

Bayesian Methods in Nonlinear Digital Image Restoration

B. R. HUNT

Abstract—Prior techniques in digital image restoration have assumed linear relations between the original blurred image intensity, the silver density recorded on film, and the film-grain noise. In this paper a model is used which explicitly includes nonlinear relations between intensity and film density, by use of the *D-log E* curve. Using Gaussian models for the image and noise statistics, a maximum *a posteriori* (Bayes) estimate of the restored image is derived. The MAP estimate is nonlinear, and computer implementation of the estimator equations is achieved by a fast algorithm based on direct maximization of the posterior density function. An example of the restoration method implemented on a digital image is shown.

Index Terms—Image restoration, nonlinear processing of images, Bayesian estimation, optimization theory, fast algorithms.

I. INTRODUCTION

DIGITAL PROCESSING of images has become in recent years a very active field of computer applications and research. One of the more active topics in this field has been digital restoration of images, as can be ascertained from recent survey papers [1], [2]. A number of different techniques have been proposed for digital image restoration, utilizing a number of different models and assumptions. About the only point of similarity in most of the existing methods has been their common use of models that do not account for the nonlinearities of image recording by film and the associated behavior of film-grain noise. In this paper we shall try to remove some of these previous restrictions and thereby extend the generality of digital image restoration to more realistic models of image formation and recording processes. The tool we shall employ in this task is that of Bayesian analysis: the characterization of the image and noise by statistical descriptions with associated probability densities, and the manipulation of conditional densities to develop an estimate of the restored image.

Other approaches to Bayesian analysis for digital image restoration have been formulated in recursive form [3], [4], [5]. The approach used herein will be to characterize image formation and recording processes in a form that does not lead to recursion. This makes easier the description of nonlinear effects, but also makes the resulting equations considerably more difficult to solve.

II. BASIC MODEL FORMULATION

The basic model which underlies the discussion within this paper is presented in Fig. 1. An object is perceived by

Manuscript received May 10, 1975; revised April 1, 1976. This work was performed under the auspices of the U.S. Energy Research and Development Administration.

The author is with the Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721.

virtue of an intensity distribution of some radiated energy (e.g., light-waves, penetrating radiation). The radiated energy is acted upon by an image formation system (e.g., optical lenses, or opaque pinhole collimators for penetrating radiation) and an image is formed. The image must be recorded by some means in the plane of image formation, which is the sensor in Fig. 1.

Fig. 1 is describable in block-diagram schematic form as shown in Fig. 2. The original distribution of object radiated intensity is acted upon by a linear system, with point-spread function h , to produce a blurred image b . The blurred image is then sensed and recorded. The recording process involves a nonlinear transformation, however, which we symbolize by the function s in Fig. 2 and which we discuss below. Finally, the image sensor is not ideal and introduces additive noise into the recorded image. Thus we assume the only noise source to be that due to image sensing and recording (and is usually the noise source that is most objectionable). The description of the creation of the blurred image is given in terms of the point-spread function by the equation

$$b(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y,x_1,y_1)f(x_1,y_1) dx_1 dy_1. \quad (1)$$

The recorded image is given by

$$g(x,y) = s(b(x,y)) + n(x,y). \quad (2)$$

The image restoration problem is to estimate f given the point-spread function h , the recorded image g , the response function s , and some amount of *a priori* knowledge (usually of a statistical nature) about the noise and/or f .

The form of the model assumed for (2) is justified by what is known about image sensing and recording processes. There are two basic categories of image sensing and recording systems: photochemical systems (chiefly film), and photoelectronic systems (TV and its various refinements). For film recording systems the relation between the variable g and the variable b is that between film optical density and the logarithm of incident intensity [6]. Consequently, the function s in (2) is the so-called film characteristic curve, or *D-log E* curve, and can be approximated by the logarithm function in regions where film response is linear with the logarithm of incident intensity [6]. For photoelectronic systems a power-law behavior governs the relationship between incident light intensity and the output of the photoelectronic sensor, the output usually being current flowing in an external circuit [7]. Thus for photoelectronic systems the function s is raising to the power γ the variable b . (The power γ is also used in film systems to describe the slope of the linear portion of the *D-log E* curve.)

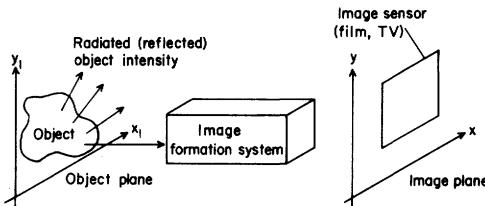


Fig. 1. Image formation and recording geometry.

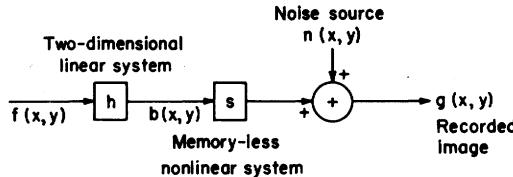


Fig. 2. Block diagram of image formation and recording process.

The additive nature of the noise in (2) is also justified on the basis of the nature of noise in image sensors. In photoelectronic systems the output current of the sensor is usually subjected to the usual sources of thermal noise, amplifier electronic noise, etc. The use of an additive model for such noise is a conventional practice [8]. For film recording systems, the source of noise is principally that of film-grain. Film-grain noise is a very complex signal-dependent form of noise, and can take on a number of forms of variation with signal [6], [9]. A simplifying assumption that does not do a grave injustice to reality is to assume the noise is additive in film optical density with a constant variance, which in turn corresponds to a multiplicative noise source in the incident intensities if exposure is in the linear response region of the D -log E curve [9]. The most recent results of attempting to encompass a signal-dependent noise model into algorithms for image restoration show that very little is gained by a signal-dependent model [9], further justifying this simplifying assumption. Thus the model of (2) is sufficient to account for known nonlinearities in both photoelectric and photochemical image recording systems. For photoelectric systems the response variable g will be current (typically) and the input variable b will be incident intensities; for photochemical systems the response variable g will be film optical density and b will be incident intensities.

Equations (1) and (2) describe the image formation and recording processes in terms of continuous variables, since the actual processes are basically continuous in space. The digital restoration of an image requires the sampling of the recorded image and discrete processing. The underlying linear systems of equations can be described in a vector-matrix formulation using lexicographical orderings of the sampled images [10]. Thus one can describe (1) in a discrete version:

$$\mathbf{b} = \mathbf{Hf} \quad (3)$$

and the corresponding analogy to (2) is

$$\begin{aligned} \mathbf{g} &= \mathbf{s}(\mathbf{b}) + \mathbf{n} \\ &= \mathbf{s}(\mathbf{Hf}) + \mathbf{n} \end{aligned} \quad (4)$$

where the notation for \mathbf{s} as a function of the vector \mathbf{b} means that each component of \mathbf{b} is transformed by the same function s to generate a new vector, i.e.,

$$\mathbf{c} = \mathbf{s}(\mathbf{b})$$

where

$$c_i = s(b_i)$$

for each component. Equation (4) is the basic structure we shall be assuming for the digital image restoration problem, which is to estimate the *digital restored image* \mathbf{f} from the data \mathbf{g} and knowledge of the point-spread function.

III. ESTIMATION METHODS AND MODELS

If we treat (4) as a deterministic problem, then it is a matter requiring solution of a nonlinear system of equations. Viewing it as a stochastic problem, however, leads one to estimate the digital restored image given the digital recorded image and some (statistical) knowledge of the noise. We adopt the latter view, which leads directly to the concept of conditional estimates: to estimate f conditioned on a knowledge of g . Bayes' law leads immediately to the description of the *a posteriori* conditional density:

$$p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{g})}. \quad (5)$$

The use of the posterior density for estimation is well known [11]. Minimum mean-square error (MMSE) estimates are the mean of the posterior density. Maximum *a posteriori* (MAP) estimates are the mode of the posterior density. Maximum likelihood (ML) estimates may be viewed as a special case of MAP where the posterior density is equal to the prior density $p(\mathbf{g}|\mathbf{f})$.

The form of specific estimates depends upon the form of the probability density functions chosen to model the

quantities \mathbf{f} and \mathbf{n} . In modeling these quantities we are guided by two principles.

1) Physical reality. The noise model for \mathbf{n} is determined by the physical processes that introduce noise into image sensing and recording systems. Fortunately, it is possible to model the underlying noise as a Gaussian process in both film and photoelectronic image recording systems [6], [7]. Thus the sampled vector \mathbf{n} will possess a multivariate Gaussian probability density, which is easy to work with. On the other hand, probabilistic descriptions of image intensities may be difficult to obtain. First, image probability densities are dependent upon the specific image at hand; second, they must be positive, which limits the choice of probability density. Faced with such uncertainty and constraints, we fall back upon a second guideline, namely,

2) Mathematical tractability. Simplifying assumptions for image probability densities are justified if the resulting simplification of the mathematics is profitable, and the assumption does not do damage to underlying reality. Again the Gaussian assumption is natural. The mathematics are greatly simplified, as we shall see below, the real-world images are probably sufficiently varied so as to make the Gaussian assumption applicable to a chosen (but unknown) subset of such images.

We formalize these considerations in the following assumptions for probability densities. First, we assume the noise process to be described by a multivariate normal probability density with mean zero and covariance \mathbf{R}_n :

$$p(\mathbf{n}) = ((2\pi)^{N/2} |\mathbf{R}_n|^{1/2})^{-1} \exp\left(-\frac{1}{2} \mathbf{n}^T \mathbf{R}_n^{-1} \mathbf{n}\right). \quad (6)$$

The image to be estimated as a restored quantity we describe by the multivariate normal density with mean $\bar{\mathbf{f}}$ and covariance \mathbf{R}_f :

$$\begin{aligned} p(\mathbf{f}) &= ((2\pi)^{N/2} |\mathbf{R}_f|^{1/2})^{-1} \\ &\cdot \exp\left(-\frac{1}{2} (\mathbf{f} - \bar{\mathbf{f}})^T \mathbf{R}_f^{-1} (\mathbf{f} - \bar{\mathbf{f}})\right). \end{aligned} \quad (7)$$

We assume the components of the mean vector $\bar{\mathbf{f}}$ are each positive, and are each sufficiently larger than the corresponding variance components (diagonal elements of \mathbf{R}_f) so that the positivity restrictions on light intensity are violated only very infrequently by this model. In (6) and (7) we have chosen N as the number of components in both vectors, implying H to be of size N by N ; the possibility of zero extension to satisfy these dimension requirements should be borne in mind [10].

The natural question to ask in conjunction with (7) is, how realistic is (7) in modeling a class of images \mathbf{f} ? In answering this question it is necessary to make the following points.

1) A stationary image model may have very little utility in modeling a situation. A second-order stationary model would have the properties that

$$E(\mathbf{f}) = \mu$$

$$E[(\mathbf{f} - \mu)(\mathbf{f} - \mu)^T] = \mathbf{R}_f$$

where μ is a constant vector, \mathbf{R}_f is a Toeplitz matrix,¹ and E is over the ensemble of random images described by (7). We see that in a stationary model, all the variability in the images must be conveyed by the covariance properties, since the ensemble mean has been fixed as a constant. It is known that computing the sample covariance statistics of an image (say a picture of scenery or a face), and then simulating random two-dimensional fields, about a fixed mean value, which possess the computed covariance statistics, does not generate images that have any resemblance to the original image [11]. Thus, in making probabilistic image models, covariance is a very weak property, particularly if the model is stationary with a constant mean value.

2) Assume, however, that we are given a model which consists of a nonuniform mean but with a stationary covariance, i.e.,

$$E(\mathbf{f}) = \bar{\mathbf{f}}$$

$$E[(\mathbf{f} - \bar{\mathbf{f}})(\mathbf{f} - \bar{\mathbf{f}})^T] = \mathbf{R}_f$$

where $\bar{\mathbf{f}}$ is not a constant vector, but \mathbf{R}_f is a Toeplitz matrix; this describes a process of stationary fluctuations about a nonuniform mean value. This can be a most realistic model of images. For example, consider an ensemble of images consisting of full-face photographs of individuals made for driver's licenses. Assuming all faces are registered by the photographer to be of approximately the same size and orientation in the image frame, then we can conclusively state that the ensemble mean should not be a uniform value. Instead, it should be approximately an oval, with suitable shadings to represent eyes, nose, mouth, etc. The mean of this ensemble is very structured, and the structure in the mean is directly related to the visible features in the image. Thus the covariance properties of the ensemble would represent the random perturbations about this structured mean of each individual image in the ensemble. As we shall see in the following discussions, the development of the Bayesian estimate is sufficiently powerful enough to allow a model, in the form of (7), in which the variations about the mean are stationary, even though the mean is not uniform and the underlying process is, therefore, not stationary.

Completing the evaluation of factors in the posterior density of (5) requires a description for $p(\mathbf{g})$. Unfortunately, the density of \mathbf{g} depends upon the nonlinear transformation s , and we cannot make a general description for \mathbf{g} . However, the MAP and ML estimates do not require $p(\mathbf{g})$. Consequently, we shall concentrate upon these estimates and not compute the MMSE estimate,

¹ In a Toeplitz matrix \mathbf{T} , the elements t_{ij} are such that $t_{ij} = t_{mn}$ if $i - j = m - n$.

which requires $p(\mathbf{g})$ in order to determine the conditional mean of (5). This is typical in nonlinear problems where the difficulty of evaluating the conditional mean is usually avoided for MAP results [12].

IV. DERIVATION OF ESTIMATOR EQUATIONS

As is conventional [12], we work with the logarithm of the posterior density:

$$\ln(p(\mathbf{f}|\mathbf{g})) = \ln(p(\mathbf{g}|\mathbf{f})) + \ln(p(\mathbf{f})) - \ln(p(\mathbf{g})) \quad (8)$$

and, since the last term on the right does not depend on \mathbf{f} , we neglect it in maximization with respect to \mathbf{f} . The MAP estimate is given by

$$\left[\frac{\partial \ln(p(\mathbf{g}|\mathbf{f}))}{\partial \mathbf{f}} + \frac{\partial \ln(p(\mathbf{f}))}{\partial \mathbf{f}} \right] \Big|_{\mathbf{f}=\hat{\mathbf{f}}_{\text{MAP}}} = 0 \quad (9)$$

where $\mathbf{f} = \hat{\mathbf{f}}_{\text{MAP}}$ is the solution at the maximum. Likewise, the maximum likelihood estimate is

$$\left[\frac{\partial \ln(p(\mathbf{g}|\mathbf{f}))}{\partial \mathbf{f}} \right] \Big|_{\mathbf{f}=\hat{\mathbf{f}}_{\text{ML}}} = 0, \quad (10)$$

and $\mathbf{f} = \hat{\mathbf{f}}_{\text{ML}}$ is the solution at the maximum. In (9) and (10) there remains only the specification of the *a priori* density $p(\mathbf{g}|\mathbf{f})$. Given \mathbf{f} , the variation in \mathbf{g} is that due to the noise \mathbf{n} ; therefore,

$$p(\mathbf{g}|\mathbf{f}) = ((2\pi)^{N/2} |\mathbf{R}_n|^{1/2})^{-1} \cdot \exp\left(-\frac{1}{2} (\mathbf{g} - s(\mathbf{H}\mathbf{f}))^T \mathbf{R}_n^{-1} (\mathbf{g} - s(\mathbf{H}\mathbf{f}))\right) \quad (11)$$

which follows directly from (6). Equations (9) and (10) are necessary conditions only, of course.

Substituting first from (11) into (10) then we can show (see the Appendix) that

$$\mathbf{H}^T \mathbf{S}_b \mathbf{R}_n^{-1} [\mathbf{g} - s(\mathbf{H}\mathbf{f})] \Big|_{\mathbf{f}=\hat{\mathbf{f}}_{\text{ML}}} = 0 \quad (12)$$

is the equation to be satisfied by the ML estimate, where \mathbf{S}_b is a diagonal matrix of derivatives with the form

$$\mathbf{S}_b = \begin{bmatrix} \frac{\partial s(u)}{\partial u} \Big|_{u=\hat{b}_1} & & & \\ & \frac{\partial s(u)}{\partial u} \Big|_{u=\hat{b}_2} & & \\ & & \ddots & \\ & & & \frac{\partial s(u)}{\partial u} \Big|_{u=\hat{b}_N} \end{bmatrix} \quad (13)$$

where

$$\mathbf{b} = \mathbf{H}\hat{\mathbf{f}}_{\text{ML}}. \quad (14)$$

Equation (12) is always satisfied if there exists a value of \mathbf{f} , which we call $\hat{\mathbf{f}}_{\text{ML}}$, such that

$$\mathbf{g} = s(\mathbf{H}\hat{\mathbf{f}}_{\text{ML}}), \quad (15)$$

and the corresponding value of $\hat{\mathbf{f}}_{\text{ML}}$ we call the maximum likelihood estimate of the digital restored image.

It is also possible to have (12) satisfied if every row of the matrix

$$\mathbf{H}^T \mathbf{S}_b \mathbf{R}_n^{-1} \Big|_{\mathbf{f}=\hat{\mathbf{f}}_{\text{ML}}}$$

is orthogonal to the vector

$$[\mathbf{g} - s(\mathbf{H}\mathbf{f})] \Big|_{\mathbf{f}=\hat{\mathbf{f}}_{\text{ML}}}.$$

Such an occurrence is possible, of course, if the matrix product $\mathbf{H}^T \mathbf{S}_b \mathbf{R}_n^{-1}$ is singular, and the solution to (12) is the entire null space of the matrix $\mathbf{H}^T \mathbf{S}_b \mathbf{R}_n^{-1}$. The choice of a single solution for the associated homogeneous equation can then be made by techniques such as the least squares form of the generalized-inverse solution. In either case, the solution obtained by (12) is usually of little value, as discussed in the next section.

Likewise, substituting from (7) and (11) into (9), we can show that the MAP estimate satisfies:

$$\hat{\mathbf{f}}_{\text{MAP}} = \bar{\mathbf{f}} + \mathbf{R}_f \mathbf{H}^T \mathbf{S}_b \mathbf{R}_n^{-1} [\mathbf{g} - s(\mathbf{H}\hat{\mathbf{f}}_{\text{MAP}})] \quad (16)$$

where \mathbf{S}_b is as defined in (13), but the derivatives are evaluated as

$$\mathbf{b} = \mathbf{H}\hat{\mathbf{f}}_{\text{MAP}}. \quad (17)$$

The nonlinear nature of (16) does not allow it to be simplified in the fashion of (15). Equations (15) and (16) are the basic estimator equations whose properties we consider in the next section. Details of the derivation of (12)–(17) are contained in the Appendix.

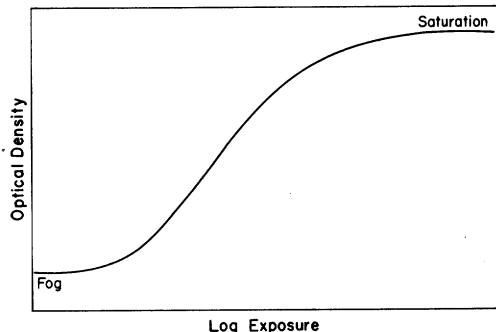
V. PROPERTIES OF ESTIMATORS

Consider first the special case where the function s is linear. Since s is memoryless it can, without loss of generality, be represented as a scalar multiplier α . If \mathbf{H} is nonsingular, then from (15):

$$\hat{\mathbf{f}}_{\text{ML}} = \frac{1}{\alpha} \mathbf{H}^{-1} \mathbf{g} \quad (18)$$

which is a solution by the *inverse filter*, in image restoration terminology, and can be implemented by fast Fourier transform [10]. The relation between the inverse filter and least squares solutions has been noted before [13], and the Gaussian assumption relating ML and least squares is well known; (18) is not surprising, therefore. This is not to say that the solution of (18) is desirable, however. Given the generally ill-conditioned nature of \mathbf{H} [10], [13], the ML inverse filter solution can possess excessively large variance, even though it is unbiased; the resulting restored images consist mostly of noise [10].

In the general nonlinear case, the function s may be such that the mapping from the domain to the range of s is not one-to-one. There is no unique s^{-1} , as a result, and construction of $\hat{\mathbf{f}}_{\text{ML}}$ is not unique. In image recording by film, the function s is usually of the general form in Fig. 3, the *D-log E* or characteristic curve. For exposures lying in the fog or saturation region, the mapping is not one-to-one. Similar types of curves are often found for electronic image recording systems [7]. However, if exposures sufficiently

Fig. 3. D - $\log E$ curve for typical film.

removed from fog or saturation are used, then the resulting function s is monotonic increasing and one-to-one, so that a unique s^{-1} exists. Under such assumptions

$$\hat{\mathbf{f}}_{\text{ML}} = \mathbf{H}^{-1}s^{-1}(\mathbf{g}). \quad (19)$$

The estimation of a restored image by a process such as (19) has been described before [14], but without recognition of its relation to ML estimation. Equation (19) can suffer from ill-conditioned behavior in \mathbf{H} , just as in (18), and the resulting restorations will be extremely noisy.

Of course, the point-spread function may be such that \mathbf{H} is singular, in which case the ML estimate cannot be computed by (18) or (19).

The MAP estimate possesses a more interesting set of properties than does the ML estimate. We note the following.

1) Assume the function s is linear, e.g., multiplication by $\alpha = 1$ without loss of generality. Then the matrix \mathbf{S}_b becomes the identity matrix. Equation (16) becomes

$$\hat{\mathbf{f}}_{\text{MAP}} = \bar{\mathbf{f}} + \mathbf{R}_f \mathbf{H}^T \mathbf{R}_n^{-1} [\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}_{\text{MAP}}]$$

which can be rearranged so that:

$$\begin{aligned} \hat{\mathbf{f}}_{\text{MAP}} &= (\mathbf{R}_n \mathbf{H}^{T-1} \mathbf{R}_f^{-1} + \mathbf{H})^{-1} \mathbf{g} \\ &+ (\mathbf{R}_n \mathbf{H}^{T-1} \mathbf{R}_f^{-1} + \mathbf{H})^{-1} (\mathbf{R}_n \mathbf{H}^{T-1} \mathbf{R}_f^{-1}) \bar{\mathbf{f}} \\ &= (\mathbf{R}_f^{-1} + \mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{g} \\ &+ (\mathbf{R}_f^{-1} + \mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H})^{-1} \mathbf{R}_f^{-1} \bar{\mathbf{f}}. \end{aligned} \quad (20)$$

We recognize this as a discrete Wiener filter with *a priori* mean included. If the matrices \mathbf{R}_n , \mathbf{R}_f , and \mathbf{H} are assumed to be Toeplitz forms (equivalent to stationarity of the random processes and spatial invariance of the point-spread function) then circulant approximations may be used and the corresponding estimation carried out by discrete Fourier transforms; the frequency domain form of (20) can be constructed directly and fast Fourier transforms used for computational efficiency [10], [13]. The equivalence of MAP and MMSE estimates for linear problems with symmetric densities is illustrated by (20) [12].

2) Equation (16), which the MAP estimate satisfies, is a nonlinear matrix equation. It is, however, a direct analogy to the nonlinear integral equation which describes the MAP estimate of a continuous, time-dependent random process [12, p. 435]; a discrete approximation to the corresponding integral equation would yield an equation of

the form of (16). However, by adopting the discrete approach to the restoration, we have avoided the problems associated with extension of the continuous one-dimensional problem to two dimensions. Indeed, a continuous two-dimensional formulation in an integral equation could be derived by letting the sample spacing approach zero so that the sums in (16) become integrations in the limit.

3) Just as in the continuous case, (16) has a structure that can be described in a feedback system, as we see in Fig. 4. Such systems for time functions are not physically realizable, because of the noncausal nature of the covariance function filters [12]. Causal restrictions do not exist in image processing, however, and the underlying feedback structure should be implementable in a straightforward fashion. Implementation and computational details are to be discussed in a subsequent paper concerning special hardware [17]. Note that digitally programmable light modulators, for the nonlinear function generation, and optical filtering and feedback could theoretically be employed to generate the MAP restored image estimate in a hybrid digital/optical system.

4) In (16) the *a priori* mean value is included directly in the estimation process. Suppose that one possessed no *a priori* knowledge about the restored image; then $\bar{\mathbf{f}}$ in (16) could be set to a positive constant (recall the discussion relevant to (7) above). However, suppose some *a priori* knowledge about the restored image is known, e.g., the original nonblurred image was a man's face; then the *a priori* knowledge could be used as the image $\bar{\mathbf{f}}$, and the extent of knowledge about the original image would determine how detailed the *a priori* image became. For example, a crude model of a face could be constructed if knowledge was limited to the existence of the face as the object; if the *a priori* knowledge extended to specific features, e.g., bushy eyebrows, mustache, then such specific features could be imposed on the model. Such can be done in the linear Wiener estimate of (20) also, but the author knows of no place where such *a priori* knowledge has been used in image restoration. This is unfortunate because there is frequently much specificity in *a priori* knowledge in image analysis: the medical diagnostician is not looking for airplane shapes in his images, the military photo-interpreter usually is.

5) Finally, we note that a problem which arises so frequently in many image restoration methods does not ap-

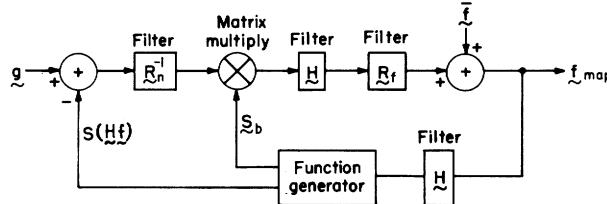


Fig. 4. Feedback structure of the MAP estimate.

pear in (16): the ill-conditioned properties of the matrix \mathbf{H} . Many image restoration methods require the computation of inverses of the matrix \mathbf{H} or inverses of matrices involving \mathbf{H} [cf., (20)]. In (16) the only explicit matrix inverse required is that of \mathbf{R}_n^{-1} . Since \mathbf{R}_n is the noise covariance matrix, the generally assumed "white" properties for the noise will make \mathbf{R}_n diagonal and the inversion is trivial. Even if the noise process is not white it is usually very wide band, making the matrix \mathbf{R}_n nearly a diagonal matrix, again with good inversion properties. Thus the MAP estimate possesses no difficulties in dealing with ill-conditioned or even singular matrices, whether for the point-spread function or for other matrices in the estimation process.

VI. COMPUTATIONAL ASPECTS

The solution of a general space-variant, nonlinear image restoration problem is posed in the ML and MAP equations, (15) and (16). However, for images of interest, the sampling produces a great number of points; the direct solution of (15) or (16) becomes impossible with conventional computers and current algorithms. Consequently, it is space-invariant restoration which we shall consider here, in which case the point-spread function takes the form

$$h(x, y, x_1, y_1) = h(x - x_1, y - y_1). \quad (21)$$

The corresponding matrix \mathbf{H} in (15) and (16) then has a block Toeplitz structure, and matrix-vector products with \mathbf{H} can be implemented by use of the block circulant approximation for a Toeplitz form [10]; the fast Fourier transform (FFT) algorithm is then used to provide rapid computation. Likewise, if the matrices \mathbf{R}_f and \mathbf{R}_n are Toeplitz, then matrix products indicated in (16) can also be carried out by circulant approximations and the FFT algorithm. The matrices \mathbf{R}_f and \mathbf{R}_n are Toeplitz, provided the underlying random processes are stationary about the ensemble mean, which is consistent with the discussion and assumptions of Section III. At the close of this section we shall list in detail the steps to go through in carrying out the indicated matrix computations by FFT.

Equation (16) represents the greatest computational difficulty, since it is a nonlinear system of equations possessing a feedback structure, as discussed above. Two general techniques appear feasible for computing the MAP estimate, and both techniques require iteration: implicit

iterative schemes (e.g., Picard's iteration), and explicit maximization of the posterior density. The former scheme utilizes (16) directly, whereas the latter scheme makes use of the relation between (16) and the associated maximization of the posterior density.

Picard's iteration relies upon the direct formation of (16), which can be written as

$$\hat{\mathbf{f}}_{\text{MAP}} = \phi(\hat{\mathbf{f}}_{\text{MAP}}) \quad (22)$$

where the vector function ϕ is the right-hand side of (16). We then form the sequence $\hat{\mathbf{f}}_n$ by taking an initial estimate of $\hat{\mathbf{f}}_{\text{MAP}}$, which we call $\hat{\mathbf{f}}_0$, and then computing

$$\left. \begin{aligned} \hat{\mathbf{f}}_1 &= \phi(\hat{\mathbf{f}}_0) \\ \hat{\mathbf{f}}_2 &= \phi(\hat{\mathbf{f}}_1) \\ &\vdots \\ \hat{\mathbf{f}}_n &= \phi(\hat{\mathbf{f}}_{n-1}) \end{aligned} \right\} \quad (23)$$

Provided a set of convergence conditions are met, then it is possible to show that the process converges in the limit, i.e.,

$$\lim_{n \rightarrow \infty} \hat{\mathbf{f}}_n = \hat{\mathbf{f}}_{\text{MAP}}. \quad (24)$$

The conditions under which the iteration sequence of (23) will converge depend upon the function ϕ and the initial guess $\hat{\mathbf{f}}_0$. A necessary condition for convergence is the following: If $\hat{\mathbf{f}}_{\text{MAP}}$ is the solution to (16), and if in a region R about $\hat{\mathbf{f}}_{\text{MAP}}$ it is true that

$$\sum_{j=1}^N \left| \frac{\partial \phi_i}{\partial f_j} \right| < 1, \quad \text{for } i = 1, 2, \dots, N \quad (25)$$

and if the initial guess $\hat{\mathbf{f}}_0$ lies in R , then the sequence of (23) will converge to $\hat{\mathbf{f}}_{\text{MAP}}$.

The test of (25) is not of much practical utility. First, it is necessary to have some estimate of the solution so that one knows approximately over what region to carry out the test in (25). Second, the actual elements of the Jacobian matrix are very complex, as indicated by repeated application of the differentiation rules in the Appendix; the resulting expressions for each element of the Jacobian are so complex that no simple procedure for carrying out the test in an analytical (closed-form) manner has been found. This is, of course, a common occurrence in working with nonlinear equations; often convergence cannot be proved but can be demonstrated by numerical experiment [15]. Numerical simulations of the sequence in (23) have dem-

onstrated that $\hat{\mathbf{f}}_{\text{MAP}}$ is a fixed point of ϕ , i.e., if one constructs a blurred image from a known \mathbf{f} , then using the known \mathbf{f} as $\hat{\mathbf{f}}_0$ results in $\hat{\mathbf{f}}_n = \hat{\mathbf{f}}_{n-1}$ for all n . Slight deviations from the known \mathbf{f} result in iteration divergence, indicating the region R is very small for the particular simulation experiment. Numerical experiments of this kind have shown that the Picard's iteration scheme is too problem-dependent to be of general use in solving (16). We turn to another approach instead.

Equation (16) maximizes the posterior density function, and was derived by taking the logarithm, computing the gradient, and solving after equating the gradient to zero. The logarithm of the posterior density function is the sum of two quadratic forms:

$$\begin{aligned} \ln p(\mathbf{f}|\mathbf{g}) = K - \frac{1}{2} [\mathbf{g} - s(\mathbf{H}\mathbf{f})]^T \mathbf{R}_n^{-1} [\mathbf{g} - s(\mathbf{H}\mathbf{f})] \\ - \frac{1}{2} [\mathbf{f} - \bar{\mathbf{f}}]^T \mathbf{R}_f^{-1} [\mathbf{f} - \bar{\mathbf{f}}] \end{aligned} \quad (26)$$

where K includes all terms that are not a function of \mathbf{f} . The direct method is to find the value of \mathbf{f} that minimizes the right-hand side of (26), since this maximizes the posterior density. A number of iterative schemes to minimize functions such as (26) are known; we shall discuss those that fall under the general heading of iterative gradient methods.

Let a scalar function of a vector \mathbf{x} be given, $\psi(\mathbf{x})$. Then the minimum of ψ can be computed by the iterative sequence:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla \psi(\mathbf{x}_k) \quad (27)$$

where α_k is determined by various schemes to speed convergence [15], [16]. The convergence of (27) to the minimum is guaranteed under quite general conditions if a unique minimum exists [15]. To apply (27) to the quadratic forms in the posterior density function, we define

$$\begin{aligned} \psi(\mathbf{f}) = K - \frac{1}{2} [\mathbf{g} - s(\mathbf{H}\mathbf{f})]^T \mathbf{R}_n^{-1} [\mathbf{g} - s(\mathbf{H}\mathbf{f})] \\ - \frac{1}{2} [\mathbf{f} - \bar{\mathbf{f}}]^T \mathbf{R}_f^{-1} [\mathbf{f} - \bar{\mathbf{f}}] \end{aligned} \quad (28)$$

and so we have that

$$\nabla \psi(\mathbf{f}) = \mathbf{H}^T \mathbf{S}_b \mathbf{R}_n^{-1} [\mathbf{g} - s(\mathbf{H}\mathbf{f})] - \mathbf{R}_f^{-1} [\mathbf{f} - \bar{\mathbf{f}}].$$

Thus the iterative scheme to compute the MAP estimate by direct minimization of $\psi(\mathbf{f})$ is

$$\begin{aligned} \hat{\mathbf{f}}_{k+1} = \hat{\mathbf{f}}_k - \alpha_k (\mathbf{H}^T \mathbf{S}_b \mathbf{R}_n^{-1} [\mathbf{g} - s(\mathbf{H}\hat{\mathbf{f}}_k)]) \\ - \mathbf{R}_f^{-1} [\hat{\mathbf{f}}_k - \bar{\mathbf{f}}]. \end{aligned} \quad (29)$$

To utilize (29) we choose an initial guess, $\hat{\mathbf{f}}_0$, and then perform the indicated computations. The decision as to when convergence is achieved can be judged by the criterion of the magnitude of the gradient term. At the minimum (or any local minimum) of $\psi(\mathbf{f})$, the gradient is zero.

Thus as $\hat{\mathbf{f}}_k$ approaches the solution, $\nabla \psi(\mathbf{f})$ becomes smaller. The solution can be said to be achieved if

$$\|\nabla \psi(\hat{\mathbf{f}}_k)\| < \epsilon \quad (30)$$

where ϵ is chosen as a tolerance on the error in the final gradient. When (30) is achieved then we have that

$$\hat{\mathbf{f}}_{\text{MAP}} = \hat{\mathbf{f}}_K \quad (31)$$

and the solution is achieved.

The only difficulty in (29) lies in the gradient computations themselves. However, \mathbf{H} is Toeplitz and approximation by FFT is possible. Likewise, the matrices \mathbf{R}_f^{-1} and \mathbf{R}_n^{-1} are approximately Toeplitz, and the approximation is quite good for large matrices such as encountered in image processing [18]. Thus we can employ FFT calculations with these matrices and construct the following algorithm.

- 1) Set $k = 0$ and construct an initial guess \mathbf{f}_0 (e.g., make a Wiener filter restoration based upon a linear approximation to the nonlinear characteristic s).
- 2) Compute $\mathbf{H}\hat{\mathbf{f}}_k$ by computing the discrete Fourier transform (DFT) of the point-spread function, the DFT of $\hat{\mathbf{f}}_k$, multiplying the two, and computing the inverse DFT of the product. Perform DFT's by fast Fourier transform.
- 3) Using the sensor response function s , compute $s(\mathbf{H}\hat{\mathbf{f}}_k)$ and subtract from the recorded image to generate $\mathbf{x}_k = (\mathbf{g} - s(\mathbf{H}\hat{\mathbf{f}}_k))$.
- 4) Compute $\mathbf{R}_n^{-1} \mathbf{x}_k$ using DFT's (see step 2) of \mathbf{x}_k and the inverse autocovariance sequence (note that \mathbf{R}_f^{-1} can be computed from \mathbf{R}_f using DFT's [10]). Let $\mathbf{y}_k = \mathbf{R}_n^{-1} \mathbf{x}_k$.
- 5) Multiply each component of \mathbf{y}_k by the appropriate derivative in \mathbf{S}_b , i.e.,

$$y_{kj}' = \left(\frac{\partial s}{\partial u} \Big|_{u=\sum_i h_{ji} \hat{f}_{ki}} \right) y_{kj}.$$

- 6) Compute \mathbf{H}^T times \mathbf{y}_k' using DFT's, analogous to step 2) above.
- 7) Compute the difference $(\hat{\mathbf{f}}_k - \bar{\mathbf{f}})$ and then compute $\mathbf{R}_f^{-1}(\hat{\mathbf{f}}_k - \bar{\mathbf{f}})$ using DFT's as in step 2) above.
- 8) Subtract the result of step 7) from the result of step 6). This quantity is $\nabla \psi(\hat{\mathbf{f}}_k)$.
- 9) If (30) is satisfied, then stop. Otherwise multiply $\nabla \psi(\hat{\mathbf{f}}_k)$ by α_k and subtract from $\hat{\mathbf{f}}_k$ to generate the next estimate $\hat{\mathbf{f}}_{k+1}$.
- 10) Set $k = k + 1$ and return to step 2).

We make the following comments in concluding the discussion of this section. First, the optimum choice of α_k can be determined theoretically and used to take a maximum-length step in the minimum gradient direction for each step k . For complex nonlinear functions, however, the solution for optimum α_k cannot be made in a closed form, and it is more reasonable to use a fixed value of α_k for all iterations. That such a suitable fixed value exists can also

be proved [15], but again the numerical determination may be difficult. Practical experience has been to choose a fixed α and then test it in an algorithm such as the above. If the algorithm diverges after a few steps, then a smaller α should be tried until the process exhibits convergence. The ideal fixed value of α is that which is the largest without causing the algorithm to diverge (when the algorithm is converging, then $\|\hat{\mathbf{f}}_{k+1} - \hat{\mathbf{f}}_k\|$ decreases with each step).

A second comment concerns the choice of the steepest descent process for maximization of the posterior density. Other techniques are possible (conjugate gradient, Davidon-Fletcher-Powell, etc.); however, the more powerful techniques require more complex computations at each iteration step. As the number of variables in the problem increases, the simpler methods become competitive (because of fewer computations per step) than the more complex methods. The steepest descent process seems to offer a reasonable compromise between the number of iterations required and the computational complexity required at each iteration, particularly for the large number of variables in a digital image restoration [15].

VII. AN EXAMPLE

In this final section we demonstrate the use of a steepest descent algorithm to generate the MAP estimate of the restored image by fast transform computations. In Fig. 5 is a two-dimensional digital image, sampled as a 128 by 128 matrix. The display of Fig. 5 is proportional to intensity, i.e., the display represents the image that would be observed by a photometer or other instrument which senses intensity directly. In Fig. 6 we see the same image after being blurred by a motion blur point-spread function of 5 points in width; the blur is in the intensity domain and the image of Fig. 6 is, like Fig. 5, an intensity display. Finally, we see in Fig. 7 the result of transforming the blurred image through a D -log E curve, followed by the addition of zero-mean uncorrelated Gaussian noise, to simulate film-grain noise. The D -log E curve used in the generation of the image in Fig. 7 is seen in Fig. 8; the original image in Fig. 5 spanned a range of (dimensionless) E values from 11.0 to 200.0. Thus Fig. 7 spans a range from approximately 0.4 to 1.7. These are optical density values and thus Fig. 7 is a density display, i.e., the values displayed are proportional to film optical density. The signal-to-noise ratio of the image in Fig. 7 is 23 dB, measured in terms of image and noise variance; that is, the variance of the image divided by the variance of the noise gives an SNR of 23 dB.

Fig. 9 shows the result of restoring the image in Fig. 7 by means of the MAP estimation process discussed in this paper; the MAP estimate was computed by means of the fast algorithm, discussed in Section VI, using fast transform procedures to compute the gradient values for the steepest-descent process. Since the MAP estimate directly reconstructs the original intensity, Fig. 9 is again an in-

tensity display. The restoration of features and details in Fig. 9 is apparent, as comparison to Fig. 7 and Fig. 5 indicates. The restoration process requires the ensemble mean $\bar{\mathbf{f}}$ to generate the restoration, as discussed above. To generate the estimate of Fig. 9, the assumption was made that the blurred recorded image, Fig. 7, was representative of the ensemble mean. Thus Fig. 7 was transformed back into the intensity domain (by backward interpolation through the D -log E curve) and used as the $\bar{\mathbf{f}}$ for the restoration process. This produced stable computational behavior and algorithm convergence; if other knowledge was available on $\bar{\mathbf{f}}$, then it could be used in place of the blurred recorded image, as was done above. This value of $\bar{\mathbf{f}}$ was also used for the first estimate in the iterative process.

Of interest in any iterative scheme is the speed of convergence. Fig. 10 shows the mean-square error as a function of the number of iterations for different values of α . We see that the error drops rapidly initially and then becomes quite slow. In this plot $\alpha = 5$ is near optimum, and was used in the example above. Larger values of α result in algorithm divergence, as discussed in the previous section. The optimum value of α is sensitive to the D -log E curve primarily, since the term containing S_b usually dominates the other term in (29), i.e., if the sensor response has steeper slope a smaller value of α may be required to produce convergence. For values of α near optimum, the flattening of the error curve, as in Fig. 10, means that approximately 40 to 60 iterations suffice to produce a restoration. Fig. 9 was achieved after 60 iterations.

The iterative algorithm is quite stable in convergence behavior, provided α is not too large, and convergence in approximately 60 iterations has been achieved for different types of image blurs (motion blur and focus blur), and for different signal-to-noise ratios (from 15 to 30 dB). The algorithm has always shown convergence when $\bar{\mathbf{f}}$ and the initial estimate are both derived from the recorded image, as discussed above. However, for the most noisy images, it has been found beneficial to slightly smooth the image after backward interpolation through the D -log E curve, in order to suppress the effects of noise in both $\bar{\mathbf{f}}$ and the initial estimate; this generates a restoration of more pleasing visual appearance (since noise in $\bar{\mathbf{f}}$ propagates into the restoration, as can be seen in (29)).

The restored image of Fig. 9 shows more visual noise than the original, and the SNR in Fig. 9 is about 16 dB. This decrease in SNR is typical of image restoration schemes and is primarily a function of the ill-conditioned nature of the restoration process causing an amplification of noise in the image.

The images in this example were scanned, displayed, and processed on equipment at the Los Alamos Scientific Laboratory (LASL). The image computations were implemented in the LADIES image processing software at LASL on a CDC 7600 computer. The generation of the MAP estimate required approximately 16 min of CDC 7600 computing time.



Fig. 5. Original image (intensity representation).



Fig. 6. Image of Fig. 5 after motion blur (intensity representation).

Fig. 7. Image of Fig. 6 after transformation by D -log E curve (density representation).

APPENDIX

DETAILED DERIVATION OF ESTIMATOR EQUATIONS

The derivation of the estimator equations is not difficult but is tedious in notation, because differentiation of a quadratic form with a nonlinear function is involved. The quadratic forms result from the logarithm of the posterior density, (28) in the main text of the paper:

$$\begin{aligned} \psi(\mathbf{f}) = K - \frac{1}{2} [\mathbf{g} - s(\mathbf{H}\mathbf{f})]^T \mathbf{R}_n^{-1} [\mathbf{g} - s(\mathbf{H}\mathbf{f})] \\ - \frac{1}{2} [\mathbf{f} - \bar{\mathbf{f}}]^T \mathbf{R}_f^{-1} [\mathbf{f} - \bar{\mathbf{f}}]. \quad (28) \end{aligned}$$

Expanding the indicated products, we must differentiate

$$\begin{aligned} \psi(\mathbf{f}) = K - \frac{1}{2} [\mathbf{g}^T \mathbf{R}_n^{-1} \mathbf{g} \\ - 2s(\mathbf{H}\mathbf{f})^T \mathbf{R}_n^{-1} \mathbf{g} + s(\mathbf{H}\mathbf{f})^T \mathbf{R}_n^{-1} s(\mathbf{H}\mathbf{f})] \\ - \frac{1}{2} [\mathbf{f}^T \mathbf{R}_f^{-1} \mathbf{f} - 2\mathbf{f}^T \mathbf{R}_f^{-1} \bar{\mathbf{f}} + \bar{\mathbf{f}}^T \mathbf{R}_f^{-1} \bar{\mathbf{f}}]. \quad (A1) \end{aligned}$$

We wish to concentrate upon terms in the first square bracket that are functions of \mathbf{f} . Therefore, we must differentiate the terms

$$2s(\mathbf{H}\mathbf{f})^T \mathbf{R}_n^{-1} \mathbf{g} = \sum_{m=1}^N \sum_{p=1}^N s \left(\sum_{t=1}^N h_{mt} f_t \right) r_{mp} g_p \quad (A2)$$

and

$$\begin{aligned} s(\mathbf{H}\mathbf{f})^T \mathbf{R}_n^{-1} s(\mathbf{H}\mathbf{f}) \\ = \sum_{m=1}^N \sum_{p=1}^N s \left(\sum_{t=1}^N h_{mt} f_t \right) r_{mp} s \left(\sum_{u=1}^N h_{pu} f_u \right) \quad (A3) \end{aligned}$$

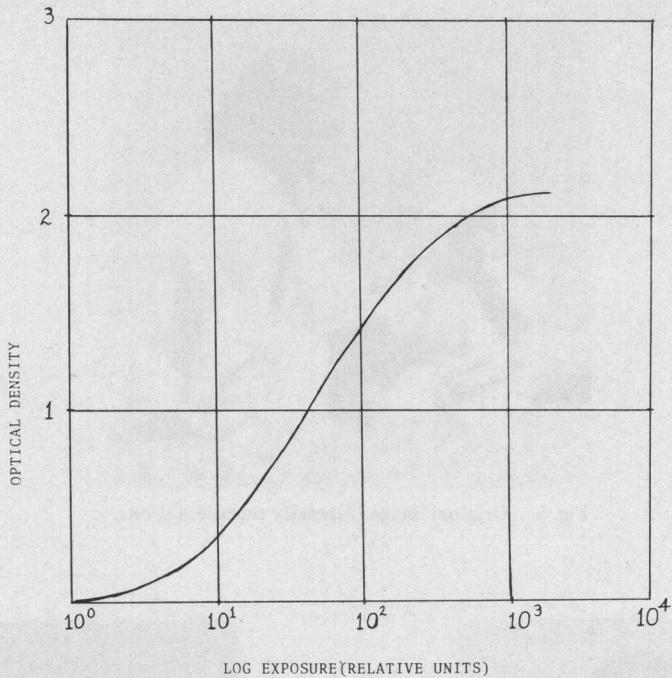
Fig. 8. D -log E curve used in example.

Fig. 9. Image restored by MAP process.

where r_{mp} is the m th element of the matrix \mathbf{R}_n^{-1} . Taking derivatives with respect to f_c for $c = 1, 2, \dots, N$, we have

$$\frac{\partial(A - 2)}{\partial f_c} = \sum_{m=1}^N \sum_{p=1}^N \frac{\partial s(b_m)}{\partial b_m} \frac{\partial b_m}{\partial f_c} r_{mp} g_p \quad (\text{A4})$$

where

$$b_m = \sum_{t=1}^N h_{mt} f_t \quad (\text{A5})$$

$$\begin{aligned} \frac{\partial(A - 3)}{\partial f_c} = & \sum_{m=1}^N \sum_{p=1}^N \left[\frac{\partial s(b_m)}{\partial b_m} \frac{\partial b_m}{\partial f_c} r_{mp} s(b_p) \right. \\ & \left. + s(b_m) r_{mp} \frac{\partial s(b_p)}{\partial b_p} \frac{\partial b_p}{\partial f_c} \right] \quad (\text{A6}) \end{aligned}$$

where

$$b_p = \sum_{u=1}^N h_{pu} f_u. \quad (\text{A7})$$

The indicated differentiations are an obvious application of the chain-rule. In (A4) we have

$$\frac{\partial s(b_m)}{\partial b_m} \frac{\partial b_m}{\partial f_c} = \frac{\partial s(b_m)}{\partial b_m} h_{mc} \quad (\text{A8})$$

and in (A6) we likewise have

$$\frac{\partial s(b_p)}{\partial b_p} \frac{\partial b_p}{\partial f_c} = \frac{\partial s(b_p)}{\partial b_p} h_{pc}. \quad (\text{A9})$$

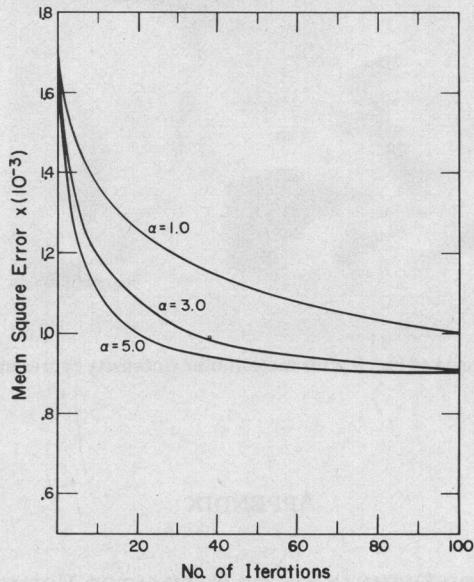


Fig. 10. Error behavior curves.

For all $c = 1, 2, \dots, N$, there are equations like (A4) and (A6), and it is possible to write (A4) and (A6) using vector matrix products. Substituting into (A4) from (A8), we recognize that

$$\frac{\partial(A-2)}{\partial f} = H^T S_b R_n^{-1} g \quad (\text{A10})$$

where S_b is a diagonal matrix of derivatives

$$S_b = \begin{bmatrix} \frac{\partial s}{\partial u} \Big|_{u=b_1} & & \\ & \ddots & \\ & & \frac{\partial s}{\partial u} \Big|_{u=b_N} \end{bmatrix} \quad (\text{A11})$$

(A11)

and the b_i are as defined in (A8) or (A9).

Equation (A6) is more complex, since there are two terms within the square brackets. Substituting into (A6) from (A8) and (A9) we have

$$\frac{\partial(A-3)}{\partial f_c} = \sum_{m=1}^N \sum_{p=1}^N \left[\frac{\partial s(b_m)}{\partial b_m} h_{mc} r_{mp} s(b_p) + s(b_m) r_{mp} \frac{\partial s(b_p)}{\partial b_p} h_{pc} \right]. \quad (\text{A12})$$

We see that it is possible to write (A12) as a vector matrix-product and we have

$$\begin{aligned} \frac{\partial(A-3)}{\partial f} &= H^T S_b R_n^{-1} s(Hf) + [s(Hf)^T R_n^{-1} S_b H]^T \\ &= H^T S_b R_n^{-1} s(Hf) + H^T S_b^T [R_n^{-1}]^T s(Hf). \end{aligned}$$

But R_n is a covariance matrix and hence symmetric. Thus $R_n^{-1} = [R_n^{-1}]^T$. S_b is diagonal, so

$$\frac{\partial(A-3)}{\partial f} = 2 H^T S_b R_n^{-1} s(Hf). \quad (\text{A13})$$

Using (A10) and (A12) it is now possible to write

$$\nabla \psi(f) = H^T S_b R_n^{-1} [g - s(Hf)] - R_f^{-1} [f - \bar{f}] \quad (\text{A14})$$

and we see that $\nabla \psi = \phi$ can be rearranged to yield (16) in the main body of the paper.

ACKNOWLEDGMENT

The author wishes to thank the referees for their helpful comments.

REFERENCES

- [1] H. C. Andrews, "Digital image restoration: A survey," *Computer*, vol. 7, pp. 36–45, May 1974.
- [2] B. R. Hunt, "Digital image processing," *Proc. IEEE*, vol. 63, pp. 693–708, Apr. 1975.
- [3] N. E. Nahi and T. Assefi, "Bayesian recursive image estimation," *IEEE Trans. Computers*, vol. C-21, pp. 734–738, July 1972.
- [4] A. Habibi, "Two-dimensional Bayesian estimate of images," *Proc. IEEE*, vol. 60, pp. 878–883, July 1972.
- [5] A. K. Jain and E. Angel, "Image restoration, modelling, and reduction of dimensionality," *IEEE Trans. Computers*, vol. C-23, pp. 470–476, May 1974.
- [6] C. E. K. Mees, *The Theory of the Photographic Process*. New York: Macmillan, 1954.
- [7] L. M. Biberman and S. Nudelman, *Photoelectronic Imaging Devices*, vols. 1 and 2. New York: Plenum, 1971.
- [8] M. Schwartz, *Information Transmission, Modulation and Noise*. New York: McGraw-Hill, 1962.
- [9] J. F. Walkup and R. C. Choens, "Image processing in signal-dependent noise," *Opt. Eng.*, vol. 13, pp. 258–266, May 1974.
- [10] B. R. Hunt, "The application of constrained-least-squares estimation to image restoration by digital computer," *IEEE Trans. Computers*, vol. C-22, pp. 805–812, Sept. 1973.
- [11] P. Wintz, "Transform picture coding," *Proc. IEEE*, vol. 60, pp. 809–820, 1972.
- [12] H. L. VanTrees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968.
- [13] B. R. Hunt and H. C. Andrews, "Comparison of different filter structures for restoration of images," *Proc. HICCS Conf.*, Honolulu, HI, Univ. of Hawaii, Jan. 1973.
- [14] T. S. Huang, W. F. Schreiber, and O. J. Tretiak, "Image processing," *Proc. IEEE*, vol. 59, pp. 1586–1609, Nov. 1971.
- [15] T. Saaty and J. Bram, *Nonlinear Mathematics*. New York: McGraw-Hill, 1964.
- [16] R. Fox, *Optimization Methods for Engineering Design*. New York: Addison-Wesley, 1971.
- [17] B. R. Hunt, "Nonlinear hybrid digital/optical processors for restoration of images," in *Proc. 3rd Int. Conf. Optical Computing*, Washington, DC, Apr. 1975.
- [18] R. M. Gray, "Toeplitz and circulant matrices: A review," Stanford Univ., Stanford, CA, Rep. SU-SEL-71-032, June 1971.



B. R. Hunt was born on August 24, 1941, in McAllister, OK. He received the B.Sc. degree (magna cum laude) in aeronautical engineering from Wichita State University, Wichita, KS, 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.

In 1967 he joined the staff of Sandia Laboratory in Albuquerque, NM, where he was engaged in systems analysis of strategic weapons problems. In 1968 he joined the staff of the Los Alamos Scientific Laboratory of the University of California in Los Alamos, NM, where he worked in the field of research and applications of digital image processing. In 1973 he became Assistant and then Alternate Group Leader of image processing, and has been responsible for directing and conducting research in image processing, supervision of software, and systems development for image processing laboratories, and development of sources of funding for image processing research. In the fall of 1975 he joined the University of Arizona as Associate Professor in the Department of Systems and Industrial Engineering. In 1976 he was also appointed as Associate Professor in the Optical Sciences Center at Arizona. He has authored numerous papers on image processing and the related areas of computing and system theory. He has lectured at many seminars before government, industrial, and university groups, and in intensive university short courses. He has just completed coauthoring (with Dr. H. C. Andrews of the University of Southern California) a book on digital image restoration and enhancement, which will be the first book-length treatment devoted to this topic. He has been a leader in the transfer of research results in working applications in image processing. He has also been a leader in the development of general-purpose software for image processing applications on large-scale digital computers.