The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer

B. R. HUNT

Abstract—Constrained least squares estimation is a technique for solution of integral equations of the first kind. The problem of image restoration requires the solution of an integral equation of the first kind. However, application of constrained least squares estimation to image restoration requires the solution of extremely large linear systems of equations. In this paper we demonstrate that, for convolution-type models of image restoration, special properties of the linear system of equations can be used to reduce the computational requirements. The necessary computations can be carried out by the fast Fourier transform, and the constrained least squares estimate can be constructed in the discrete frequency domain. A practical procedure for constrained least squares estimation is presented, and two examples are shown as output from a program for the CDC 7600 computer which performs the constrained least squares restoration of digital images.

Index Terms—Block circulant matrices, constrained estimation, diagonalization of block circulant matrices, Fourier transforms, image restoration, integral equations, least squares estimation.

I. Introduction

IN A RECENT special issue of the IEEE Proceedings devoted to digital picture processing, a paper by Sondhi [1] reviewed the problem of image restoration as being equivalent to solving a Fredholm integral equation of the first kind. The solution of such equations is made difficult by their inherent ill-conditioned behavior [2]. Sondhi called attention to a method of Phillips as an approach to the solution of such equations that arise in image restoration. However, the detailed application of Phillips' method to image restoration was left unresolved. In this paper we demonstrate the successful application of Phillips' method in image restoration.

II. Discrete Restoration Model

The fundamental model describing the image restoration problem is given by [3]

$$g(x,y) = \int_0^y \int_0^x h(x - x_1, y - y_1) f(x_1, y_1)\, dx_1\, dy_1 + \epsilon(x,y) \quad (1)$$

where g is the recorded or degraded image, h is the *point-spread function* of the image formation system, f is the *ideal image*, and ϵ is a two-dimensional random noise process. We are assuming that the indicated images are defined over finite rectangular intervals, where the origin of coordinates is positioned such that

$$\begin{aligned} f(x,y), &\quad \text{for } 0 \leqslant x \leqslant A \\ &\qquad\quad 0 \leqslant y \leqslant B \\ h(x,y), &\quad \text{for } 0 \leqslant x \leqslant C \\ &\qquad\quad 0 \leqslant y \leqslant D \end{aligned}$$

where these definitions imply that

$$\begin{aligned} g(x,y), &\quad \text{for } 0 \leqslant x \leqslant A + C \\ &\qquad\quad 0 \leqslant y \leqslant B + D. \end{aligned}$$

The discrete analogy is formed by constructing an appropriate quadrature over the integrals in (1). Equal interval quadrature is desired since image digitizing devices are usually constructed to sample images on a mesh of equal spaced points. The "rectangular rule" is the quadrature we choose and the discrete approximation to (1) has the form

$$g(j\Delta x, k\Delta y) \cong \Delta x \Delta y \sum_{m=0}^{j} \sum_{n=0}^{k} h((j-m)\Delta x, (k-n)\Delta y) \times f(m\Delta x, n\Delta y) + \epsilon(j\Delta x, k\Delta y). \quad (2)$$

We assume $\Delta x = \Delta y = 1$ and represent this approximation in the more compact notational form

$$g(j,k) = \sum_{m=0}^{j} \sum_{n=0}^{k} h(j-m, k-n) f(m,n) + \epsilon(j,k). \quad (3)$$

The matrices h, f, ϵ, and g are constructed from sampling the corresponding functions, and are sized as follows: f is a matrix of size $M \times N$; h is a matrix of size $J \times K$; and g and ϵ are matrices of size $M + J - 1 \times N + K - 1$.

The digital image restoration problem is posed in terms of this model: to estimate the ideal image matrix f given a recorded image matrix g and a knowledge of the point-spread function matrix h. We assume that our knowledge about ϵ is limited to information of a statistical nature.

III. Vector-Matrix Formulation

The form of the double summations in (3) is that of a two-dimensional convolution, a linear system of equations in the unknowns $f(m, n)$ which must be solved. We will rewrite (3)

Manuscript received October 30, 1972; revised March 30, 1973. This work was supported in part by the U.S. Atomic Energy Commission.
The author is with the Los Alamos Scientific Laboratory, University of California, Los Alamos, N. Mex. 87544.

in an equivalent, but more familiar, vector-matrix form. The steps by which we express (3) as a vector-matrix relation are as follows.

Step 1: Choose $U \geqslant M + J - 1$ and $V \geqslant N + K - 1$. Then construct four new *extended* matrices of size $U \times V$, according to the definitions

$$f_e(j,k) = \begin{cases} f(j,k), & \text{for } 0 \leqslant j \leqslant M-1 \text{ and } 0 \leqslant k \leqslant N-1 \\ 0, & \text{for } M \leqslant j \leqslant U-1 \text{ or } N \leqslant k \leqslant V-1 \end{cases}$$

$$h_e(j,k) = \begin{cases} h(j,k), & \text{for } 0 \leqslant j \leqslant J-1 \text{ and } 0 \leqslant k \leqslant K-1 \\ 0, & \text{for } J \leqslant j \leqslant U-1 \text{ or } K \leqslant k \leqslant V-1 \end{cases}$$

$$g_e(j,k) = \begin{cases} g(j,k), & \text{for } 0 \leqslant j \leqslant M+J-1 \text{ and } 0 \leqslant k \leqslant N+K-1 \\ 0, & \text{for } M+J \leqslant j \leqslant U-1 \text{ or } N+K \leqslant k \leqslant V-1 \end{cases}$$

$$\epsilon_e(j,k) = \begin{cases} \epsilon(j,k), & \text{for } 0 \leqslant j \leqslant M+J-1 \text{ and } 0 \leqslant k \leqslant N+K-1 \\ 0, & \text{for } M+J \leqslant j \leqslant U-1 \text{ or } N+K \leqslant k \leqslant V-1. \end{cases}$$

Step 2: Create column vectors of length UV by lexicographically ordering the matrices f_e, g_e, and ϵ_e. In this construction, the first row of a matrix becomes the first column partition in a vector, the second row becomes the second partition, etc. Thus

$$f_e = \begin{bmatrix} f_{e0} \\ \hline f_{e1} \\ \hline f_{e2} \\ \hline \vdots \\ f_{e,U-1} \end{bmatrix}$$

where the column partition f_{ej} is formed by transposing the jth row of the matrix $f_e(j,k)$. In a similar fashion we construct column vectors g_e and ϵ_e.

Step 3: Construct a matrix H according to the following rules. H is of size $UV \times UV$ and consists of U^2 partitions, each partition being of size $V \times V$, and ordered according to

$$H = \left[\begin{array}{c|c|c|c} H_0 & H_{U-1} & H_{U-2} \cdots & H_1 \\ \hline H_1 & H_0 & H_{U-1} \cdots & H_2 \\ \hline H_2 & H_1 & H_0 & \cdot \\ \hline \vdots & & \ddots & \vdots \\ \hline H_{U-1} & H_{U-2} & \cdots & H_0 \end{array}\right]. \tag{4}$$

Further, each partition matrix H_j is constructed from the jth row of $h_e(j,k)$ according to

$$H_j = \begin{bmatrix} h_e(j,0) & h_e(j,V-1) & h_e(j,V-2)\cdots & h_e(j,1) \\ h_e(j,1) & h_e(j,0) & h_e(j,V-1)\cdots & h_e(j,2) \\ h_e(j,2) & h_e(j,1) & h_e(j,0) \quad \cdots & \vdots \\ & & \ddots & \\ h_e(j,V-1) & h_e(j,V-2) & \cdots & h_e(j,0) \end{bmatrix}. \tag{5}$$

Step 4: It can be verified directly that (3) is represented as the vector-matrix equation

$$g_e = Hf_e + \epsilon_e. \tag{6}$$

That is, by forming the indicated products in (6) and then unscrambling the lexicographic order, we can see that all the elements of the matrix $g(j,k)$ in (3) are contained in (6) in the vector g_e. If $U > M + J - 1$ or $V > N + K - 1$, the vector g_e contains zero elements in addition to the elements of $g(j,k)$.

Equation (6) makes explicit the linear system that needs to be solved for the unknown elements of the ideal image f. However, the difficulty of solving (6) directly is evident in the extremely large size of the linear system. Let $U = V = 500$, and the system of equations in (6) is of order 250 000. Fortunately, the matrix H has a special structure that will be exploited in a later section of the paper and will lead to a computationally feasible solution.

IV. Constrained Least Squares Estimation

A version of the constrained least squares estimation technique was first formulated by Phillips [4]. Phillips noted that due to the inherent ill conditioning, numerical solutions to integral equations of the first kind are obscured by wildly oscillating noise. The gross oscillations are usually recognized as being not consistent with *a priori* knowledge that the solution is smooth (or at least smoother than the noise oscillations actually obtained). Phillips proposed that the solution should encompass a smoothness measure of some kind and chose the inner product of the second difference of the solution vector [4], which is quantitatively sensitive to the relative smoothness of the solution vector. Thus in the following we will be describing the extension to two dimensions of this technique.

Let $\hat{f}_e$ be an estimate of the solution to (6). Then we derive $\hat{f}_e$ by solving the following problem. Minimize

$$\hat{f}_e' C' C \hat{f}_e \tag{7}$$

subject to

$$(g_e - H\hat{f}_e)'(g_e - H\hat{f}_e) = \epsilon_e' \epsilon_e. \tag{8}$$

The prime is used to denote the transpose of a vector or a matrix. A straightforward Lagrangian minimization yields the solution

$$\hat{f}_e = (H'H + \gamma C'C)^{-1} H' g_e \tag{9}$$

where $\gamma = 1/\lambda$ and λ is a Lagrange multiplier. We will postpone discussion of the determination of γ for several paragraphs.

In the original one-dimensional work of Phillips, the matrix C must be such as to compute the second difference of a vector when multiplied into the vector. Thus, in one dimension, the second-difference matrix is

$$\begin{bmatrix} 1 & & & & & & & \\ -2 & 1 & & & & \emptyset & & \\ 1 & -2 & 1 & & & & & \\ & 1 & -2 & 1 & & & & \\ & & 1 & -2 & 1 & & & \\ & & & & \ddots & & & \\ & & & & 1 & -2 & 1 \\ & \emptyset & & & & 1 & -2 \\ & & & & & & 1 \end{bmatrix}.$$

The second-difference structure used by Phillips in one dimension must be retained or approximated in two dimensions. There are two approaches.

Approach 1—Use of the Laplacian: The discrete Laplacian operator can be expressed as a matrix:

$$C_L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{10}$$

This matrix must be convolved in two dimensions [an operation in the form of (3)] in order to compute the Laplacian operator on a matrix. However, in (9) the matrix H is of size $UV \times UV$, and C_L is of size 3×3. We must construct a matrix from C_L that is of size $UV \times UV$. This can be done in two steps. First, make C_L into an extended matrix C_{Le} of size $U \times V$ by adding rows and columns of zeros, i.e., repeat Step 1 of Section III, using C_L as the matrix to be extended. Second, construct a matrix C of size $UV \times UV$ in the same way that the matrix H was constructed, i.e., repeat Step 3 of Section III, and (4) and (5), using the extended matrix C_{Le} rather than the extended matrix of the point-spread function.

Approach 2—Use of Radial Symmetry: The Laplacian operator possesses the property that it is not radially symmetric in the two-dimensional discrete frequency domain. However, radial symmetry is a property that is often assumed for image formation systems and referred to as *isotropic response*. To construct a radially symmetric representation of the second-difference operator, the following steps should be used. First, choose U or V, whichever is greater, and compute the one-dimensional transform of U or V points of the second-difference sequence $\{1, -2, 1, 0, 0, \cdots, 0\}$. Assuming V is greater, call this transform $C_1(p)$ for $p = 0, 1, 2, \cdots, V-1$. Second, using only the magnitude of the transform $|C_1(p)|$, construct a two-dimensional transform matrix of size $U \times V$ according to the rule

$$C_2(m,n) = |C_1([r])|$$

for $m = 0, 1, 2, \cdots, U/2$, $n = 0, 1, 2, \cdots, V/2$, where

$$r = \left(\frac{V^2}{U^2} m^2 + n^2\right)^{1/2}.$$

$[r]$ denotes the greatest integer less than r; let $[r] = V/2$ if $[r] \geqslant V/2$. This constructs one quadrant of the matrix $C_2(m, n)$. The remaining three quadrants are constructed by reflecting the first quadrant about the midpoint: $(U/2, V/2)$. Third, compute the inverse two-dimensional discrete Fourier transform of $C_2(m, n)$ and let this be the matrix C_s, of size $U \times V$. Use the matrix C_s to construct a matrix C of size $UV \times UV$, as discussed in Step 3 of Section III.

Either of the two approaches just presented results in a matrix which is conformable in size with the matrix H and the vectors f_e and g_e.

The parameter γ must be found, and the value of γ is dependent upon the constraint of (8) being satisfied. Given estimate $\hat{f}_e$ we define the *residual* to be the quantity

$$g_e - H\hat{f}_e = \rho.$$

If it is true that $\hat{f}_e = f_e$, then (6) implies that

$$\rho'\rho = \epsilon_e'\epsilon_e.$$

Of course, in general $\hat{f}_e \neq f_e$; however, a property that we can seek for an estimate is that it have a residual that is characteristic of the noise process, i.e., we seek to satisfy (8). This leads us, finally, to consider how to estimate $\epsilon_e'\epsilon_e$. We see that we can write the unbiased estimate of the variance

$$S_\epsilon^2 = \frac{\epsilon_e'\epsilon_e - \mu_\epsilon^2}{(M+J-1)(N+K-1)}$$

where the denominator reflects the zeros in the extended matrix, and

$$\mu_\epsilon = \frac{\sum_{j,k} \epsilon_e(j,k)}{(M+j-1)(N+K-1)}.$$

Therefore,

$$\epsilon_e'\epsilon_e = (M+J-1)(N+K-1)S_\epsilon^2 + \mu_\epsilon^2 \tag{11}$$

is an estimate for $\epsilon_e'\epsilon_e$. The importance of (11) is that it allows one to establish a value for the residual constraint in terms of an easily measured quantity, the noise variance.

The parameter γ is found by iteration until the constraint of (8) is satisfied, which we discuss later in Section VI.

Superficially, (9) is as difficult to deal with as (6), because the equation represents an extremely large number of equations and unknowns. However, the structure of matrices H and C is of a special kind and can be exploited to reduce the matrix inversion in (9) to a quotient of sequences.

V. Diagonalization of Block Circulants

The matrices H and C, as defined in (4) and (5), have a special structure and are called *block circulant* matrices. A circulant matrix is diagonalized by the discrete Fourier transform [5]. Likewise, as we now demonstrate, a block circulant matrix can be diagonalized by the two-dimensional discrete Fourier transform. In the following we use the notation:

$$w_U^{jm} = \exp\left(\frac{i2\pi}{U}\,jm\right) \qquad w_V^{kn} = \left(\frac{i2\pi}{V}\,kn\right)$$

to symbolize raising to a power the U or V roots of unity. We now define a matrix W that is of size $UV \times UV$, and is partitioned into a total of U^2 partitions of size $V \times V$. The jmth partition of W is symbolized as W_{jm} and is defined as

$$W_{jm} = w_U^{jm}\, W_V$$

where W_V is a $V \times V$ matrix, and the knth element of W_V is

$$\{W_V\}_{kn} = w_V^{kn}$$

for $j, m = 0, 1, 2, \cdots, U-1$ and $k, n = 0, 1, 2, \cdots, V-1$. Likewise, we will make use of the matrix W^{-1}, which has the property

$$WW^{-1} = W^{-1}W = I \tag{12}$$

where I is the identity matrix. We can demonstrate directly that W^{-1} is also of size $UV \times UV$ and is composed of U^2 partitions of size $V \times V$. The jmth partition of W^{-1} is symbolized as W_{jm}^{-1} and is defined as

$$W_{jm}^{-1} = \frac{w_U^{-jm}}{U}\, W_V^{-1}$$

where W_V^{-1} is a $V \times V$ matrix whose knth element is given by

$$\{W_V^{-1}\}_{kn} = \frac{w_V^{-kn}}{V}.$$

Substitution from the definitions of W and W^{-1} into (12) verifies that W and W^{-1} are properly defined and are the inverses of each other.

The matrix W^{-1} has special meaning when multiplied into a lexicographically ordered image vector such as f_e. We use a script letter to denote the product of the vector of f_e with W^{-1}:

$$\mathcal{F}_e = W^{-1} f_e.$$

The vector $\mathcal{F}_e$ is a lexicographic ordering on a transformation of the extended matrix $f_e(j, k)$. In terms of the column partitions, the mth column partition of $\mathcal{F}_e$ is related to the jth column partitions of f_e by

$$\mathcal{F}_{em} = \sum_{j=0}^{U-1} \frac{w_U^{-jm}}{U}\, W_V^{-1} f_{ej}.$$

However, the product $W_V^{-1} f_{ej}$ is an operation on the elements of the jth column partition. In terms of the original extended matrix $f_e(j, k)$, the kth element of f_{ej} is

$$\{f_{ej}\}_k = f_e(j, k).$$

Thus in the transformed vector the nth element of the mth column partition is given as

$$\{\mathcal{F}_{em}\}_n = \sum_{j=0}^{U-1} \frac{w_U^{-jm}}{U} \sum_{k=0}^{V-1} \frac{w_V^{-kn}}{V} f_e(j, k) = \frac{1}{UV} \sum_{j=0}^{U-1} \sum_{k=0}^{V-1} f_e(j, k) \cdot \exp\left(-i2\pi\left(\frac{jm}{U} + \frac{kn}{V}\right)\right) \tag{13}$$

for $j, m = 0, 1, 2, \cdots, U-1$ and $k, n = 0, 1, 2, \cdots, V-1$. This we recognize as the discrete Fourier transform in two dimensions of the original extended matrix $f_e(j, k)$, and $\mathcal{F}_e$ is therefore a lexicographically ordered vector constructed from the matrix $\mathcal{F}_e(m, n)$ which is the two-dimensional transform of $f_e(j, k)$.

We examine the action of W on a block circulant matrix such as H. Thus

$$D = W^{-1}HW$$

where D is of size $UV \times UV$ and is composed of U^2 partitions of size $V \times V$. Let D_{jk} be the jkth partition of D. We can express D_{jk} in terms of the partitions of W, W^{-1}, and the partitions H_j of H

$$D_{jk} = \sum_{m=0}^{U-1} \sum_{n=0}^{U-1} \frac{w_U^{-jm}}{U}\, W_V^{-1} H_{(m-n) \bmod U}\, w_U^{kn} W_V$$

for $j, k = 0, 1, 2, \cdots, U-1$. The modular notation on the subscript of the partitions H_j results from the circulant partition structure seen in (4).

Each partition H_j is a circulant matrix, and can be diagonalized by the matrix W_V [5]. In fact for any j

$$B_j = W_V^{-1} H_j W_V$$

is a diagonal matrix. Since H_j is a circulant constructed from the jth row of the extended matrix $h_e(j, k)$, then we can compute the rrth element (on the diagonal) of B_j as [5]

$$\{B_j\}_{rr} = \sum_{k=0}^{V-1} h_e(j, k) \exp\left(-\frac{i2\pi}{V}\,kr\right). \tag{14}$$

Using the B_j matrices, we can write

$$D_{jk} = \sum_{m=0}^{U-1} \sum_{n=0}^{U-1} B_{(m-n) \bmod U} \frac{w_U^{(kn-jm)}}{U}.$$

From the orthogonality properties of the complex exponential, we see that

$$D_{jk} = \phi, \qquad \text{if } j \neq k$$

where ϕ is the zero matrix. However, in the case where $j = k$, then

$$D_{jj} = \sum_{m=0}^{U-1} \sum_{n=0}^{U-1} B_{(m-n) \bmod U} \frac{w_U^{-j(m-n)}}{U} = \sum_{p=0}^{U-1} B_p \exp\left(-\frac{i2\pi}{U}\,jp\right) \tag{15}$$

where (15) follows from the periodicity in the complex exponential, i.e.,

$$w_U^{-j(m-n)} = w_U^{-[j(m-n)] \bmod U}.$$

Following the definition of (13),

$$\mathcal{H}_e(m,n) = \frac{1}{UV} \sum_{j=0}^{U-1} \sum_{k=0}^{V-1} h_e(j,k) \exp\left(-i2\pi\left(\frac{jm}{U} + \frac{kn}{V}\right)\right)$$

then we can see that D is a diagonal matrix and the elements on the diagonal correspond to a lexicographic order of the transformed matrix $H_e(m,n)$. That is, the kkth element of D is

$$\{D\}_{kk} = UV\mathcal{H}_e\left(\left[\frac{k}{V}\right], k \bmod V\right)$$

for $k = 0, 1, 2, \cdots, UV - 1$, where $[x]$ again denotes the greatest integer less than x. All other elements (off diagonal) of D are zero.

By similar arguments, one can show that the matrix H' is also a block circulant and is diagonalized by W, in the form $\overline{D} = W^{-1}H'W$ where the overbar denotes complex conjugate.

In (9) the matrix C will also be a block circulant, from the way it was constructed in the discussion of the previous section. Therefore, C is also diagonalized by W and we symbolize this as a diagonal matrix E:

$$E = W^{-1}CW$$
$$\overline{E} = W^{-1}C'W.$$

As was the case with H, the elements on the diagonal of E are related to the two-dimensional discrete Fourier transform of the extended matrix as constructed in the previous section:

$$\{E\}_{kk} = UV\mathcal{C}_e\left(\left[\frac{k}{V}\right], k \bmod V\right)$$

where

$$\mathcal{C}_e(m,n) = \frac{1}{UV} \sum_{k=0}^{V-1} \sum_{j=0}^{U-1} c_e(j,k) \exp\left(-i2\pi\left(\frac{jm}{U} + \frac{kn}{V}\right)\right)$$

and $c_e(j, k)$ is an extended matrix for the second difference constructed in one of the two ways discussed in Section IV.

Equation (9) involves the sums, products, and inverses of block circulants, and the results of these operations will also be block circulants. Hence, (9) can be written in the form

$$W^{-1}\hat{f}_e = (D\overline{D} + \gamma E\overline{E})^{-1}\overline{D}W^{-1}g_e. \tag{16}$$

The products $W^{-1}f_e$ and $W^{-1}g_e$ have a special meaning, as proved earlier: the two-dimensional discrete Fourier transform of lexicographically ordered image vectors. Furthermore, in (16) all the indicated matrices are diagonal and products of the diagonal matrices with the vector $W^{-1}g_e$ can be written as a product of sequences. Using script letters to denote discrete Fourier transforms, (16) can be represented in the form

$$\hat{\mathcal{F}}_e(m,n) = \frac{1}{UV} \frac{\overline{\mathcal{H}_e(m,n)}\mathcal{G}_e(m,n)}{(\mathcal{H}_e(m,n)\overline{\mathcal{H}_e(m,n)} + \gamma\mathcal{C}_e(m,n)\overline{\mathcal{C}_e(m,n)})} \tag{17}$$

for $m = 0, 1, 2, \cdots, U - 1$ and $n = 0, 1, 2, \cdots, V - 1$.

In the two-dimensional discrete frequency domain, (17) is equivalent to the space domain form of (9). Equation (9) uses lexicographically ordered image vectors and (17) shows the relationship between the ordered vectors expressed in terms of Fourier transforms of the original unordered matrices. Most important, (17) can be used in place of (9) for great savings in computational requirements. Equation (17) requires only two-dimensional discrete Fourier transforms, which are not difficult on most large-scale scientific computers.

VI. A Practical Algorithm for Constrained Least Squares Estimation

Equation (9) represents only one equation of the pair that is involved in solving the minimization problem of (7) and (8) by Lagrangian methods; the other equation is a nonlinear equation, as a familiarity with Lagrangian methods demonstrates. We avoid solving *simultaneously* a nonlinear equation and the linear equation (9) if it is possible to find γ by iterative techniques. We now demonstrate that this is possible.

Again we consider the residual

$$\rho = g_e - H\hat{f}_e \tag{18}$$

and also the inner product of the residual as a function of γ:

$$\phi(\gamma) = \rho'\rho \tag{19}$$

where $\hat{f}_e$ is as given in (9). We substitute from (9) in (18), differentiate with respect to γ and have

$$\frac{\partial\phi}{\partial\gamma} = 2g_e'H[(H'H + \gamma C'C)^{-2} - (H'H + \gamma C'C)^{-1}(H'H)(H'H + \gamma C'C)^{-2}]H'g_e.$$

This is a quadratic form in the transformed variables $H'g_e$, which we symbolize as

$$\frac{\partial\phi}{\partial\gamma} = 2g_e'HQH'g_e.$$

The matrix Q determines the nature of the quadratic form. We note that matrix Q is composed of sums, products, and inverses of block circulant matrices, and so we see that the matrix Q can be reduced to diagonal form by the transformation W discussed in Section V. Using the notation $\lambda(Q)$ to denote the eigenvalues of Q, then

$$\lambda(Q) = \frac{1}{(\mathcal{H}_e(m,n)\overline{\mathcal{H}_e(m,n)} + \gamma\mathcal{C}_e(m,n)\overline{\mathcal{C}_e(m,n)})^2} - \frac{\mathcal{H}_e(m,n)\overline{\mathcal{H}_e(m,n)}}{(\mathcal{H}_e(m,n)\overline{\mathcal{H}_e(m,n)} + \gamma\mathcal{C}_e(m,n)\overline{\mathcal{C}_e(m,n)})^3}$$
$$= \frac{\gamma\mathcal{C}_e(m,n)\overline{\mathcal{C}_e(m,n)}}{(\mathcal{H}_e(m,n)\overline{\mathcal{H}_e(m,n)} + \gamma\mathcal{C}_e(m,n)\overline{\mathcal{C}_e(m,n)})^3}$$

for $m = 0, 1, 2, \cdots, U - 1$ and $v = 0, 1, 2, \cdots, V - 1$. We see that, regardless of the signs of $\mathcal{H}_e(m, n)$ and $\mathcal{C}_e(m, n)$, the eigenvalues of Q are such that

$$\lambda(Q) \geqslant 0, \quad \text{if } \gamma \geqslant 0.$$

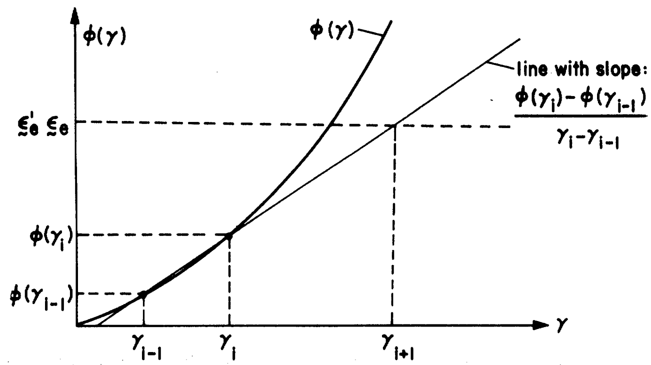


Fig. 1. Iterative scheme for determination of γ. (*Note:* Variable with wavy underline is denoted by boldface in text.)

Hence the corresponding quadratic form is positive semidefinite and the function $\phi(\gamma)$ is monotonically increasing as a function of γ.

The monotonicity of $\phi(\gamma)$ means that there is only one unique value of γ such that $\phi(\gamma) = \epsilon_e'\epsilon_e$ and an iterative process on γ is sufficient to find that unique value. Quick convergence to that unique value can be achieved by a Newton–Raphson-like procedure. Fig. 1 shows $\phi(\gamma)$ versus γ. To find the value $\phi(\gamma) = \epsilon_e'\epsilon_e$, the two previous values of $\phi(\gamma)$ are used to project the next value of γ. That the function $\phi(\gamma)$ is always positive for $\gamma > 0$ and that $\phi(0) = 0$ is seen directly from (9), (18), and (19).

The following steps summarize a practical algorithm for image restoration by constrained least squares estimation.

Step 1: Choose and fix an initial value of γ. Estimate the quantity $\epsilon_e'\epsilon_e$ from examination of the noise process by using (11).

Step 2: Form all the required extended matrices corresponding to choices for U and V. If radix-2 fast Fourier transform algorithms are used, then U and V must be exact powers of 2.

Step 3: Compute all the required Fourier transforms and compute an estimate of $\hat{\mathfrak{F}}_e$ using (17). Compute the inverse transform to get $\hat{f}_e$.

Step 4: For the fixed value of γ, compute the residual by (18) and the value $\phi(\gamma)$ by (19). Since the product $H\hat{f}_e$ in (18) is a convolution, do this convolution by the fast Fourier transform method.

Step 5: Increment or decrement γ.

a) $\phi(\gamma) < \epsilon_e'\epsilon_e - \alpha$. Increment γ according to an appropriate algorithm (such as the Newton–Raphson-like procedure).

b) $\phi(\gamma) > \epsilon_e'\epsilon_e + \alpha$. Decrement γ according to an appropriate algorithm.

Step 6: Return to Step 3 and continue, unless Step 7 is the case.

Step 7: $\phi(\gamma) = \epsilon_e'\epsilon \ \pm \alpha$, where α determines the accuracy with which the constraint is satisfied. Stop iteration, with $\hat{f}_e$ for the associated γ being the desired estimate.

Final comments concerning the choice of U and V in Step 2 are in order. Initially in Section III we specified $U \geqslant M + J - 1$ and $V \geqslant N + K - 1$, in order to demonstrate the convolution in vector-matrix form in (6). The zeros added to the matrices in the formation of extended matrices are necessary to suppress the effects of circular convolution using the fast Fourier transform [6]. We see in (9), however, double convolution, i.e., the product $H'H$. In the case of (9), the number of zeros that must be added must be increased to provide sufficient zeros to suppress the double convolution on the point-spread function. Thus for the algorithm above we require the choices

$$U \geqslant M + 2J - 1, \qquad U \text{ a power of 2}$$

$$V \geqslant N + 2K - 1, \qquad V \text{ a power of 2.}$$

In closing this section we point out that the lexicographic ordering of the image vectors, while used in proving the validity of (9) and (17), is also fundamental in the storage of images and their manipulation. Virtually all image digitizing systems of which the author is aware generate row or column one of the image matrix, followed by row or column two, row or column three etc. Hence, the data structure produced by an image digitizing device is basically a lexicographically ordered sequence from the image matrix. This data structure must be accounted for in implementing computer programs for the above algorithm [7].

VII. Two Examples

A computer program for the CDC 7600 computer was written to implement the algorithm in Section VI. The program was written to accommodate maximum choices for U and V of $U_{\max} = V_{\max} = 512$. Since the CDC 7600 computer is equipped with 512 000 words of large core memory, these choices for U and V meant that all transforms of image matrices could be done in large core memory, with a consequent reduction in magnetic disk I/O.

The computing time required to complete one iteration of the program, Steps 3 through 6 of the algorithm presented in the previous section, was approximately 20 s.

Using a Newton–Raphson-like procedure such as that given in Fig. 1, convergence to the proper value of γ is usually achieved in 4–7 iterations, depending on the initial choice for γ. The minimum number of iterations ever observed has been 3; the maximum number observed has been 12. The sensitivity parameter α, used in Steps 5 and 7 of the algorithm, has been set to reflect a 5 percent total error in the satisfaction of the constraint, i.e., $\alpha = 0.025(\epsilon_e'\epsilon_e)$. A C matrix based on radial symmetry, as in Approach 2 of Section IV, was used.

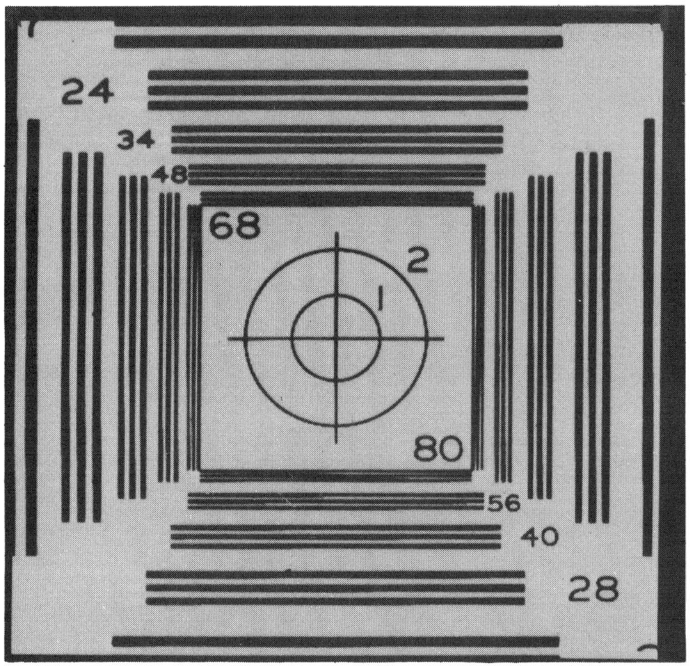


Fig. 2. Resolution chart image.

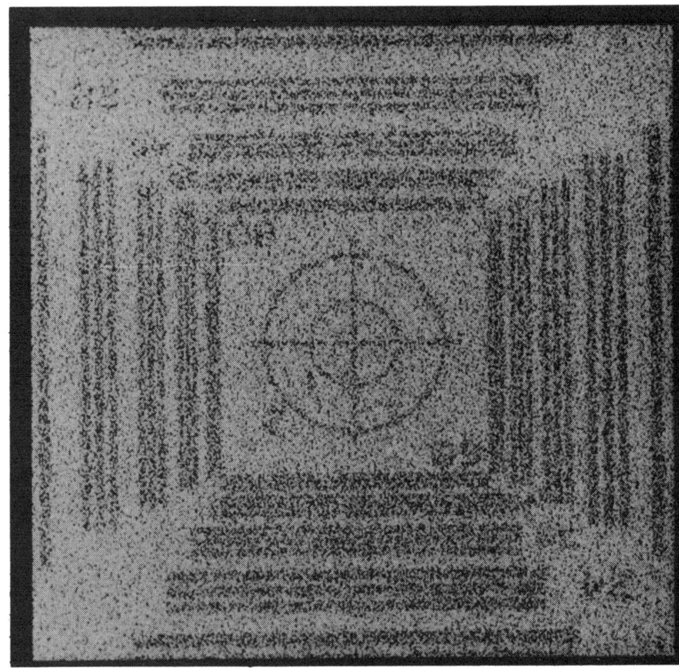


Fig. 4. Restoration with $\gamma = 0$.

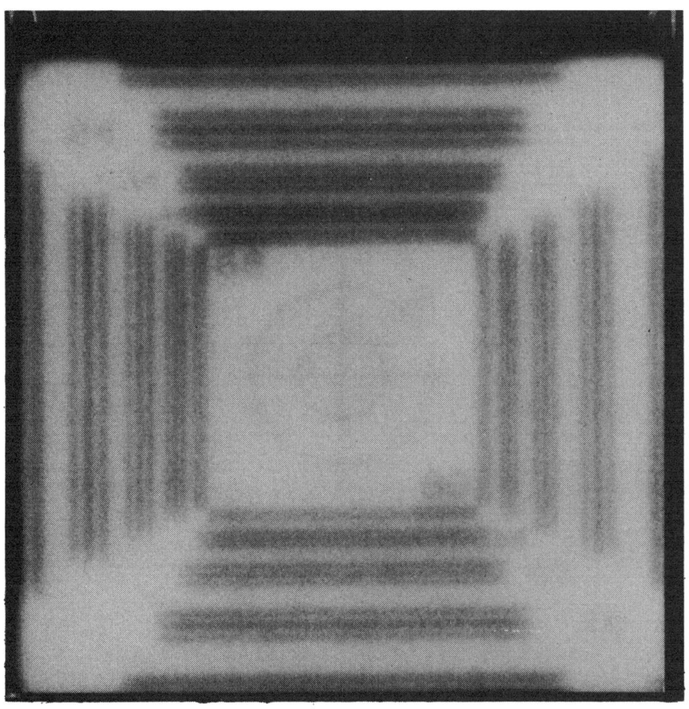


Fig. 3. Blurred resolution chart image with additive noise.

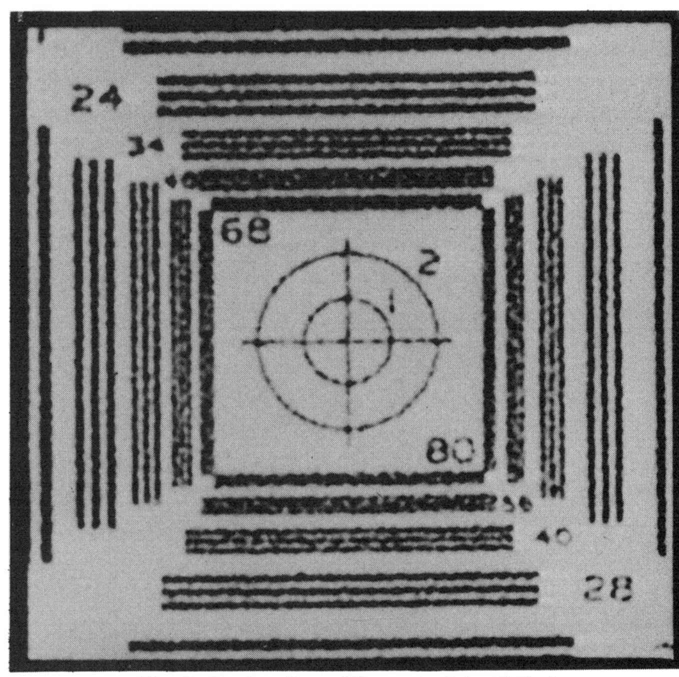


Fig. 5. Restoration with constraint satisfied.

Fig. 2 is an image of a standard National Bureau of Standards resolution chart. The image was scanned at 100-μ raster spacing using a 100-μ^2 aperture. The magnification of the image was arbitrary, so the numbers on the chart do not reflect line pairs per centimeter of resolution. The chart was digitized as a 450×450 matrix.

Fig. 3 is the chart of Fig. 2 after being blurred by a Gaussian-shaped point-spread function with a standard deviation of 2400 μ.

$$h(r) = \exp\left(-\left(\frac{r}{2400}\right)^2\right)$$

where r is a radial variable. Then random independent noise (white noise) samples were drawn from a uniform distribution on the interval [0, 0.5] and added to the blurred image. Fig. 3 is the result.

Fig. 4 is the image obtained by setting $\gamma = 0$ in (17). Since this is ill conditioned, the noise is greatly increased by this action. Allowing the program to increment γ to satisfy the constraint produced the image in Fig. 5. The variance and mean

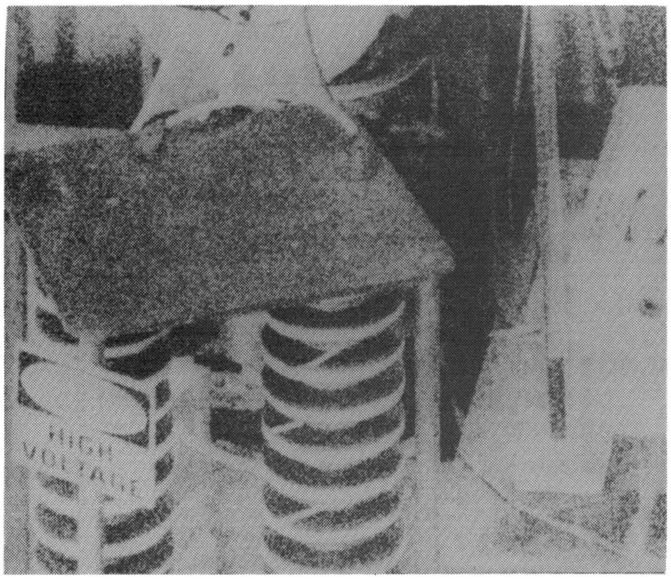


Fig. 6. Image with additive noise.

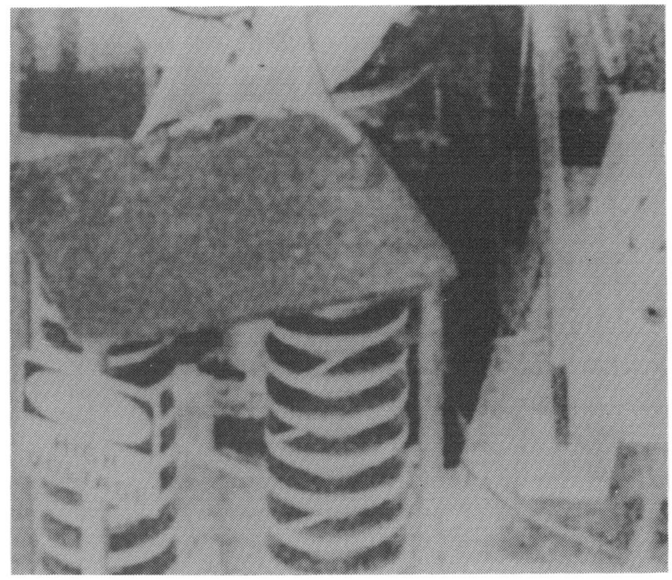


Fig. 7. Image with noise smoothed and constraint satisfied.

of the uniform density on the interval [0, 0.5] were used in (11) to estimate the quantity $\epsilon_e'\epsilon_e$. The improvement of Fig. 5 over Fig. 4 is readily discerned.

Fig. 6 is another image, a picture of some laboratory equipment at Los Alamos Scientific Laboratory, digitized as a 500×500 matrix. The picture was not blurred, but noise samples with a variance $\frac{1}{4}$ the variance of the image were drawn independently from a uniform density and added to the image (signal/noise ratio = 4, based on variances). Fig. 6 is the result.

The constrained least squares estimation technique is valid even if the point-spread function is a Dirac impulse. Further, the image itself may be used to estimate the noise variance. This was done in Fig. 6 by using an area where the picture was relatively constant. The flat table-like surface atop the choke coils was examined and the variance of the image densities computed in a region on this constant surface. The noise was assumed to be zero mean, and this computed variance was used in (11) to estimate $\epsilon_e'\epsilon_e$. The program was then allowed to iterate until γ satisfied the constraint on the residual. Fig. 7 is the result. This illustrates that the image on hand can be used to estimate the noise variance needed for the constraint. It also illustrates the use of the constrained least squares estimation technique for smoothing.

All images were digitized on a Photometric Data Systems image digitizer/image display unit at the Los Alamos Scientific Laboratory; the images were displayed on the same facility.

References

[1] M. M. Sondhi, "Image restoration: The removal of spatially invariant degradations," *Proc. IEEE*, vol. 60, pp. 842–853, July 1972.

[2] B. R. Hunt, "A theorem on the difficulty of numerical deconvolution," *IEEE Trans. Audio Electroacoust.* (Corresp.), vol. AU-20, pp. 94–95, Mar. 1972.

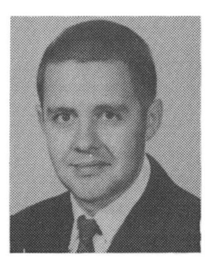
[3] H. C. Andrews, *Computer Techniques in Image Processing*. New York: Academic, 1970.

[4] D. L. Phillips, "A technique for numerical solution of certain integral equations of the first kind," *J. Ass. Comput. Mach.*, vol. 9, pp. 97–101, 1962.

[5] B. R. Hunt, "A matrix theory proof of the discrete convolution theorem," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 285–288, Dec. 1971.

[6] T. G. Stockham, Jr., "High speed convolution and correlation," in *1966 Spring Joint Comput. Conf., AFIPS Conf. Proc.*, vol. 28. Washington, D.C.: Spartan, 1966, pp. 229–233.

[7] B. R. Hunt, "Data structures and computational organization in digital image enhancement," *Proc. IEEE*, vol. 60, pp. 884–887, July 1972.



B. R. Hunt (S'66–M'67) was born in McAlester, Okla., on August 24, 1941. He received the B.Sc. degree in aeronautical engineering from Wichita State University, Wichita, Kans., in 1964, the M.Sc. degree in electrical engineering from Oklahoma State University, Stillwater, in 1965, and the Ph.D. degree in systems engineering from the University of Arizona, Tucson, in 1967.

From 1967 to 1968 he was employed at Sandia Laboratories, Albuquerque, N. Mex., performing systems analyses of strategic weapons. Since 1968 he has been associated with the University of California, Los Alamos Scientific Laboratory, Los Alamos, N. Mex., in the Division of Computer Science. He is currently conducting research in filtering and processing images by digital computer, and is also performing research in the associated areas of systems theory and numerical techniques. He has been a leader in the application of large-scale scientific computers to image processing.