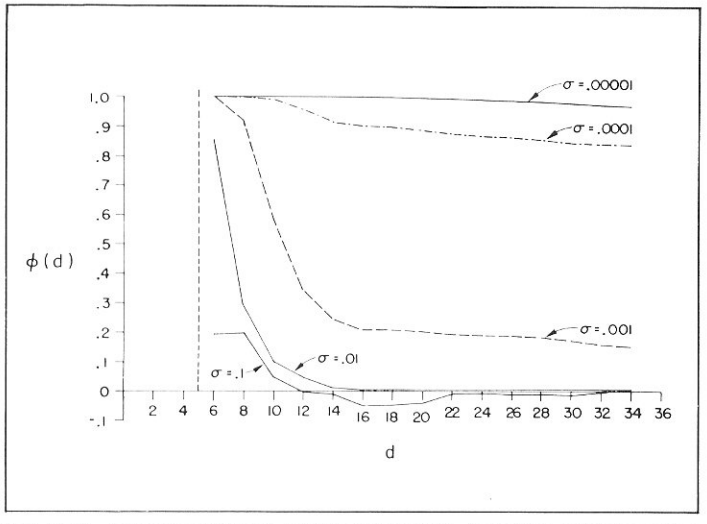


Fig. 17(f). Figures of merit for the noisy sinc function extrapolation results.

$$\mathscr{C} = k(x)\, \mathscr{B}_\Omega ; \tag{22c}$$

$$\mathscr{C} = 1 - [\, \mathscr{B}_\Omega - k(x)] ; \tag{22d}$$

$$\mathscr{C} = [\, \mathscr{B}_\Omega + k(x)] - 1. \tag{22e}$$

The possibilities are obviously limitless. If we can invert $\mathscr{C}$, then, from Eq. (20), f(x) can be found. That is,

$$f(x) = \mathscr{D}\, g(x) , \tag{23}$$

where

$$\mathscr{D} = \mathscr{C}^{-1}$$

is the deconvolution operator, Using Eq. (21), we can equivalently write

$$f(x) = \mathscr{B}_\Omega\, \mathscr{D}\, g(x) . \tag{24}$$

Let us examine some specific cases:

1) For $\mathscr{C} = k(x)$ in Eq. (22a), the solution to Eq. (23) is

$$f(x) = g(x)/k(x) .$$

This is a familiar deconvolution form with its familiar problems at those points where $k(x) = 0$.

2) Digitizing the $\mathscr{C}$ in Eq. (22b) gives

$$\underline{D} = [\underline{I} - \{\underline{I} - \underline{k}\}\underline{B}_\Omega]^{-1} .$$

For $\underline{k} = \underline{r}_T$, $\underline{D} = \underline{E}$ is a Sabri-Steenaart extrapolation matrix.[31]

3) Digitizing Eq. (22c) and using Eq. (24) gives

$$\vec{f} = \underline{B}_\Omega(\underline{k}\underline{B}_\Omega)^{-1}\vec{g} ,$$

which, for $\underline{k} = \underline{r}_T$, is recognized as Cadzow's one-step extrapolation algorithm.[21]

To place the algorithms in iterative form, define the operator $\mathscr{A}$ such that

$$\mathscr{C} = 1 - \mathscr{A} . \tag{25}$$

Then, using Eq. (23) and a generalized geometric series, we have

$$f(x) = [1 - \mathscr{A}]^{-1} g(x)$$

$$= \sum_{n=0}^{\infty} \mathscr{A}^n g(x) . \tag{26}$$

In the spirit of iteration, define

$$f_N(x) = \sum_{n=0}^{N} \mathscr{A}^n g(x) , \tag{27}$$

which, in turn, can be shown to be equivalent to

$$f_N(x) = g(x) + \mathscr{A} f_{N-1}(x) \tag{28}$$

with initialization $f_0(x) = g(x)$. Going through a similar derivation using Eq. (24) gives

$$f_N(x) = \mathscr{B}_\Omega g(x) + \mathscr{B}_\Omega \mathscr{A} f_{N-1}(x) . \tag{29}$$

The f_N's in Eqs. (28) and (29) are not necessarily equal, although both ideally should converge to the same target function. Consider some specific examples:

1) For $\mathscr{C} = k(x)$ in Eq. (22a), Eq. (28) becomes

$$f_N(x) = g(x) + [1 - k(x)] f_{N-1}(x) ,$$

which is the Van Cittert deconvolution algorithm.[23-25]

2) For $\mathscr{C}$ in Eq. (22b), Eq. (28) becomes

$$f_N(x) = g(x) + [1 - k(x)]\, \mathscr{B}_\Omega f_{N-1}(x) ,$$

which is the deconvolution algorithm of Prost and Goutte.[37] For $k(x) = r_T(x)$, it is recognized as Gerchberg's iterative extrapolation algorithm.[17]

3) For $\mathscr{C}$ in Eq. (22c), Eq. (29) becomes

$$f_N(x) = \mathscr{B}_\Omega g(x) + \{1 - \mathscr{B}_\Omega k(x)\} f_{N-1}(x) ,$$

where we have recognized that

$$\mathscr{B}_\Omega f_N(x) = f_N(x) .$$

This is recognized as the iterative form of Cadzow's algorithm. Equivalently,

$$f_N(x) = \mathscr{B}_\Omega [g(x) + \{1 - k(x)\} f_{N-1}(x)] ,$$

which is simply Van Cittert's deconvolution algorithm with a low-pass filter operation in each iteration.

3.2. Algorithm convergence

Not all the iterative algorithms formulated by the techniques of the previous sections converge in the absence of noise. Here we present necesssary and sufficient conditions for convergence for the case of extrapolation. In the preceding chapter, Fienup presents a methodology by which the convergence of certain iterative algorithms can be accelerated.

Let $k(x) = r_T(x)$ be a centered unit-amplitude rectangle over the interval $|x| \le a/2$. Let Ω be the interval defined by $|u| \le W$. Then, from Eq. (A4) in Appendix A, we can express $g_T(x)$ as

$$g_T(x) = \sum_{m=0}^{\infty} \hat{f}_m \psi_m(x) r_T(x) .$$

Equation (27) can then be written as

$$f_N(x) = \sum_{m=0}^{\infty} \hat{f}_m \Theta_N^{(m)}(x) , \tag{30}$$

where

$$\Theta_N^{(m)}(x) = \sum_{n=0}^{N} \theta_n^{(m)}(x) \tag{31}$$

and

$$\theta_n^{(m)}(x) = \mathscr{A}^n \psi_m(x) r_T(x) . \tag{32}$$

In order for Eq. (30) to converge to f(x) via Eq. (A4), we must have

$$\lim_{N \to \infty} \Theta_N^{(m)}(x) = \psi_m(x) . \tag{33}$$

Our purpose in this section is to examine the nature of this convergence.

We can write $\theta_n^{(m)}$ as

$$\theta_n^{(m)}(x) = \left[a_n^{(m)} r_T(x) + b_n^{(m)} \right] \psi_m(x) , \tag{34}$$

where $b_n^{(m)}$ is the contribution of $\psi_m(x)$ to the estimate of the tails of f(x) in the nth iteration, and $a_n^{(m)} + b_n^{(m)}$ is the corresponding residual value contained in the $|x| \le a/2$ interval. As is shown in Appendix C, the performance of the algorithm is completely determined by $a_1^{(m)}$ and $b_1^{(m)}$. For example, we show that, in order for the algorithm to converge, it is necessary that the roots of the polynomial

$$D(z) = z^2 - \left(a_1^{(m)} + 1 \right) z + \left(a_1^{(m)} + b_1^{(m)} \right) \tag{35}$$

lie within the unit circle, $|z| < 1$.

Let us examine the nature of convergence more closely. Substituting Eq. (34) into Eq. (31), we obtain

$$\Theta_N^{(m)}(x) = \left[A_N^{(m)} r_T(x) + B_N^{(m)} \right] \psi_m(x) , \tag{36}$$

where

$$\begin{bmatrix} A_N^{(m)} \\ \\ B_N^{(m)} \end{bmatrix} = \sum_{n=0}^{N} \begin{bmatrix} a_n^{(m)} \\ \\ b_n^{(m)} \end{bmatrix} . \tag{37}$$

To be consistent with Eq. (33), we require

$$\lim_{N \to \infty} \begin{bmatrix} A_N^{(m)} \\ \\ B_N^{(m)} \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix} . \tag{38}$$

This significance of this relationship is made clearer by substituting Eq. (36) into Eq. (30):

$$f_N(x) = \sum_{m=0}^{\infty} \left[A_N^{(m)} r_T(x) + B_N^{(m)} \right] \hat{f}_m \psi_m(x) . \tag{39}$$

As is shown in Appendix C, these coefficients can be found from

$$\begin{bmatrix} A_N^{(m)} \\ \\ B_N^{(m)} \end{bmatrix} \leftrightarrow \begin{bmatrix} \dfrac{z^2}{D(z)} \\ \\ \dfrac{z^2 b_1^{(m)}}{(z-1)D(z)} \end{bmatrix} , \tag{40}$$

where the double arrow denotes a conventional unilateral z-transform pair[38]:

$$c_N \leftrightarrow \sum_{N=0}^{\infty} c_N z^{-N} . \tag{41}$$

Consider some specific examples:

1) For the $\mathscr{C}$ in Eq. (22b), we have Gerchberg's algorithm, and, from Eq. (A2),

$$\mathscr{A} \psi_m(x) r_T(x) = [1 - r_T(x)] \lambda_m \psi_m(x) .$$

Thus, $a_1^{(m)} = -\lambda_m$ and $b_1^{(m)} = \lambda_m$. Since $|\lambda_m| < 1$, the roots of D(z) lie within the unit circle and the algorithm converges. From Eq. (40), we have

$$\begin{bmatrix} A_N^{(m)} \\ \\ B_N^{(m)} \end{bmatrix} \leftrightarrow \begin{bmatrix} \dfrac{z}{z-(1-\lambda_m)} \\ \\ \dfrac{\lambda_m z}{(z-1)[z-(1-\lambda_m)]} \end{bmatrix} .$$

Since

$$c^n \leftrightarrow \frac{z}{z-c} ,$$

we obtain, after a partial fraction expansion,

$$\begin{bmatrix} A_N^{(m)} \\ \\ B_N^{(m)} \end{bmatrix} = \begin{bmatrix} (1-\lambda_m)^N \\ \\ 1-(1-\lambda_m)^N \end{bmatrix} , \tag{42}$$

a result obtained in a different way by Papoulis.[20]

2) For the $\mathscr{C}$ in Eq. (22d), we have

$$\mathscr{A} \psi_m(x) r_T(x) = [\lambda_m - r_T(x)] \psi_m(x) .$$

Thus, $a_1^{(m)} = -1$ and $b_1^{(m)} = \lambda_m$. The algorithm converges. After some computation, we find that, for odd N,

$$B_N^{(m)} = 1 - (1-\lambda_m)^{\frac{N+1}{2}} ,$$

which, as required, converges to unity, but more slowly than Eq. (42).

3) For the $\mathscr{C}$ in Eq. (22c), we obtain $a_1^{(m)} = 1 - \lambda_m$ and $b_1^{(m)} = 0$. $B_N^{(0)} = 0$, and there is a zero estimate for $|x| \geq a/2$. This algorithm, however, displays a different type of convergence. Note that

$$A_N^{(m)} \leftrightarrow \frac{z^2}{(z-1)[z-(1-\lambda_m)]} ,$$

from which we obtain

$$A_N^{(m)} = \frac{1}{\lambda_m} \left[1-(1-\lambda_m)^{N-1} \right] .$$

Substituting into Eq. (39) and filtering gives

$$\mathscr{B}_\Omega f_N(x) = \sum_{m=0}^{\infty} \left[1-(1-\lambda_m)^{N-1} \right] \hat{f}_m \psi_m(x) ,$$

which tends to f(x) as $N \to \infty$. This algorithm thus iteratively builds a signal on $|x| \leq a/2$ that, when filtered, gives the Nth estimate of the extrapolated signal.[21]

4) For $\mathscr{C}$ in Eq. (22e), we have $a_1^{(m)} = 1$ and $b_1^{(m)} = -\lambda_m$. The roots of D(z) lie outside of the unit circle. The corresponding iterative algorithm thus diverges.

3.2. Algorithm stability

It is instructive to examine the stability (or posedness) of the deconvolution algorithms of the previous section. In the following stability analysis, we draw freely from the work of Youla.[19] An algorithm is said to be globally stable if the supremum (least upper bound) of the ratio of the rms output to input error can be bound. If $\mathscr{D}$ exists, it is linear. The output error $\Delta f(x)$ due to an input perturbation $\Delta g(x)$ is thus

$$\Delta f(x) = \mathscr{D} \Delta g(x) . \tag{43}$$

Define the L_2 norm by

$$\|h(x)\| \equiv \sqrt{\int_{-\infty}^{\infty} |h(x)|^2 dx} .$$

Then, the desired quantity to be bound is

$$\sup \frac{\|\Delta f(x)\|}{\|\Delta g(x)\|} < \infty ,$$

where "sup" denotes the supremum. This relation, however, is simply the definition of the operator norm[39]:

$$\|\mathscr{D}\| \equiv \sup \frac{\|\mathscr{D}\,\Delta g(x)\|}{\|\Delta g(x)\|} < \infty . \tag{44}$$

Thus, if $\mathscr{D}$ is a bounded operator, deconvolution is globally stable. If $\mathscr{D}$ is bounded, it is also continuous.

A second useful stability criterion arises from the following analysis:

$$\begin{aligned} \|\Delta g(x)\| &= \|\mathscr{C}\,\Delta f(x)\| \\ &= \|\Delta f(x) - \mathscr{A}\,\Delta f(x)\| \\ &\geq \|\Delta f(x)\| - \|\mathscr{A}\,\Delta f(x)\| \\ &\geq \|\Delta f(x)\|[1-\|\mathscr{A}\|] , \end{aligned} \tag{45}$$

where, in the second step we have used the triangle inequality and, in the last step, the definition of the operator norm. From Eqs. (44) and (45), it follows that

$$\|\mathscr{D}\| \leq \frac{1}{1-\|\mathscr{A}\|} .$$

Thus, if

$$\|\mathscr{A}\| < 1, \tag{46}$$

we are also assured of global stability. Note that this criterion is also a sufficient condition for the geometric series expansion in Eq. (26).

We now illustrate application of Eq. (46) to the case where f(x) is known over the interval $|x| \leq a/2$ and Ω is given by $|u| \leq W$. From Appendix A,

$$\begin{aligned} \mathscr{A} f(x) &= [1-r_T(x)]f(x) \\ &= [1-r_T(x)] \sum_{n=0}^{\infty} \hat{f}_n \psi_n(x) . \end{aligned}$$

Thus,

$$\|\mathscr{A}\| = \sup_{\|f(x)\|=1} \|\mathscr{A} f(x)\|$$

$$= \sup_{\|f(x)\|=1} \sqrt{\sum_{n=0}^{\infty} (1-\lambda_n)|\hat{f}_n|^2} . \tag{47}$$

Since, from Parseval's theorem,

$$\|f(x)\|^2 = \sum_{n=0}^{\infty} |\hat{f}_n|^2 ,$$

the sup in Eq. (47) occurs when $|\hat{f}_\infty|^2 = 1$, which gives $\lambda_\infty = 0$. Thus,

$$\|\mathscr{A}\| = 1 ,$$

and the algorithm is globally unstable.

Consider next the case where T consists of all $|x| \geq a/2$ (interpolation). The problem here is to "fill up" the $|x| \leq a/2$ interval. Following the above development, we have

$$\|\mathscr{A}\| = \sup_{\|f(x)\|=1} \sqrt{\sum_{n=0}^{\infty} \lambda_n |\hat{f}_n|^2} ,$$

which occurs for $|\hat{f}_o|^2 = 1$. Thus,[4]

$$\|\mathscr{A}\| = \sqrt{\lambda_o} < 1 ,$$

and the problem is globally stable.

Stability criteria to date have considered the output error over the entire domain.[19,40-42] Empirical results[33] and intuition[43-45] dictate that extrapolation results are better near to where the image is known. Possibly new local stability measures are in order.

3.4. Incorporation of further *a priori* knowledge

In this section, we present two methods by which further object information can be incorporated into the deconvolution algorithms.

First, consider the case where we can place a bound on the object's energy. That is, we know an ϵ such that

$$\int_{-\infty}^{\infty} |f(x)|^2 dx \leq \epsilon^2 . \tag{48}$$

With this we can bound f(x). Since f(x) is bandlimited, it is not affected by filtering:

$$f(x) = \int_{-\infty}^{\infty} f(\xi) w_\Omega(x-\xi) d\xi ,$$

where

$$w_\Omega(x) = \mathscr{F}^{-1}[W_\Omega(u)] .$$

Thus, using Schwarz's inequality,

$$|f(x)|^2 \leq \int_{-\infty}^{\infty} |f(\xi)|^2 d\xi \int_{-\infty}^{\infty} |w_\Omega(x-\xi)|^2 d\xi .$$

From Parseval's theorem,

$$\int_{-\infty}^{\infty} |w_\Omega(\xi)|^2 d\xi = \int_{-\infty}^{\infty} |W_\Omega(u)|^2 du ,$$

or, since $W_\Omega = W_\Omega{}^2$, we have from Eqs. (1c) and (48) the bound

$$|f(x)| \leq \epsilon\sqrt{|\Omega|} \equiv A . \tag{49}$$

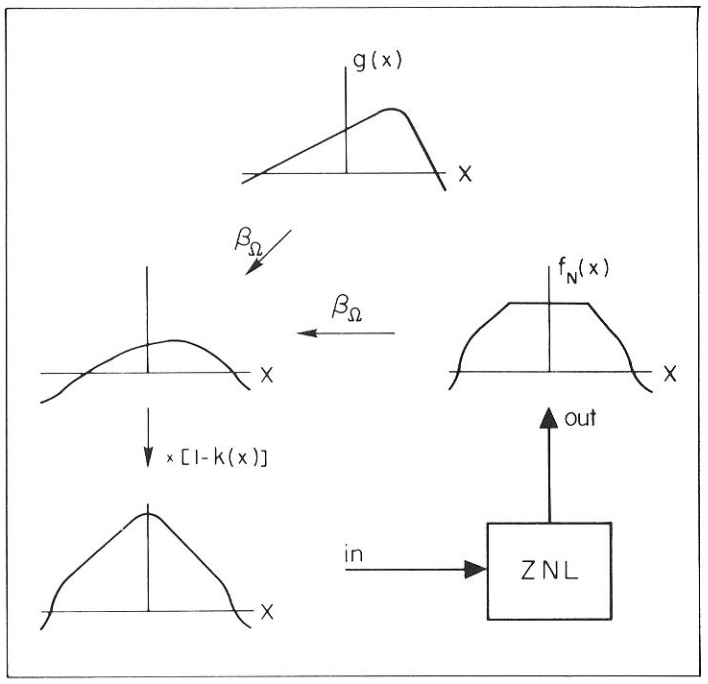


Fig. 18. Gerchberg's iterative algorithm with a zero-memory nonlinearity (ZNL). Compare with Fig. 2.

This bound can be placed in the iterative algorithm as a zero-memory nonlinearity (ZNL), as shown in Fig. 18 for Gerchberg's iterative algorithm. The ZNL corresponding to Eq. (49) is shown in Fig. 19(a). If, in addition, we know that the object in positive, the ZNL in Fig. 19(b) would be appropriate. Rushford and Frost[16] used the ZNL in Fig. 19(c), which corresponds to knowledge of a positive target function and no energy bound. Due to the nonlinear ZNL, this algorithm alteration is not placed in closed form.

A second algorithm alteration is possible when a portion of the object's spectrum is known. In this case, one merely inserts the known portion of the spectrum in each iteration as well as performing the bandlimiting truncation. This algorithm has been iteratively implemented on tomographic data by Sato et al.[46] Using previous notations, the algorithm can also easily be reformulated in closed form. Stark et al.[47-48] have recently shown that finite-energy *non*-bandlimited images can be restored in many cases from limited spatial and spectral information.

Other methods of incorporating *a priori* knowledge into iterative restoration algorithms are presented in the chapter by Fienup.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of this work by the National Science Foundation under grant ENG 79 08009.

APPENDIX A: PROLATE SPHEROIDAL WAVE FUNCTIONS

Let T be the interval defined by $|x| \leq a/2$ and Ω the interval defined by $|u| \leq W$. The solutions to the integral equation

$$\lambda_m \psi_m(x) = 2W \int_{-a/2}^{a/2} \psi_m(\xi) \mathrm{sinc} 2W(x-\xi) d\xi \tag{A1}$$

are prolate spheroidal wave functions as normalized by Slepian and Pollak.[3] An equivalent expression for Eq. (A1) is

$$\lambda_m \psi_m(x) = \mathcal{B}_\Omega \psi_m(x) r_T(x) . \tag{A2}$$

Since $\psi_m(x)$ is obviously bandlimited, we also have

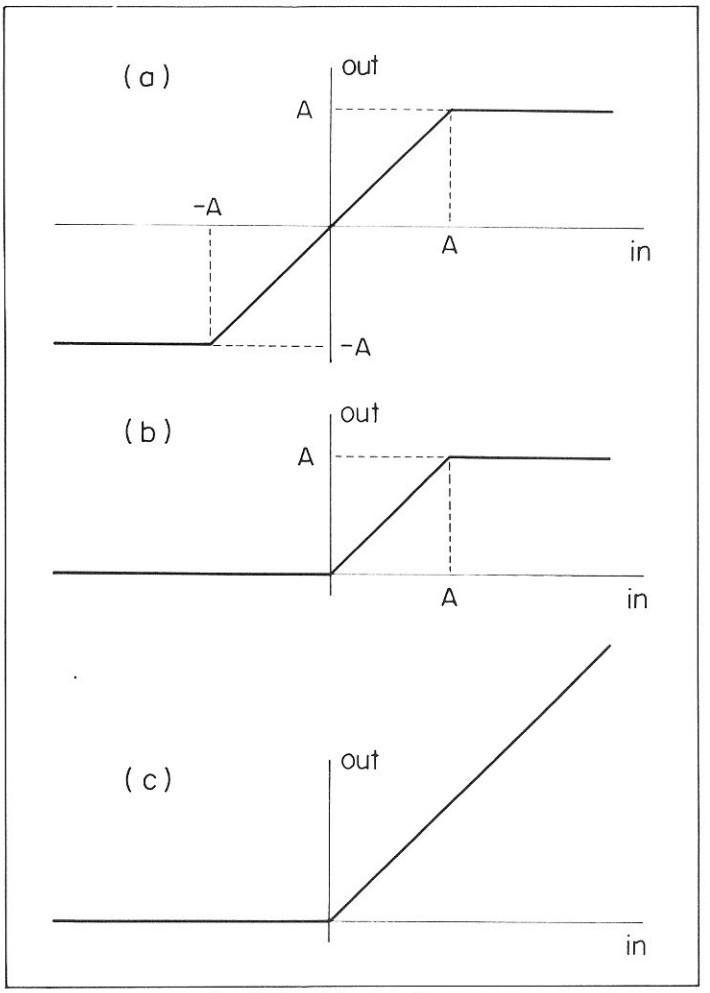


Fig. 19. Some possibe ZNLs: (a) when the image energy is known; (b) same, but with positivity constraint; and (c) with positivity constraint only.

$$\psi_m(x) = \mathcal{B}_\Omega \psi_m(x) . \tag{A3}$$

A bandlimited function f(x) with bandwidth interval Ω can be written as

$$f(x) = \sum_{m=0}^{\infty} \hat{f}_m \psi_m(x) , \tag{A4}$$

where

$$\hat{f}_m = \int_{-\infty}^{\infty} f(x)\psi_m(x)dx$$

$$= \frac{1}{\lambda_m} \int_{|x| \leq a/2} f(x)\psi_m(x)dx \tag{A5}$$

$$= \frac{1}{1-\lambda_m} \int_{|x| > a/2} f(x)\psi_m(x)dx .$$

The eigenvalues are positive and real and ordered as

$$1 > \lambda_o > \lambda_1 > \ldots > 0 . \tag{A6}$$

The functions are orthonormal on $(-\infty,\infty)$:

$$\int_{-\infty}^{\infty} \psi_n(x)\psi_m(x)dx = \delta_{n-m} ,$$

and orthogonal on T and its complement:

$$\int_{-a/2}^{a/2} \psi_n(x)\psi_m(x)dx = \lambda_n\delta_{n-m} ;$$

$$\int_{|x|\geq a/2} \psi_n(x)\psi_m(x)dx = (1-\lambda_n)\delta_{n-m} ,$$

where δ_n denotes the Kronecker delta.

APPENDIX B: ABSORPTIVE LOSS ANALYSIS IN A COHERENT OPTICAL EXTRAPOLATOR

Here we derive Eq. (4) and the corresponding expression for t_m. Equation (3) can be written as

$$\bar{f}_N(x) = \sum_{n=0}^{N} \{p[1-r_T(x)]\,\mathscr{B}_\Omega\}^n g_T(x) . \tag{B1}$$

Using the prolate spheroidal wave function expansion, we have

$$g_T(x) = r_T(x) \sum_{m=0}^{\infty} \hat{f}_m\psi_m(x),$$

$$[1-r_T(x)]\,\mathscr{B}_\Omega g_T(x) = [1-r_T(x)] \sum_{m=0}^{\infty} \lambda_m\hat{f}_m\psi_m(x) ,$$

$$\{[1-r_T(x)]\,\mathscr{B}_\Omega\}^2 g_T(x) = [1-r_T(x)] \sum_{m=0}^{\infty} \lambda_m(1-\lambda_m)\hat{f}_m\psi_m(x) ,$$

$$\vdots$$

$$\{[1-r_T(x)]\,\mathscr{B}_\Omega\}^n g_T(x) = [1-r_T(x)] \sum_{m=0}^{\infty} \lambda_m(1-\lambda_m)^{n-1}\hat{f}_m\psi_m(x) .$$

As $N \rightarrow \infty$ Eq. (B1) becomes

$$\bar{f}(x) = \lim_{N\rightarrow\infty} \bar{f}_N(x)$$

$$= g_T(x) + [1-r_T(x)] \sum_{n=1}^{\infty} p^n \left[\sum_{m=0}^{\infty} \lambda_m(1-\lambda_m)^{n-1}\hat{f}_m\psi_m(x) \right] .$$

Reversing summation order and applying a geometric series gives

$$\bar{f}(x) = g_T(x) + [1-r_T(x)] \sum_{m=0}^{\infty} \frac{p\lambda_m\hat{f}_m\psi_m(x)}{1-p(1-\lambda_m)} . \tag{B2}$$

Note that, for the ideal case of $p = 1$, the summation corresponds to the expansion of $f(x)$ in Eq. (A4), and $\bar{f}(x) = f(x)$. This amounts to Papoulis' proof of Gerchberg's algorithm.[20]

The deviation of $\bar{f}(x)$ from $f(x)$ is measured by

$$\epsilon_A^2 = \int_{-\infty}^{\infty} |f(x)-\bar{f}(x)|^2 dx$$

$$= \sum_{m=0}^{\infty} t_m|\hat{f}_m|^2 ,$$

where we have used Eqs. (A4) and (B2) and

$$t_m = \frac{(1-p)^2(1-\lambda_m)}{[1-p(1-\lambda_m)]^2} . \tag{B3}$$

Note, for $p = 1$, $t_m \equiv 0$. The plot in Fig. 9 was generated using the numerical values of λ_m given by Slepian and Sonnenblick.[7] Note that for the unit-step approximation in Eq. (6), Eq. (B3) becomes Eq. (5).

APPENDIX C: ITERATIVE EXTRAPOLATION ALGORITHM CONVERGENCE

Here we derive the convergence expressions for iterative extrapolation algorithms, specifically Eqs. (35) and (40). We begin with examination of a recursion relation for $\theta_n^{(m)}$ in Eq. (32):

$$\theta_{n+1}^{(m)}(x) = \mathscr{A}\theta_n^{(m)}(x)$$

$$= \mathscr{A}\left[a_n^{(m)}r_T(x) + b_n^{(m)} \right]\psi_m(x) . \tag{C1}$$

Using Eqs. (20) and (25) with $f = \psi_m$ gives

$$\mathscr{A}\psi_m(x) = [1-r_T(x)]\psi_m(x) .$$

Equation (C1) thus becomes

$$\theta_{n+1}^{(m)}(x) = a_n^{(m)}\,\mathscr{A}\,r_T(x)\psi_m(x)$$

$$+ b_n^{(m)}[1-r_T(x)]\psi_m(x)$$

$$= a_n^{(m)}\left[a_1^{(m)}r_T(x) + b_1^{(m)} \right]\psi_m(x) \tag{C2}$$

$$+ b_n^{(m)}[1-r_T(x)]\psi_m(x) ,$$

where we have used Eqs. (32) and (34) with $n = 1$. Also,

$$\theta_{n+1}^{(m)}(x) = \left[a_{n+1}^{(m)}r_T(x) + b_{n+1}^{(m)} \right]\psi_m(x) . \tag{C3}$$

Equating coefficients of $r_T\psi_m$ in Eqs. (C2) and (C3) and then the coefficients of ψ_m, yields the following recurrence relationship for the coefficients:

$$\begin{bmatrix} a_{n+1}^{(m)} \\ b_{n+1}^{(m)} \end{bmatrix} = \begin{bmatrix} a_1^{(m)} & -1 \\ b_1^{(m)} & 1 \end{bmatrix} \begin{bmatrix} a_n^{(m)} \\ b_n^{(m)} \end{bmatrix}. \quad \text{(C4)}$$

Note that, from Eqs. (32) and (34), we obtain the following initial conditions for n = 0:

$$\begin{bmatrix} a_o^{(m)} \\ b_o^{(m)} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The coupled difference equations in Eq. (C4) can be solved by conventional Z-transform methods.[38] Using the transform definition in Eq. (41), with n instead of N, we obtain

$$\begin{bmatrix} a_n^{(m)} \\ b_n^{(m)} \end{bmatrix} \leftrightarrow \begin{bmatrix} \dfrac{z(z-1)}{D(z)} \\ zb_1^{(m)}/D(z) \end{bmatrix}, \quad \text{(C5)}$$

where D(z)—the characteristic equation of the matrix in (C4)—is given in Eq. (35). Note that, since

$$\begin{bmatrix} \dfrac{z(z-1)}{D(z)} \\ \dfrac{zb_1^{(m)}}{D(z)} \end{bmatrix} = \sum_{n=0}^{\infty} \begin{bmatrix} a_n^{(m)} z^{-n} \\ b_n^{(m)} z^{-n} \end{bmatrix},$$

we have, from Eq. (37),

$$\lim_{N\to\infty} \begin{bmatrix} A_N^{(m)} \\ B_N^{(m)} \end{bmatrix} = \left. \begin{bmatrix} \dfrac{z(z-1)}{D(z)} \\ \dfrac{zb_1^{(m)}}{D(z)} \end{bmatrix} \right|_{z=1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{(C6)}$$

which is consistent with Eq. (38). Equation (C6), however, assumes that the transforms converge at z = 1. In general, the Z-transform of a causal sequence converges outside of a circle. The circle's radius in this case is determined by the pole furthest from the origin. Thus, in order for Eq. (C6) to be valid, all of the zeroes of D(z) must lie within the unit circle. Otherwise, the algorithm diverges.

The Z-transform pair

$$\sum_{n=0}^{N} c_n \leftrightarrow \frac{z}{z-1} \sum_{N=0}^{\infty} c_N z^{-N}$$

follows from Eq. (41). Application to Eq. (C4) yields, via Eq. (37), the Z-transform pairs in Eq. (40).

REFERENCES

1. G. C. Temes, V. Barcilon, and F. C. Marshall, III, "The optimization of bandlimited systems," Proc. IEEE 61, 196-234 (1973).
2. Y. I. Khurzin and V. P. Yakovev, "Progress in the Soviet Union on the theory and applications of bandlimited functions," Proc. IEEE 65, 1005-1029 (1977).
3. D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty I," Bell Syst. Tech. J. 40, 43-63 (1961).
4. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty II," Bell Syst. Tech. J. 40, 65-84 (1961).
5. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty III: The dimension of the space of essentially time- and band-limited signals," Bell Syst. Tech. J. 41, 1295-1336 (1962).
6. D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty IV: Extensions to many dimensions; generalized prolate spheroidal wave functions," Bell Syst. Tech. J. 43 3009-3057 (1964).
7. D. Slepian and E. Sonnenblick, "Eigenvalues associated with prolate spheroidal wave functions of zero order," Bell Syst. Tech. J. 44, 1745-1758 (1965).
8. D. Slepian, "Some asymptotic expansions for prolate spheroidal wave functions," J. Math. Phys. 44, 99-140 (1965).
9. G. Toraldo di Francia, "Degrees of freedom of an image," J. Opt. Soc. Am. 59, 799-804 (1969).
10. B. R. Frieden, "Restoring with maximum likelihood and maximum entropy," J. Opt. Soc. Am. 62, 511-518 (1972).
11. B. R. Frieden, "Image Enhancement and Restoration," in *Topics in Applied Physics,* Edited by T. S. Huang, Vol. 6, pp. 177-248, Springer-Verlag, New York (1975).
12. B. R. Frieden and D. C. Wells, "Restoring with maximum entropy III: Poisson sources and backgrounds," J. Opt. Soc. Am. 68, 93-103 (1978).
13. B. R. Frieden, "Image restoration using a norm of maximum information," Opt. Eng. 19(3), 290-296 (1980).
14. S. J. Howard, "Method for continuing Fourier spectra given by the fast Fourier transform," J. Opt. Soc. Am. 71, 95-98 (1981); Erratum, 71, 614 (1981).
15. S. J. Howard, "Continuation of discrete Fourier spectra using a minimum negativity constraint," J. Opt. Soc. Am. 71, 819-824 (1981).
16. C. K. Rushford and R. L. Frost, "Comparison of some algorithms for reconstructing space-limited images," J. Opt. Soc. Am. 70, 1539-1544 (1980).
17. R. W. Gerchberg, "Super-resolution through error energy reduction," Opt. Acta 21, 709-720 (1974).
18. J. A. Villa, "sur le prolongement des signaux a spectre born," Cables et Transmission no. 1, 44-52 (1956).
19. D. C. Youla, "Generalized image restoration by method of alternating orthogonal projections," IEEE Trans. Circuits and Systems CAS-25, 694-702 (1978).
20. A. Papoulis, "A new algorithm in spectral analysis and bandlimited signal extrapolation," IEEE Trans. Circuits and Systems CAS-22, 735-742 (1975).
21. J. A. Cadzow, "An extrapolation procedure for band-limited signals," IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-27, 4-12 (1979).
22. M. S. Sabri and W. Steenaart, "Comments on 'An extrapolation procedure for bandlimited signals,'" IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-28, 254 (1980).
23. P. H. Van Cittert, "Zum Einfluss der Splaltbreite auf die Intensitätswerteilung in Spektrallinien II," Z. Für Phys 69, 298-308 (1931).
24. N. R. Hill and G. E. Ioup, "Convergence of the van Cittert iterative method of deconvolution," J. Opt. Soc. Am. 66, 487-489 (1976).
25. D. E. Dudgeon, "An iterative implementation for 2-D digital filters," IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-28, 666-671 (1980).
26. R. J. Marks, II, "Coherent optical extrapolation of 2-D band-limited signals: processor theory," Appl. Opt. 19, 1670-1672 (1980).
27. R. J. Marks, II and D. K. Smith, "An iterative coherent processor for bandlimited signal extrapolation," Proc. SPIE 231, 106-111 (1980).
28. J. Cederquist and S. H. Lee, "Confocal feedback systems with space-variance, time sampling, and secondary feedback loops," J. Opt. Soc. Am. 71, 643-650 (1981).
29. D. Smith, "Extrapolation of 2-D bandlimited images," Thesis, University of Washington, Department of Electrical Engineering (1980).
30. B. R. Frieden, "The extrapolating pupil, image synthesis, and some thought applications," Appl. Opt. 9, 2489-2496 (1970).
31. M. S. Sabri and W. Steenaart, "An approach to bandlimited signal extrapolation: the extrapolation matrix," IEEE Trans. Circuits and Systems CAS-25, 74-78 (1978).
32. R. J. Marks, II, "Gerchberg's extrapolation algorithm in two dimensions," Appl. Opt. 20, 1815-1820 (1981).
33. D. K. Smith and R. J. Marks, II, "Closed form bandlimited image extrapolation," Appl. Opt. 20, 2476-2483 (1981).
34. M. S. Sabri and W. Stenaart, "Discrete Hilbert transform filtering," IEEE Trans. Acoustics, Speech and Signal Processing ASSP-25, 452-454 (1977).

35. K. C. Tam and V. Perez-Mendez, "Tomographical imaging with limited-angle input," J. Opt. Soc. Am. 71, 582-592 (1981).
36. T. K. Sarkar, D. D. Weiner, and V. K. Jain, "Some mathematical considerations in dealing with the inverse problem," IEEE Trans. Antennas Propag. AP-29, 373-379 (1981).
37. R. Prost and R. Goutte, "Deconvolution when the convolution kernel has no inverse," IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-25, 542-548 (1977).
38. A. Papoulis, *Circuits and Systems—A Modern Approach,* Holt, Rinehart and Winston, New York (1980).
39. D. G. Luenberger, *Optimization by Vector Space Methods,* Chap. 6.2, John Wiley & Sons, New York (1969).
40. G. A. Viano, "On the extrapolation of optical image data," J. Math Phys. 17, 1160-1165 (1976).
41. R. G. Wiley,"On an iterative technique for recovery of bandlimited signals," IEEE Trans. Comm. COM-66, 522-523 (1978).
42. R. G. Wiley, "Concerning the recovery of a bandlimited signal or its spectrum from a finite segment," IEEE Trans. Comm. COM—27, 251-252 (1979).
43. W. K. Pratt, *Digital Image Processing,* pp. 437-440, John Wiley and Sons, New York (1968).
44. C. Pask, "Simple optical theory of super-resolution," J. Opt. Soc. Am. 66, 68-69 (1976).
45. D. Slepian, "On bandwidth," Proc. IEEE 64, 292-300 (1973).
46. T. Sato, S. J. Norton, M. Linzer, O. Ikeda, and M. Hirama, "Tomographic image reconstruction from limited projections using iterative revisions in image and transform spaces," Appl. Opt. 20, 395-399 (1981).
47. H. Stark, D. Cahana, and H. Webb, "Restoration of arbitrary finite-energy optical objects from limited spatial and spectral information," J. Opt. Soc. Am. 71, 635-642 (1981).
48. H. Stark, D. Cahana, and G. J. Habetler, "Is it possible to restore an optical object from its low pass spectrum and its truncated image?," Opt. Lett. 6, 259-260 (1981).

Note in Print: Since submission of this chapter, a wealth of new publications have appeared on our subject or have been brought to the author's attention. Of particular interest is the tutorial in Ref. 50 and the independent proof of Gerchberg's algorithm using prolate spheroidal wave functions in Ref. 49. This latter paper was submitted within a month of the more widely referenced Ref. 20.

49. P. DeSantis and F. Gori, "On an iterative method for super-resolution," Optica Acta 22, 691-695 (1975).
50. R. W. Schafer, R. M. Mersereau and M. A. Richards, "Constrained iterative restoration algorithms," Proc. IEEE 69, 432-450 (1981).
51. G. Thomas, "A modified version of Van Cittert's iterative deconvolution procedure," IEEE Trans. Acoust. Speech and Signal Processing ASSP-29, 938 (1981).
52. T. Sato, K. Saski, Y. Nakamura, M. Linzer and S. J. Norton," Tomographic image reconstruction from limited projections using coherent optical feedback," Appl. Opt. 20, 3073-3076 (1981).
53. H. Maitre, "Iterative superresolution: Some new fast methods," Optica Acta 28, 973-980 (1981).
54. A. Lent and H. Tuy, "An iterative method for the extrapolation of band-limited functions," J. Math. Anal. Appl. 83, 554-565 (1981).
55. V. T. Tom, T. F. Quatieri, M. H. Hayes and J. H. McClellan, "Convergence of iterative nonexpansive signal reconstruction algorithms," IEEE Trans. Accoust. Speech and Signal Processing ASSP-29, 1052-1058 (1981).
56. R. J. Marks, II, and M. J. Smith, "Closed-form object restoration from limited spatial and spectral information," Opt. Lett. 6, 522-524 (1981).
57. J. A. Cadzow, "Observations on the extrapolation of a band-limited signal," IEEE Trans. Acoust. Speech and Signal Processing ASSP-29, 1208-1209 (1981).
58. M. S. Sabri and W. Steenaart, "Rebuttal to 'Observations on the extrapolation of a band-limited signal problem,'" IEEE Trans. Acoust. Speech and Signal Processing ASSP-29, 1209 (1981).
59. D. Cahana and H. Stark, "Bandlimited image extrapolation with faster convergence," Appl. Opt. 20, 2780-2786 (1981).
60. M. A. Fiddy and T. J. Hall, "Nonuniqueness of superresolution techniques applied to sampled data," J. Opt. Soc. Am. 71, 1406-1407 (1981).
61. A. K. Jain and S. Ranganath, "Extrapolation algorithms for discrete signals with application in spectral estimation," IEEE Trans. Acoust. Speech and Signal Processing ASSP-29, 830-845 (1981).
62. H. Stark, S. Cruze and G. Habetler, "Restoration of optical objects subject to nonnegative spatial or spectral constraints," J. Opt. Soc. Am. 72, 993-1000 (1982).
63. T. F. Quatieri and D. E. Dudgeon, "Implementation of 2-D digital filters by iterative methods," IEEE Trans. Acoust. Speech and Signal Processing ASSP-30, 473-487 (1982).
64. A. Sonnenschein and B. W. Dickinson, "On a recent extrapolation procedure for band-limited signals," IEEE Trans. Circuits and Systems CAS-22, 116-117 (1982).
65. W. D. Montgomery, "Optical applications of von Neuman's alternating-projection theorem," Opt. Lett. 7, 1-3 (1982).
66. R. J. Marks, II, "Posedness of a band-limited image extention problem in tomography," Opt. Lett. 7, 376-377 (1982).
67. Y. Yamakoshi and T. Sato, "Iterative image restoration from data available in multiple restricted regions," Appl. Opt. 21, 4473-4480 (1982).
68. D. Kaplan and R. J. Marks, II, "Noise sensitivity of interpolation and extrapolation matrices," Appl. Opt. 21, 4489-4492 (1982).