LA-6516-T
Thesis

c.3

UC-32

Issued: September 1976

Improved Methods of Maximum A Posteriori Image Restoration

by

Henry J. Trussell

los alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87545

An Affirmative Action/Equal Opportunity Employer

UNITED STATES
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
CONTRACT W-7405-ENG. 36

This thesis was submitted to the University of New Mexico, Albuquerque, NM, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics and Computer Science. It is the independent work of the author and has not been edited by the Technical Information staff.

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161
Price: Printed Copy $5.00 Microfiche $2.25

TABLE OF CONTENTS

List of Figures

List of Tables

Notations

a, A	lower or upper case letters represent scalar quantities
$\underset{\sim}{a}$	lower case letters with tilde represent vectors
$\underset{\sim}{A}$	upper case letters with tilde represent matrices
$\mathcal{A}$	script letters represent the Fourier transform of the function represented by the same nonscript character
$\|\underset{\sim}{a}\|$	double bars represent the Euclidian norm of the vector $\underset{\sim}{a}$

ABSTRACT

IMPROVED METHODS OF MAXIMUM A POSTERIORI
IMAGE RESTORATION

Henry Joel Trussell
Department of Electrical Engineering and Computer Science
The University of New Mexico, 1976

This dissertation offers an explanation of why and how maximum a posteriori (MAP) image restoration works. A new numerical iteration scheme is derived on the basis of this explanation with a convergence criterion related to the noise process. This numerical scheme is very similar to the modified Picard's method given in standard numerical analysis texts. The effects of the parameters of the MAP restoration method are predicted by the explanation and confirmed by experiment.

The MAP restoration method is further improved by using local processing, i.e., processing small sections of the image sequentially and piecing them together to form the restored image. It was found, as predicted, that smaller section sizes result in better restorations. Computing costs increase as section size decreases.

Although the improved MAP restoration scheme is over ten times more efficient than the old scheme, the method remains costly compared to more standard techniques. The method can be used as a refinement technique by postprocessing the output of other restoration methods.

CHAPTER 1

Introduction and Background

1.1 Introduction

The field of image processing has expanded considerably in recent years. The most highly publicized results from this field are the digitally enhanced images of the moon and Mars taken from space exploration vehicles. This work was done by the Jet Propulsion Laboratory (JPL) of the California Institute of Technology. While image processing has long been a necessary tool at JPL, many other installations are beginning to use results from image processing as an important part of their standard data analysis. These applications vary widely in complexity and sophistication. Among the most elementary applications of image processing is the automatic analysis of oscilloscope traces. At the other end of the image processing spectrum is restoration of an image degraded by unknown atmospheric turbulence in the presence of noise. Applications of image processing techniques are being studied in the biomedical field and in industrial quality control.

While new applications for standard image processing techniques are being discovered at an increasing rate, image processors have noted an increasing demand for solutions to problems requiring new techniques and models. In response to these demands, research is currently being conducted toward improving old algorithms, modifying existing methods to fit new conditions, and developing new algorithms using new models. This dissertation represents research to improve an algorithm developed for use with a new image model.

1.2 Description of Problem

Most restoration schemes in the field of image processing use a linear system for the image formation model as Fig. 1 illustrates,

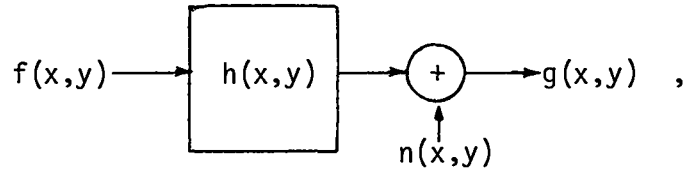

Fig. 1

Linear Model

where the image g(x,y) and the object f(x,y) may be considered intensity functions of two spatial dimensions; h(x,y) is the point spread function; n(x,y) is noise. The noise is modeled as additive and signal independent, although this does not reflect reality in most cases. The most important example of the failure of this model is visible light photography where n(x,y) is film grain noise.

Because greater accuracy was desired, Hunt [10] developed a restoration algorithm using a more nearly correct image formation model, given by Fig. 2,

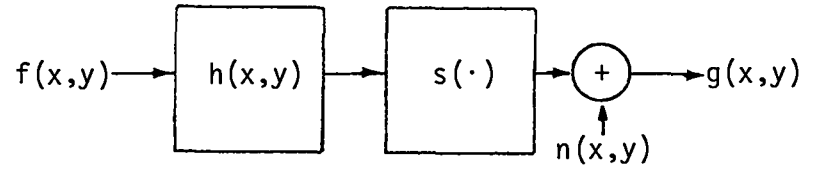

Fig. 2

Nonlinear Model

where s(·) is a point function operating on values at a point of a two-dimensional function, independent of the values at neighboring points. This

model still considers only signal-independent noise, but the noise is now additive in the correct domain.

Hunt used this model in developing the maximum a posteriori (MAP) restoration method. The attempt was made to use the maximum amount of a priori knowledge and the most realistic image formation model that mathematical tractability would allow. In [10] the restoration problem of maximizing the posterior probability density function was solved given Gaussian distributions on the undegraded image ensemble and on the noise. While the problem was defined and a mathematically derived estimate of the solution was given, several problem areas remained: (1) the solution algorithm was so slow that only small problems could feasibly be attempted on one of the largest and fastest computers available today, (2) the method lacked physical understanding of the effects of the parameters of importance, and (3) the solution was not an outstanding restoration in spite of the more accurate image formation model. These are the problems addressed by this dissertation.

1.3 Summary of Dissertation

Background on previous work using the linear model is presented in the next section of this chapter. A description of the restoration problem and a review of the previous work on MAP restoration is discussed in Chapter 2. The nonlinear problem and the MAP solution to it are also the subject of Chapter 2.

The new work presented in this dissertation falls into three main parts: (1) heuristic explanation of the MAP restoration method, (2) application of this understanding to derive a more efficient algorithm for

computing the MAP restoration, and (3) application of local processing to the MAP restoration method to achieve further improvements in the restored image. The heuristic explanation enables us to predict correctly the influence of the MAP parameters on the solution and on the behavior of the iteration scheme. The explanation also enables us to derive a more efficient iteration scheme than the one used by Hunt [10] and to propose a convergence criterion related to the noise process. The new solution method converges faster and gives a more pleasing restoration than the original method. The faster computation has made the MAP restoration method a practical tool for the image processing community. The convergence criterion can be used to judge the quality of the restoration. Preliminary results from current studies at the Los Alamos Scientific Laboratory indicate that the new MAP restoration method produces images of as good as or better visual quality for nonsingular blurs than restorations produced by more conventional processing techniques. The MAP restorations from images degraded by singular blurs, such as out-of-focus lens blurs, were much superior to conventional restorations. The new heuristic explanation of the MAP restoration method and the derivation of the new iteration scheme are discussed in Chapter 3. Experimental results of studies of the MAP parameters and timing data are also presented in Chapter 3.

The iteration scheme addressed in Chapter 3 was derived from the heuristic explanation in that chapter. In Chapter 4 it is shown that this iteration scheme is nearly the same as a standard numerical technique presented in Isaacson and Keller [12]. A further discussion is made in Chapter 4 relating behavior of the iteration scheme to certain approximations which were made in the course of the derivation.

Chapter 5 presents a short introduction to local processing in contrast to global processing. In this chapter it is shown why local processing should theoretically be beneficial to the MAP restoration process. Results are presented which demonstrate the improvements gained by local processing. Finally, the trade-offs between computation time and restoration quality as a function of the size of the local area to be processed are discussed.

Conclusions and topics for further research are discussed in Chapter 6.

1.4 The Linear Problem

The problem which has been addressed by previous restoration methods is based on the model given by Fig. 1, that is, given a detected or measured image $\underset{\sim}{g}$ which was formed by the given model, find an estimate of $\underset{\sim}{f}$, the original scene or object. As noted in the model given in Fig. 1, an image is usually considered a function of two spatial variables. This function can be represented in the discrete-discrete space [1] by a matrix. This matrix can be represented as a vector by lexicographically ordering the column of the matrix, i.e., the i,j^{th} element of the mxn matrix is the $[(j-1)m+i]^{th}$ element of the vector. This ordering permits the use of the simple matrix model

$$\underset{\sim}{g} = \underset{\sim}{H}\underset{\sim}{f} + \underset{\sim}{n} \quad , \tag{1-1}$$

where $\underset{\sim}{g}$ is the recorded or measured image,

$\underset{\sim}{f}$ is the original undegraded scene,

$\underset{\sim}{n}$ is additive noise,

$\underset{\sim}{H}$ is a transformation representing the blurring phenomenon; the vectors $\underset{\sim}{g}$, $\underset{\sim}{f}$, and $\underset{\sim}{n}$ are of length N and the matrix $\underset{\sim}{H}$ is of dimension N x N.

This model could also be described as a continuous model

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y,u,v)\, F(u,v)\, du\,dv + n(x,y) \quad , \tag{1-2}$$

where g, f, n, h represent the same quantities but in continuous space and (x,y) and (u,v) represent the two independent spatial coordinates. Equation (1-1) is the discrete version of Eq. (1-2) which is implemented on digital computers. This is the form of the problem which will be addressed in this dissertation.

1.5 Solutions to the Linear Problem

Solutions to the linear problem are discussed for completeness of the dissertation. A comparison of two linear restoration techniques and the MAP restoration method is made in Chapter 3. The standard linear solution methods use the model given by Eq. (1-1) with the additional assumption that the image $\underset{\sim}{f}$ and the blur $\underset{\sim}{H}$ are stationary. This latter makes $h(x,y,u,v) = h(x-u,y-v)$, a convolution and permits the use of the Fourier transform.

The classical method of solving a nondeterministic problem is to minimize the expected value of the mean square error between the solution (original scene) $\underset{\sim}{f}$ and the estimate of the solution $\hat{\underset{\sim}{f}}$. This yields the Wiener estimate [7] which is usually described in the frequency domain as

$$\mathcal{F}(m,n) = \mathcal{H}_W(m,n)\,\mathcal{G}(m,n) \quad , \tag{1-3}$$

where the restoring filter is

$$\mathcal{H}_W(m,n) = \frac{\overline{\mathcal{H}}(m,n)\Phi_f(m,n)}{\Phi_g(m,n)} \quad .$$

The script letters denote the two-dimensional discrete Fourier transforms of the corresponding lower case quantities, the $\underset{\sim}{\Phi}$'s denote the power spectrum of the subscripted quantities and the overbar represents complex conjugate. Since $\underset{\sim}{\Phi}_g$ can be estimated from the given image $\underset{\sim}{g}$, the quantities which must be known a priori are $\mathcal{H}$ and $\underset{\sim}{\Phi}_f$. Cole [3] has shown that estimates of $\underset{\sim}{\Phi}_f$ can be made from images of similar objects. Cannon [2] has established methods for estimating $\mathcal{H}$ for certain types of blurs.

Constrained least squares estimation [9] eliminates the need for an estimate of $\underset{\sim}{\Phi}_f$. This estimate is found by solving the minimization problem

$$\begin{aligned} &\text{minimize:} && \underset{\sim}{f}^T \underset{\sim}{C}^T \underset{\sim}{C}\, \underset{\sim}{f} \\ &\text{subject to:} && [\underset{\sim}{g} - \underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}]^T [\underset{\sim}{g} - \underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}] = e \end{aligned} \qquad (1\text{-}5)$$

where $\underset{\sim}{C}$ is a constraint matrix and e is proportional to the noise variance. Heuristically, if $\underset{\sim}{C}$ represents a differentiation process, the solution $\underset{\sim}{f}$ is the smoothest estimate such that the variance of the residual $\underset{\sim}{g} - \underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}$ is equal to the variance of the noise. The solution filter for the problem (1-5) is given in the frequency domain as

$$\mathcal{H}_C(m,n) = \frac{\overline{\mathcal{H}}(m,n)}{|\mathcal{H}(m,n)|^2 + \lambda |\mathcal{C}(m,n)|^2} \quad , \qquad (1\text{-}6)$$

where λ is a Lagrange multiplier which is determined by iteration. The only quantity which must be known a priori is $\mathcal{H}$.

A method which requires even less a priori knowledge is homomorphic filtering [3]. This method equalizes the power spectrum of the solution estimate to that of a prototype of the original undegraded scene $\underset{\sim}{f}$. The homomorphic filter can be described in the frequency domain as

$$\mathcal{H}_H(m,n) = \sqrt{\frac{\Phi_f(m,n)}{\Phi_g(m,n)}} \quad . \tag{1-7}$$

Cannon [2] extended this filter to include estimation of the phase for specific point spread functions, $\underset{\sim}{H}$.

The above restoration methods use the linear model Fig. 1. All have been successfully applied to realistic digital image restoration problems.

CHAPTER 2

Previous Work on MAP

2.1 Introduction

In an attempt to use a more realistic image model and the maximum amount of a priori knowledge for image restoration, Hunt [10] proposed the maximum a posteriori (MAP) estimate. The proposed model takes into account the nonlinear nature of most image recording devices, a factor missing from most previous restoration work. The following is a review of work done in [10] presented here for background and completeness.

2.2 The Nonlinear Model

While the model can be described in any of three spaces, continuous-continuous, continuous-discrete, and discrete-discrete, we will use only the discrete-discrete space because of our implementation of a digital computer to find the MAP solution. The model in this space is given by

$$\underset{\sim}{g} = s(\underset{\sim}{\underset{\sim}{H}}\,\underset{\sim}{f}) + \underset{\sim}{n}$$

where

$\underset{\sim}{g}$ is the recorded image measured in units of film density;

$\underset{\sim}{f}$ is the original scene measured in units of light intensity;

$\underset{\sim}{n}$ is film grain noise measured in units of film density;

$\underset{\sim}{H}$ is the blurring matrix or point spread function;

s is the transformation from light intensity to film density.

A picture is usually represented as a two-dimensional matrix. Here, it is more convenient to lexicographically order the columns to produce a vector. Thus, if our original picture were LxM points, the vectors $\underset{\sim}{g}$, $\underset{\sim}{f}$, and $\underset{\sim}{n}$ are of length LM and the matrix H is LMxLM. We assume $\underset{\sim}{f}$ (or equivalently, the original picture) was appropriately padded with zeroes so the product $\underset{\sim}{H}\underset{\sim}{f}$ represents convolution.

The transformation of the LM vector $\underset{\sim}{b} = \underset{\sim}{H}\underset{\sim}{f}$ by s is the LM vector generated by transforming each element of $\underset{\sim}{H}\underset{\sim}{f}$ by $\underset{\sim}{s}$, i.e.,

$$\underset{\sim}{c} = s(\underset{\sim}{b})$$

where

$$c_i = s(b_i) \qquad i = 1,2,\ldots,LM \ .$$

2.3 Assumptions for MAP Restoration

Since we have noted above it is most convenient to work with $\underset{\sim}{g}$, $\underset{\sim}{f}$, $\underset{\sim}{n}$ as vectors, let N denote the length of these vectors.

The most basic assumption is that the noise, given by the vector $\underset{\sim}{n}$, is an additive, zero mean, Gaussian random process with constant variance. This is a simplification of the true behavior of film grain noise which is more closely modeled by

$$n_i = k\eta_i d_i^p$$

where

k is a constant dependent on film type;

η_i is a standard normal random variable, i.e., $\mu = 0$, $\sigma = 1$;

d_i is the film density at position i;

$P = 1/2$ [6] or $P = 1/3$ [8] .

However, the simplified model has been used successfully for some time [13]. By using the simple model we gain mathematical tractability and recent results [14] have shown very little loss in restoration quality. Thus, we will assume a multivariate normal probability density with zero mean and covariance $\underset{\sim}{R}_n$:

$$p(\underset{\sim}{n}) = \left((2\pi)^{N/2}|\underset{\sim}{R}_n|^{1/2}\right)^{-1} \exp\left(-1/2\underset{\sim}{n}^T\underset{\sim}{R}_n^{-1}\underset{\sim}{n}\right) . \qquad (2\text{-}1)$$

The vector $\underset{\sim}{f}$, the original scene, is assumed to have been drawn from a multivariate normal population with mean $\underset{\sim}{\bar{f}}$ and covariance $\underset{\sim}{R}_f$:

$$p(\underset{\sim}{f}) = \left((2\pi)^{N/2}|\underset{\sim}{R}_f|^{1/2}\right)^{-1} \exp\left(-1/2(\underset{\sim}{f}-\underset{\sim}{\bar{f}})^T\underset{\sim}{R}_f^{-1}(\underset{\sim}{f}-\underset{\sim}{\bar{f}})\right) . \qquad (2\text{-}2)$$

This assumption has the disadvantage that it presents the possibility of negative values and for some images is unrealistic. However, it has the important property of mathematical tractability and may be rendered plausible by the following example. Consider an ensemble of driver's license photographs. Assuming that the faces have been registered by the photographer to have approximately the same size and orientation, then we would have a nonuniform mean $\underset{\sim}{\bar{f}}$. Histograms of images show that intensities can sometimes be modeled by Gaussian statistics [11]. It is not necessary at this stage to assume stationarity of the covariance but

this assumption will be of use in making certain approximations in the numerical computation of the MAP solution.

2.4 Derivation of MAP Equation

It should be noted here that the only assumptions we have made thus far are the model form

$$\underset{\sim}{g} = s(\underset{\sim}{\underset{\sim}{H}}\,\underset{\sim}{f}) + \underset{\sim}{n}$$

and Gaussian statistics for $\underset{\sim}{n}$ and $\underset{\sim}{f}$.

Given the recorded image, sampled as $\underset{\sim}{g}$, the probability density of $\underset{\sim}{f}$ conditioned on $\underset{\sim}{g}$ is given by Bayes' law

$$p(\underset{\sim}{f}|\underset{\sim}{g}) = \frac{p(\underset{\sim}{g}|\underset{\sim}{f})\; p(\underset{\sim}{f})}{p(\underset{\sim}{g})} \quad . \tag{2-3}$$

The maximum a posteriori (MAP) estimate is derived by differentiating with respect to $\underset{\sim}{f}$ and equating the result to zero. As is usual in the case of Gaussian statistics, we will first take the logarithm of both sides and then maximize. This yields

$$\frac{\partial \ln[p(\underset{\sim}{f}|\underset{\sim}{g})]}{\partial \underset{\sim}{f}} = \frac{\partial \ln[p(\underset{\sim}{g}|\underset{\sim}{f})]}{\partial \underset{\sim}{f}} + \frac{\partial \ln[p(\underset{\sim}{f})]}{\partial \underset{\sim}{f}} - \frac{\partial \ln[p(\underset{\sim}{g})]}{\partial \underset{\sim}{f}} = 0 \tag{2-4}$$

Noting that $p(\underset{\sim}{g})$ is independent of $\underset{\sim}{f}$ and substituting in the quantities $p(\underset{\sim}{g}|\underset{\sim}{f})$ from Eq. (2-1) and $p(\underset{\sim}{f})$ from Eq. (2-2), we have

$$\frac{\partial\left[-1/2\left(\underset{\sim}{g}-s(\underset{\sim}{H}\underset{\sim}{f})\right)^T\underset{\sim}{R}_n^{-1}\left(\underset{\sim}{g}-s(\underset{\sim}{H}\underset{\sim}{f})\right)\right]}{\partial\underset{\sim}{f}} + \frac{\partial\left[-1/2(\underset{\sim}{f}-\bar{\underset{\sim}{f}})^T\underset{\sim}{R}_f^{-1}(\underset{\sim}{f}-\bar{\underset{\sim}{f}})\right]}{\partial\underset{\sim}{f}} = 0 \quad . \tag{2-5}$$

We first concentrate upon the first term. Expanding and eliminating terms independent of $\underset{\sim}{f}$ we have

$$\frac{\partial}{\partial\underset{\sim}{f}}\left[s(\underset{\sim}{H}\underset{\sim}{f})^T\underset{\sim}{R}_n^{-1}\underset{\sim}{g}-1/2s(\underset{\sim}{H}\underset{\sim}{f})^T\underset{\sim}{R}_n^{-1}s(\underset{\sim}{H}\underset{\sim}{f})\right] \quad .$$

Writing the terms as summations

$$s(\underset{\sim}{H}\underset{\sim}{f})^T\underset{\sim}{R}_n^{-1}\underset{\sim}{g} = \sum_{m=1}^{N}\sum_{p=1}^{N} s\left(\sum_{t=1}^{N} h_{mt}f_t\right) r_{mp}g_p \tag{2-6}$$

and

$$1/2\ s(\underset{\sim}{H}\underset{\sim}{f})^T\underset{\sim}{R}_n^{-1}s(\underset{\sim}{H}\underset{\sim}{f}) = 1/2\sum_{m=1}^{N}\sum_{p=1}^{N} s\left(\sum_{t=1}^{N} h_{mt}f_t\right) r_{mp} s\left(\sum_{u=1}^{N} h_{pu}f_u\right) \tag{2-7}$$

where r_{mp} is the mp^{th} element of the matrix $\underset{\sim}{R}_n^{-1}$. Taking derivatives we have

$$\frac{\partial s(\underset{\sim}{H}\underset{\sim}{f})^T\underset{\sim}{R}_n^{-1}\underset{\sim}{g}}{\partial f_k} = \sum_{m=1}^{N}\sum_{p=1}^{N} \frac{\partial s(x)}{\partial x}\bigg|_{x=b_m} \frac{\partial b_m}{\partial f_k}\ r_{mp}g_p \quad , \tag{2-8}$$

where $b_m = \sum_{t=1}^{N} h_{mt}f_t$; (2-9)

$$\frac{\partial \frac{1}{2}\, s(\underset{\sim}{H}\underset{\sim}{f})^T \underset{\sim}{R}_n^{-1}\, s(\underset{\sim}{H}\underset{\sim}{f})}{\partial f_k} = 1/2 \sum_{m=1}^{N} \sum_{p=1}^{N} \left[\left.\frac{\partial s(x)}{\partial x}\right|_{x=b_m} \frac{\partial b_m}{\partial f_k}\; r_{mp}\; s(b_p) \right.$$

$$\left. + s(b_m) r_{mp} \left.\frac{\partial s(x)}{\partial x}\right|_{x=b_p} \frac{\partial b_p}{\partial f_k} \right] \quad , \tag{2-10}$$

where b_m is as Eq. (2-9). We have from (2-9)

$$\frac{\partial b_m}{\partial f_k} = h_{mk} \quad . \tag{2-11}$$

We can write Eqs. (2-6) and (2-7) as vector matrix products. Substituting (2-9) into (2-8) we have

$$\frac{\partial s(\underset{\sim}{H}\underset{\sim}{f})^T \underset{\sim}{R}_n^{-1} \underset{\sim}{g}}{\partial \underset{\sim}{f}} = \underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1}\, \underset{\sim}{g} \quad , \tag{2-12}$$

where $\underset{\sim}{S}_b$ is a diagonal matrix of derivatives

$$\underset{\sim}{S}_b = \begin{bmatrix} \left.\frac{\partial s(x)}{\partial x}\right|_{x=b_1} & & \emptyset \\ & \left.\frac{\partial s(x)}{\partial x}\right|_{x=b_2} & \\ & & \ddots \\ \emptyset & & \left.\frac{\partial s(x)}{\partial s}\right|_{x=b_N} \end{bmatrix} \tag{2-13}$$

and the b_i are as defined in Eq. (2-9).

Substituting Eq. (2-11) into Eq. (2-10) we have

$$\frac{\partial \frac{1}{2}\, s(\underset{\sim}{H}\underset{\sim}{f})^T \underset{\sim}{R}_n^{-1}\, s(\underset{\sim}{H}\underset{\sim}{f})}{\partial f_k} = 1/2 \sum_{m=1}^{N} \sum_{p=1}^{N} \left.\frac{\partial s(x)}{\partial x}\right|_{x=b_m} h_{mk} r_{mp} s(b_p)$$

$$+ s(b_m) r_{mp} \left.\frac{\partial s(x)}{\partial s}\right|_{x=b_p} h_{pk} \quad , \tag{2-14}$$

which we can rewrite as the vector matrix product

$$\frac{\partial \frac{1}{2}\, s(\underset{\sim}{H}\underset{\sim}{f})^T \underset{\sim}{R}_n\, s(\underset{\sim}{H}\underset{\sim}{f})}{\partial \underset{\sim}{f}} = 1/2\, \underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} s(\underset{\sim}{H}\underset{\sim}{f}) + 1/2\, [s(\underset{\sim}{H}\underset{\sim}{f})^T \underset{\sim}{R}_n^{-1} \underset{\sim}{S}_b \underset{\sim}{H}]^T$$

$$= 1/2\, \underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} s(\underset{\sim}{H}\underset{\sim}{f}) + 1/2\, \underset{\sim}{H}^T \underset{\sim}{S}_b^T [\underset{\sim}{R}_n^{-1}]^T s(\underset{\sim}{H}\underset{\sim}{f})$$

$$= \underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1}\, s(\underset{\sim}{H}\underset{\sim}{f}) \quad , \tag{2-15}$$

since $\underset{\sim}{S}_b^T = \underset{\sim}{S}_b$ and $[\underset{\sim}{R}_n^{-1}]^T = \underset{\sim}{R}_n^{-1}$.

Now evaluating the second term of Eq. (2-5) we have

$$\frac{\partial\, [-1/2(\underset{\sim}{f}-\bar{\underset{\sim}{f}})^T \underset{\sim}{R}_f^{-1} (\underset{\sim}{f}-\bar{\underset{\sim}{f}})]}{\partial \underset{\sim}{f}} = -\, \underset{\sim}{R}_f^{-1} (\underset{\sim}{f}-\bar{\underset{\sim}{f}}) \quad . \tag{2-16}$$

Combining (2-12), (2-15), and (2-16) in Eq. (2-5) we get the fundamental MAP equation:

$$\underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} [\underset{\sim}{g} - s(\underset{\sim}{H}\underset{\sim}{f})] - \underset{\sim}{R}_f^{-1}\, [\underset{\sim}{f}-\bar{\underset{\sim}{f}}] = 0 \quad . \tag{2-17}$$

The nonlinear nature of s in (2-17) does not allow it to be simplified

further, that is, there is no direct closed-form solution for the restored $\underset{\sim}{f}$.

It is noteworthy that Eq. (2-17) does not contain $\underset{\sim}{H}^{-1}$; thus $\underset{\sim}{H}$ being singular or ill-conditioned does not impair our computation. If s is linear, it is possible to rearrange Eq. (2-17) to yield

$$\underset{\sim}{f} = \left(\underset{\sim}{R}_f^{-1} + \underset{\sim}{H}^T\underset{\sim}{R}_n^{-1}\underset{\sim}{H}\right)^{-1} \underset{\sim}{H}^T\underset{\sim}{R}_n^{-1}\underset{\sim}{g} + \left(\underset{\sim}{R}_f^{-1} + \underset{\sim}{H}^T\underset{\sim}{R}_n^{-1}\underset{\sim}{H}\right)^{-1} \underset{\sim}{R}_f^{-1}\,\bar{\underset{\sim}{f}}$$

which is recognized as a discrete Wiener filter with the a priori mean included. Note that the case $s(x) = x$ is sufficient to cover all linear cases. If $s(x) = ax + b$, the slope a could be included in $\underset{\sim}{H}$ and the bias b could be subtracted.

2.5 Computational Aspects and Further Assumptions

Up to this point we have assumed very little about the structure of $\underset{\sim}{H}$, $\underset{\sim}{R}_n$, and $\underset{\sim}{R}_f$. To make computation of MAP estimates for reasonable size images feasible, it is necessary to make certain assumptions which simplify the computation.

We will assume the point spread function is space-invariant. This is quite reasonable and is true for a wide variety of blurs. This causes the matrix $\underset{\sim}{H}$ to be block Toeplitz.

We will assume stationarity for the noise process and for the process yielding the original scene $\underset{\sim}{f}$. The noise process is stationary about zero. However, the scene $\underset{\sim}{f}$ is from a process which is stationary about $\bar{\underset{\sim}{f}}$. This forces the matrices $\underset{\sim}{R}_n$ and $\underset{\sim}{R}_f$ to be block Toeplitz. We will further assume the noise is uncorrelated, forcing $\underset{\sim}{R}_n$ to be diagonal.

The assumptions on the noise are reasonable, standard and need no further justification.

In the case of the random process yielding the original scene $\underset{\sim}{f}$, we have assumed a Gaussian process with a nonuniform mean $\bar{\underset{\sim}{f}}$ but stationary covariance $\underset{\sim}{R}_f$. Since we have only one member of this ensemble, we have considerable latitude in defining the ensemble from which it came. The assumption is clearly for computational convenience but is not unreasonable.

Because of the block Toeplitz forms of the matrices $\underset{\sim}{H}$, $\underset{\sim}{R}_n$, $\underset{\sim}{R}_f$ we can use block circulant approximations. This approximation permits the use of the fast Fourier transform (FFT) for rapid computation of the matrix products [9].

Since a direct solution was impossible, two iterative schemes were tried. Picard's method relies upon rewriting Eq. (2-17) to yield an equation in which the solution, $\underset{\sim}{f}_{MAP}$, is a fixed point, i.e.,

$$\underset{\sim}{f}_{MAP} = \Phi(\underset{\sim}{f}_{MAP}) = \left(\phi_1(\underset{\sim}{f}_{MAP}),\ldots,\phi_N(\underset{\sim}{f}_{MAP})\right)^T \quad .$$

Such a formulation is given by

$$\underset{\sim}{f}_{MAP} = \bar{\underset{\sim}{f}} + \underset{\sim}{R}_f \underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} [\underset{\sim}{g} - s(\underset{\sim}{H}\underset{\sim}{f}_{MAP})] \quad . \tag{2-18}$$

The solution is obtained by the iteration

$$\underset{\sim}{f}^{(k+1)} = \Phi\left(\underset{\sim}{f}^{(k)}\right) \quad . \tag{2-19}$$

Convergence of this method is guaranteed if in a neighborhood of the solution

$$\sum_{j=1}^{N} \left| \frac{\partial \phi_i}{\partial f_j} \right| < 1 \quad \text{for } i = 1,2,\ldots,N \quad , \tag{2-20}$$

and the initial supposition, $\underset{\sim}{f}_0$, lies in this neighborhood.

Unfortunately, Eqs. (2-20) are not of much utility for it is necessary to guess the solution to get the proper neighborhood. The elements of the summations in Eqs. (2-20) are very complex and would prove extremely time-consuming to compute for any guess $\underset{\sim}{f}_0$. Thus, it was decided to simply test this scheme for several cases with known solutions. Results from these tests show divergence even when slight deviations of the known solution were used for $\underset{\sim}{f}_0$.

The second scheme tried was the steepest ascent method. This iteration scheme is given by

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{(k)} + \alpha^{(k)} \underset{\sim}{\nabla}\psi\left(\underset{\sim}{f}^{(k)}\right) \quad , \tag{2-21}$$

where ψ is the function to be maximized and $\alpha^{(k)}$ is determined to speed convergence. In the MAP case ψ is given by (2-3) and (2-4), and $\underset{\sim}{\nabla}\psi$ is given by (2-17). Substituting the left-hand side of (2-17) for $\underset{\sim}{\nabla}\psi$ into (2-21) we have

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{(k)} + \alpha^{(k)} \left\{ \underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} \left[\underset{\sim}{g} - s(\underset{\sim}{H}\underset{\sim}{f}^{(k)})\right] - \underset{\sim}{R}_f^{-1} \left[\underset{\sim}{f}^{(k)} - \underset{\sim}{\bar{f}}\right] \right\} . \tag{2-22}$$

Convergence is reached if

$$\left\| \underset{\sim}{\nabla}\psi\left(\underset{\sim}{f}^{(k)}\right) \right\| < \varepsilon \quad , \tag{2-23}$$

where ε is a tolerance on the error in the final gradient. This method proved to be computationally feasible and test results showed convergence over a wide range of initial guesses.

2.6 Estimating the Parameters $\alpha^{(k)}$, $\underset{\sim}{R}_n$, $\underset{\sim}{R}_f$, and $\bar{\underset{\sim}{f}}$

The optimum choice of $\alpha^{(k)}$ can be determined theoretically and used to take the maximum step in the maximum gradient direction at each step k. However, computation of this optimum is often impossible and with functions as complex as (2-22) is extremely difficult numerically. It was decided to use a fixed value, $\alpha^{(k)} = \alpha$. The value of α was determined by trial and error, i.e., if the method diverged after a few iterations, a smaller α was chosen.

Since the noise was assumed to be an additive, white, Gaussian, zero mean process, it is necessary only to determine a variance. This was done by determining the sample variance on a small, relatively flat portion of the image.

Statistically, one way to estimate the mean of an ensemble given a single sample is $\hat{\bar{\underset{\sim}{f}}} = s^{-1}(\underset{\sim}{g})$. However, let us consider again the ensemble of drivers' license photographs. All faces are clearly not alike; but if properly registered, the eyes, ears, noses, mouths, and hair should appear in the same areas. One would presume that the mean image of this ensemble would be approximately oval with suitable shadings for eyes, mouth, nose, and hair; in other words, a blurry face. Another reasonable $\bar{\underset{\sim}{f}}$ is, then, a slightly blurred version of the intensity image derived from the measured image g. Both methods of generating $\bar{\underset{\sim}{f}}$ were tried with little effect on the final result.

When $\underset{\sim}{\bar{f}}$ was estimated by a slightly blurred version of $s^{-1}(\underset{\sim}{g})$, $\underset{\sim}{R}_f$ was estimated by computing the auto-covariance of the difference $s^{-1}(\underset{\sim}{g}) - \underset{\sim}{\bar{f}}$. This estimate requires stationarity which was already assumed. This same $\underset{\sim}{R}_f$ was used when $\underset{\sim}{\bar{f}}$ was estimated by $s^{-1}(\underset{\sim}{g})$.

2.7 Computation

This method was tested on a 128 x 128 element picture. Convergence was reached in approximately 60 iterations using about 16 minutes of computation time on a CDC-7600 computer. The restoration compared favorably with other more standard techniques.

CHAPTER 3

Application of Heuristics to Get Improved Iteration Scheme

3.1 Introduction

We have seen in Chapter 2 that Hunt [10] has laid the foundation for maximum a posteriori estimation in nonlinear image restoration. Several problem areas remained and several new ones can be brought up now. An interesting anomaly is that while the MAP method was developed on the basis of a maximum amount of a priori knowledge, very little extra knowledge was actually used in the restoration. Of the parameters required for the restoration, $\underset{\sim}{H}$, $\underset{\sim}{R}_n$, $\underset{\sim}{R}_f$, $\bar{\underset{\sim}{f}}$, only $\underset{\sim}{R}_f$ and $\bar{\underset{\sim}{f}}$, the parameters for the probability density $p(\underset{\sim}{f})$, are not determined by the physics of the image formation system. $\underset{\sim}{R}_n$ and $\underset{\sim}{H}$ can be derived from the sampled image, $\underset{\sim}{g}$ [2]. Hunt calculated both $\bar{\underset{\sim}{f}}$ and $\underset{\sim}{R}_f$ from this same image. So in actual practice, less information was required for MAP restoration than for Wiener restoration.

As part of the question of a priori knowledge we would ask about the significance of $\bar{\underset{\sim}{f}}$ and $\underset{\sim}{R}_f$. Since in practice $\bar{\underset{\sim}{f}}$ and $\underset{\sim}{R}_f$ were estimated by a reasonable heuristic rather than by hard knowledge, it is only natural to seek other methods of generating the parameters. It is also natural to inquire about the criticality of these quantities on the solution and about the effects of perturbing their values. The parameters $\bar{\underset{\sim}{f}}$ and $\underset{\sim}{R}_f$ define the distribution of the original unblurred image $\underset{\sim}{f}$ since this distribution is assumed Gaussian as in Chapter 2. By perturbing their values we imply we are using different estimates of these quantities. To make

the point clear to the reader that in the following discussion we are using estimates and not true values, we shall use the notation $\hat{\underset{\sim}{f}}$ and $\hat{\underset{\sim}{R}}_f$.

Even though Hunt [10] proved that the MAP restoration worked, the results as they stood were not outstanding and hardly justified the exorbitant amount of computing time. Sixteen minutes on one of the world's largest and fastest computers to process a small, 128 x 128, image is enough to curtail the interest of all but the few possessed of comparable computing capacity. It is desirable not only to speed the present computation but to find an algorithm which converges more rapidly.

While a convergence criterion was proposed by Hunt in the initial work on MAP restoration, Eq. (2-23), no method was indicated to choose an appropriate α, the step size parameter. In practice the eyeball served as the convergence criterion, i.e., when further computing failed to produce a better picture, the iterations were stopped. This is unsatisfactory from a mathematical as well as an aesthetic viewpoint.

3.2 Heuristic Explanation of MAP

This section is intended to give the reader insight into the reasons why the MAP restoration method works and what the solution method is trying to accomplish. The assumptions are made that the quantities $\underset{\sim}{g}$, $\underset{\sim}{H}$, and $\underset{\sim}{R}_n$ and the function s are given. This might be the case in a real restoration problem. The quantities $\hat{\underset{\sim}{f}}$ and $\hat{\underset{\sim}{R}}_f$ must always be estimated in a real problem; because of this it is useful to consider different possible estimates.

It is noted that the fundamental MAP equation derived in Chapter 2 and repeated below consists of two terms.

$$\underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} [\underset{\sim}{g}-s(\underset{\sim}{H}\underset{\sim}{f})] - \underset{\sim}{R}_f^{-1} [\underset{\sim}{f} - \bar{\underset{\sim}{f}}] = 0 \quad . \tag{3-1}$$

The first term is derived from the probability distribution, $p(\underset{\sim}{g}|\underset{\sim}{f})$. If this distribution were the only one considered, we would have the maximum likelihood estimate as the solution. We will refer to this term as the maximum likelihood term.

The second term, or MAP term, is derived from the a priori density, $p(\underset{\sim}{f})$. This term distinguishes the MAP estimate from the maximum likelihood estimate. This is also the term about which we have the least amount of information.

It would be ideal if both terms were zero. We could force this by choosing $\hat{\underset{\sim}{f}}$ to be the maximum likelihood estimate, but this would serve no purpose. Obviously, any other choice for $\hat{\underset{\sim}{f}}$ results in a solution of equation (3-1) where neither of the terms is zero. Thus, what a solution must do is to strike a balance between the maximum likelihood solution and $\bar{\underset{\sim}{f}}$. In essence, what we will have for a solution is a nonlinear weighted average of the solutions to each of the two terms.

Let us consider the iteration scheme given in Eq. (2-22) and repeated here

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{(k)} + \alpha \left\{ \underset{\sim}{H}^T \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} [\underset{\sim}{g}-s(\underset{\sim}{H}\underset{\sim}{f}^{(k)})] - \underset{\sim}{R}_f^{-1} [\underset{\sim}{f}^{(k)} - \bar{\underset{\sim}{f}}] \right\} . \tag{3-2}$$

It is reasonable to start the iteration scheme at $\underset{\sim}{f}^{(o)} = \hat{\underset{\sim}{f}}$. The method moves successive $\underset{\sim}{f}^{(k)}$'s away from $\hat{\underset{\sim}{f}}$ and toward the maximum likelihood solution, which is $\underset{\sim}{H}^{-1}s^{-1}(\underset{\sim}{g})$. It would be beneficial to know the path the iterations take from $\hat{\underset{\sim}{f}}$ to $\underset{\sim}{H}^{-1}s^{-1}(\underset{\sim}{g})$; but because of the nonlinearity of s,

this is extremely difficult to predict. The magnitude or norm of $\underset{\sim}{R}_n$ and $\hat{\underset{\sim}{R}}_f$ gives us an indication of our confidence in the two solutions $\underset{\sim}{H}^{-1}s^{-1}(\underset{\sim}{g})$ and $\hat{\bar{\underset{\sim}{f}}}$, respectively. If $\|\underset{\sim}{R}_n\| << \|\hat{\underset{\sim}{R}}_f\|$, then we would have more confidence in the maximum likelihood solution than the a priori mean, $\hat{\bar{\underset{\sim}{f}}}$. The weighting of the two terms by $\underset{\sim}{R}_n^{-1}$ and $\hat{\underset{\sim}{R}}_f^{-1}$ reflects this confidence since $\|\underset{\sim}{R}_n^{-1}\| >> \|\underset{\sim}{R}_f^{-1}\|$ and our solution to Eq. (3-1) would lie closer to $\underset{\sim}{H}^{-1}s^{-1}(\underset{\sim}{g})$.

3.3 Relation of MAP to Constrained Least Squares Estimation

In the preceding section it was noted how little actual knowledge we had about the probability density, $p(\underset{\sim}{f})$, from which the MAP term of Eq. (3-1) was derived. If the inclusion of this term were merely an artifact of clever mathematics, we would not expect the good quality restorations that resulted. The restored images shown by Hunt [10] are certainly better than we could have expected from the maximum likelihood term alone, i.e., $\underset{\sim}{H}^{-1}s^{-1}(\underset{\sim}{g})$. Since one is hardly likely to say that the $\hat{\bar{\underset{\sim}{f}}}$ and $\hat{\underset{\sim}{R}}_f$ used by Hunt were lucky guesses to the actual values, we are left with the question, "Why did those guesses work?"

Let us consider the maximum likelihood estimate, $\underset{\sim}{H}^{-1}s^{-1}(\underset{\sim}{g})$. The outstanding undesirable characteristic of this solution in image restoration is its magnification of noise. In most applied cases the point spread function matrix $\underset{\sim}{H}$ is singular or near singular. In the range of the measured image where the noise power is much greater than the signal power, (usually in the higher frequencies) the inverse filter, $\underset{\sim}{H}^{-1}$, causes the entire restoration to be dominated by the enhanced noise. It would appear

that the MAP term acts as a noise control. This is seen by considering the MAP solution as a weighted average. The solution cannot go too far toward the maximum likelihood solution and "blow up" because of the averaging in of the a priori mean. The MAP term is acting as a constraint.

Let us consider the problem of constrained least squares estimation. This was developed by Hunt [9] for the linear problem:

$$\text{minimize } \underset{\sim}{f}^T \underset{\sim}{C}^T \underset{\sim}{C} \underset{\sim}{f}$$

$$\text{subject to } (\underset{\sim}{g}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})^T (\underset{\sim}{g}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}) = \xi \ ,$$

where $\underset{\sim}{g} = \underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}+\underset{\sim}{n}$ is the measured image,

$\underset{\sim}{f}$ is the original image,

$\underset{\sim}{\underset{\sim}{H}}$ is the point spread function matrix,

$\underset{\sim}{n}$ is additive noise with variance σ^2,

$\underset{\sim}{C}$ is an arbitrary weighting matrix,

ξ is usually $N\sigma^2$ where N is the dimension of the vectors $\underset{\sim}{g}$, $\underset{\sim}{f}$, $\underset{\sim}{n}$.

The method of Langrange multipliers yields the equations

$$\underset{\sim}{C}^T \underset{\sim}{C} \underset{\sim}{f} - \lambda \underset{\sim}{\underset{\sim}{H}}^T (\underset{\sim}{g}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}) = 0$$

$$(\underset{\sim}{g}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})^T (\underset{\sim}{g}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}) - \xi = 0$$

where λ is the Lagrange multiplier.

If we restate the constrained least squares problem above to suit our nonlinear case, including the a priori mean, we have the problem:

$$\text{minimize } (\underset{\sim}{f}-\underset{\sim}{\bar{f}})^T \underset{\sim}{R}_f^{-1} (\underset{\sim}{f}-\underset{\sim}{\bar{f}}) \qquad (3\text{-}3)$$

$$\text{subject to: } [\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})]^T \underset{\sim}{\underset{\sim}{R}}_n^{-1} [\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})] = 1 \ , \tag{3-4}$$

where $\underset{\sim}{f}$, $\underset{\sim}{\bar{f}}$, $\underset{\sim}{g}$, $\underset{\sim}{\underset{\sim}{H}}$, $\underset{\sim}{\underset{\sim}{R}}_n$, $\underset{\sim}{\underset{\sim}{R}}_f$, and s are as defined in Chapter 2. Following the same mathematics as in the linear case, we have the following

$$\underset{\sim}{\underset{\sim}{R}}_f^{-1} (\underset{\sim}{f} - \underset{\sim}{\bar{f}}) - \lambda \underset{\sim}{\underset{\sim}{H}}^T \underset{\sim}{\underset{\sim}{S}}_b \underset{\sim}{\underset{\sim}{R}}_n^{-1} [\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})] = 0 \tag{3-5}$$

$$[\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})]^T \underset{\sim}{\underset{\sim}{R}}_n^{-1} [\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})] - 1 = 0$$

where S_b is defined by Eq. (2-13).

Thus, if $\lambda = 1$, Eq.(3-5) is exactly the fundamental MAP Eq. (3-1). Since the user of the MAP method has some latitude in estimating R_f, he could choose $\underset{\sim}{\underset{\sim}{\hat{R}}}_f$ to force $\lambda = 1$. This would mean the two problems would yield the same solution; however, it also changes the problem slightly by changing the weighting of the $(\underset{\sim}{f} - \underset{\sim}{\bar{f}})$ term of Eq. (3-1). It is later shown that changes in $\underset{\sim}{\underset{\sim}{\hat{R}}}_f$ have little effect on the visual quality of the solution. In any case, the forms of the equations for the two problems are nearly identical.

To summarize the new heuristic findings, the MAP solution is related to the constrained least squares defined by Eqs. (3-3) and (3-4). We are looking for the image which falls on the noise level boundary, i.e., the sphere defined by Eq. (3-4), of the maximum likelihood estimate and minimizes the weighted distance from the mean, Eq. (3-3). The noise level boundary is determined by the variance of the additive noise. While the constraint Eq. (3-4) specifies the boundary of the sphere by the equality, heuristically and computationally it is reasonable to include the interior as well by using the inequality:

$$[\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})]^T \underset{\sim}{\underset{\sim}{R}}_n^{-1} [\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})] \leq 1 \ .$$

3.4 New Iteration Scheme

When choosing an iterative method to solve nonlinear equations, one looks first for the simplest one that works. We have a simple workable scheme given by Hunt, Eq. (3-2). This scheme, however, is slow and furthermore, requires user interaction to choose a proper step size parameter α. It is desirable to find a faster method which can run without user assistance.

While there are methods to choose the optimum α, these methods all require two or more additional computations of the function to be minimized. The savings by such a method must be extraordinary to justify the additional computations of a very complex function. After preliminary studies, it was decided to abandon this approach since the additional computation outweighed the advantages of optimization of α.

Another scheme which computes its own step size is Newton's method. In one dimension this scheme is given by

$$x^{(k+1)} = x^{(k)} - \frac{\phi(x^{(k)})}{\phi'(x^{(k)})}$$

where the solution is a zero of ϕ. For a vector valued function, $\underset{\sim}{\Psi}(\underset{\sim}{f})$, the zero is found by

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{(k)} - \underset{\sim}{J}^{-1}(\underset{\sim}{f}^{(k)})\underset{\sim}{\psi}(\underset{\sim}{f}^{(k)}) \qquad (3\text{-}6)$$

where $J(f^k) = \left(\frac{\partial \psi_i(\underset{\sim}{f}^{(k)})}{\partial f_j}\right)$. The function for which we want to find the zero is the fundamental MAP Eq. (3-1). To use this method would involve taking the derivative of Eq. (3-1). This results in a much more complex function to evaluate; since the goal was to speed the computation, this direct

approach offers no further hope of reducing computing times.

As we have seen, the MAP solution is a "weighted average" of the solutions to each of the two terms of Eq. (3-1). It is natural to try solving the two terms separately using Newton's method, then doing a single weighted average. This, however, is dangerous since we know the "weighted average" concept was for physical understanding and did not imply a true relation. Since $\underset{\sim}{H}$ may be singular, we may not have a unique solution for the first term but a set of solutions determined by the null space of $\underset{\sim}{H}$. Further, we are given no clue as to how to choose the weights for averaging the two solutions.

By applying Newton's method on each of the two terms and doing a weighted average at each iteration, some of the above-mentioned problems can be avoided. This permits the iteration to move successively closer to the solution while at each step controlling the ill-conditioning of the maximum likelihood solution. This method will avoid having to choose a single set of weights for the average of the two solutions and will yield a method of selecting and adjusting the weights at each iteration. Basically, we are calculating a step to solve the MAP (or mean solution) term and a step to solve the maximum likelihood term and doing a weighted average to calculate a step to solve the MAP Eq. (3-1).

Consider the first term of Eq. (3-1). We are essentially solving

$$\psi(\underset{\sim}{f}) = \underset{\sim}{g} - s(\underset{\sim\sim}{H}\underset{\sim}{f}) = 0. \tag{3-7}$$

Letting $\underset{\sim}{b}=\underset{\sim\sim}{H}\underset{\sim}{f}$ and applying the formula given by Eq. (3-6) we get

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{(k)} + (\underset{\sim\sim}{H}\underset{\sim\sim}{S}_b)^{-1} [\underset{\sim}{g} - s(\underset{\sim\sim}{H}\underset{\sim}{f}^{(k)})] , \tag{3-8}$$

where $\underset{\sim}{S}_b$ is the same as defined by Eq. (2-13) and $\underset{\sim}{J}(\underset{\sim}{f}^{(k)}) = \underset{\sim}{H}\underset{\sim}{S}_b$. Unfortunately, if $\underset{\sim}{H}$ is ill-conditioned or singular, so is $\underset{\sim}{H}\underset{\sim}{S}_b$. As mentioned above, singularity of $\underset{\sim}{H}$ results in a set of solutions to Eq. (3-7) determined by the null space. There is no way to eliminate this problem mathematically. However, in practice by starting the iteration scheme at reasonable values $\underset{\sim}{f}_0$ and by controlling the solution of Eq. (3-7) by the mean $\hat{\underset{\sim}{f}}$, there is no detectable problem.

Let us consider a way of lessening the effect of this ill-conditioned matrix. If we solve Eq. (3-7) only for $\underset{\sim}{b} = \underset{\sim}{H}\underset{\sim}{f}$, we get

$$\underset{\sim}{b}^{(k+1)} = \underset{\sim}{b}^{(k)} + \underset{\sim}{S}_b^{-1}\left[\underset{\sim}{g} - s(\underset{\sim}{b}^{(k)})\right] \quad . \tag{3-9}$$

This alone would lead to the solution $\underset{\sim}{b} = s^{-1}(\underset{\sim}{g})$; however, we want a solution for $\underset{\sim}{f}$. Multiplying this solution by $\underset{\sim}{H}^{-1}$ would put us back where we started. The alternative is to modify the iteration.

Consider the calculation of $\underset{\sim}{H}\underset{\sim}{f}^{(k)} = \underset{\sim}{b}^{(k)}$ where

$$b_i = \sum_{j=1}^{N} h_{ij} f_j^{(k)} \quad .$$

From this it is seen that any change in $f_j^{(k)}$ will result in changes to all b_i for which $h_{ij} \neq 0$. If it is determined from $\underset{\sim}{S}_b^{-1}\left[\underset{\sim}{g} - s(\underset{\sim}{b}^{(k)})\right]$ that b_i must be changed, it would be foolish to change only f_i; but the change should be distributed over all the points which affect b_i, that is, all f_j for which $h_{ij} \neq 0$. A natural weighting for this distribution is the h_{ij}'s since it is usual to consider only energy-conserving systems where

$$\sum_{j=1}^{N} h_{ij} = 1 \qquad \text{for } i = 1,\ldots,N.$$

This means that the total energy (intensity) is unchanged during the blurring phenomena. This leads to the iteration

$$\underset{\sim}{f}^{(k-1)} = \underset{\sim}{f}^{(k)} + \underset{\sim}{H}^{T} \underset{\sim}{S}_b^{-1} \left[g - s\left(\underset{\sim}{f}^{(k)} \right) \right] \quad . \tag{3-10}$$

This scheme is well-conditioned and leads to the maximum likelihood solution but in a more restrained manner because of the smoothing by the point spread function matrix $\underset{\sim}{H}^{T}$.

The Newton's method formulation for the solution of the zero for the MAP term is

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{(k)} - \left[\underset{\sim}{f}^{(k)} - \bar{\underset{\sim}{f}} \right] \quad , \tag{3-11}$$

i.e., set the k+1 iteration equal to the mean.

We noted before that the values R_n and R_f, taken as simple variances here, represent the confidence in the solutions to the maximum likelihood and MAP terms, respectively. We combine the two iterations Eqs. (3-10) and (3-11) and weight each using the natural ratios $\frac{R_f}{R_f+R_n}$ and $\frac{R_n}{R_f+R_n}$, respectively. This yields the iteration scheme

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{(k)} + \frac{R_f}{R_f+R_n} \underset{\sim}{H}^{T} \underset{\sim}{S}_b^{-1} \left[\underset{\sim}{g} - s\left(\underset{\sim}{H} \underset{\sim}{f}^{(k)} \right) \right] - \frac{R_n}{R_f+R_n} \left[\underset{\sim}{f}^{(k)} - \bar{\underset{\sim}{f}} \right] \quad . \tag{3-12}$$

In using any iteration scheme it is important to know when the scheme has converged. We saw that the convergence criterion proposed by Hunt for the original numerical method was not well-defined. Because of the relation of MAP estimation with constrained least squares estimation, we can now propose a well-defined convergence criterion. Convergence is reached when

$$\left[\underset{\sim}{g} - s(\underset{\sim\sim}{H}\underset{\sim}{f}^{(k)})\right]^T R_n^{-1} \left[\underset{\sim}{g} - s(\underset{\sim\sim}{H}\underset{\sim}{f}^{(k)})\right] \leq 1 \quad . \tag{3-13}$$

When the iteration gets within the noise level boundary, we take that iteration as the solution.

It has not been proved that this solution satisfies either the fundamental MAP equation or the constrained least squares problem. It is conjectured that by starting the iteration scheme at $\hat{\bar{f}}$ and moving away from $\hat{\bar{f}}$ in a restrained way, the solution is very close to those of the MAP and constrained least squares problems. Results from applied problems have indicated that the solution described by Eqs. (3-12) and (3-13) is quite good.

3.5 Comparison of New and Old MAP Algorithms

A previous restoration, presented in [10], is shown here for comparison of the old and new MAP solution schemes. The original unblurred image is shown in Fig. 3-a. This image was blurred using a uniform 3 x 3 point spread function matrix; the image was transformed to the density domain by a standard D log E curve and zero mean Gaussian noise with standard deviation, $\sigma = 0.02$ was added. This image is shown in Fig. 3-b. The restoration of Fig. 3-b by Hunt's algorithm after 60 iterations is shown in Fig. 3-c. The restoration of Fig. 3-b by the new algorithm which converged in six iterations is shown in Fig. 3-d. Notice the bands on the child's cap and the separation of his eyes. Notice also the overall reduction of noise.

Fig. 3-a Original unblurred image

Fig. 3-b Blurred image plus noise

Fig. 3-c MAP restoration by Hunt's method

Fig. 3-d MAP restoration by new method

3.6 Comparison of New MAP Solution With Other Restoration Methods

Since it is clear from the results shown in Fig. 3 that the new MAP algorithm produces a better restoration, it is worthwhile to compare the MAP restoration with other standard methods as was done previously by Hunt [10]. A standard test chart was blurred by defocusing the camera lens before exposing the film. The radius of the blur, as determined by Cannon's algorithm [2], was r = 4.0 pixels. The result is shown in Fig. 4-a. Figure 4-b is the Wiener restoration. Figure 4-c is restoration by power spectrum equalization (P.S.E.) [2]. Figure 4-d is the MAP restoration after 64 iterations.

The Wiener and P.S.E. restorations used the power spectrum from an unblurred picture of the same chart, hence the restorations are the best one could expect from these methods. The MAP restoration used only information from the blurred picture in Fig. 4-a. Notice the noise control and the absence of artifacts in the MAP restoration as compared to the other two methods. The Wiener and P.S.E. restoration methods have produced small disks (actually images of the point spread function) because of imperfection in the film (dust, etc.). One is easily seen below the h in the second line This is a type of artifact avoided in the MAP restoration.

3.7 Use of the Convergence Criterion

As mentioned previously, the new iteration scheme has a natural convergence criterion given by Eq. (3-13). The algorithm is stopped when the variance of $g - s(Hf) \leq R_n$, i.e., the residual, is less than or equal to the variance of the noise. We could go further with our convergence criterion and require that the residuals be uncorrelated. This would lead to a rather complex set of tests to be carried out at each iteration.

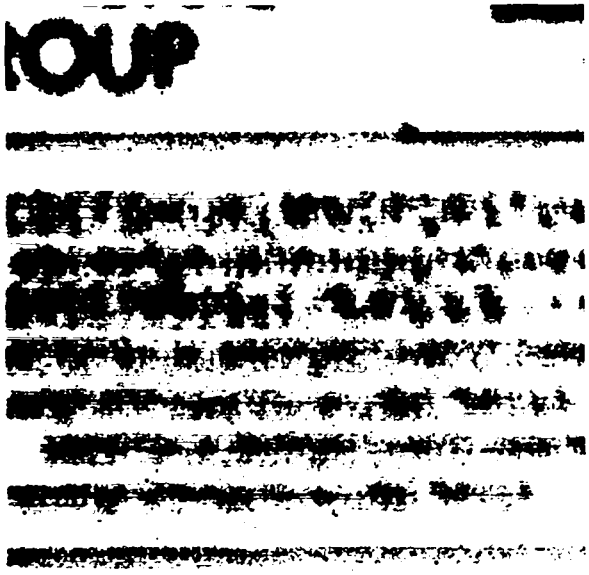

Fig. 4-a Original blurred image

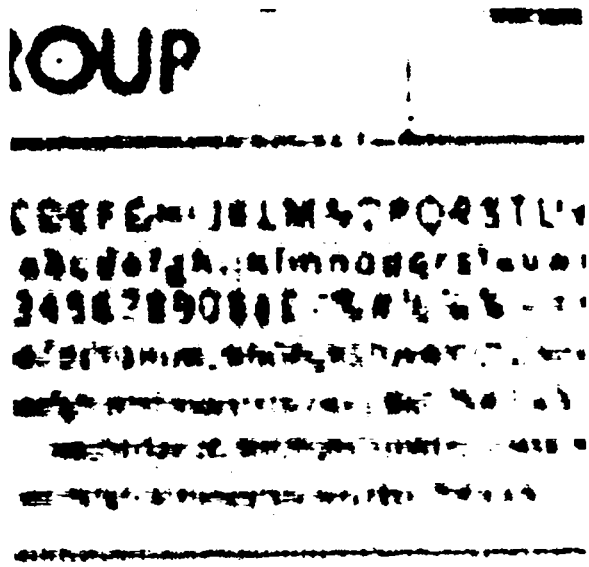

Fig. 4-b Restoration by Wiener filter

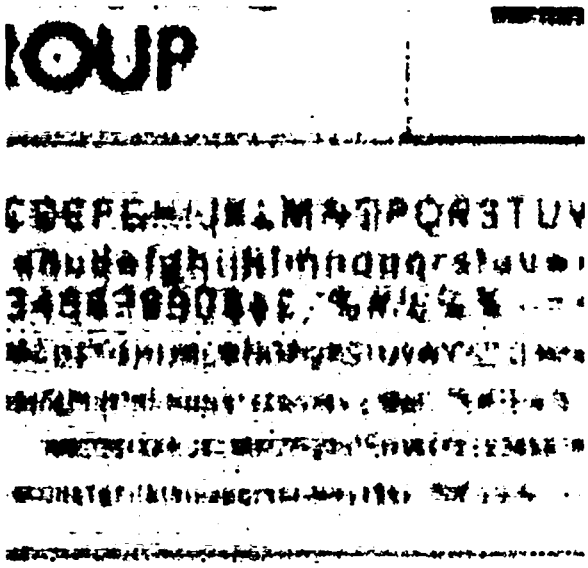

Fig. 4-c Restoration by power spectrum equalization

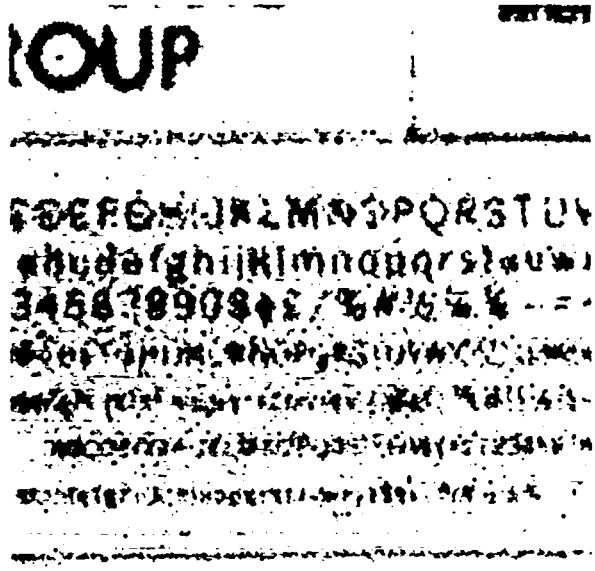

Fig. 4-d Restoration by MAP

Such tests would be costly and in the light of other results from local processing are unwarranted.

We can, however, make good use of the residual, $\underset{\sim}{g}-s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})$. Since we would like to have an uncorrelated residual image, we can easily check for correlations with the eye. If we display the residual image, we would hope to see white noise. If we see patterns in this image, we know further processing with MAP would be beneficial. We also know exactly where in the image we can make a better restoration. This allows us to extract only that part of the image that can benefit from further processing and use the algorithm on a smaller image, thus saving computation time.

The residual images associated with the MAP restorations in Fig. 3-d and Fig. 4-d are shown in Fig. 5-a and Fig. 5-b, respectively. While both restorations have converged and the variances of the residual images are less than the variances of the respective noise processes, one can easily see patterns in Fig. 5-a around the child's cap and at the edge of the mother's jacket, whereas patterns are much less discernable in Fig. 5-b.

3.8 Properties of the New Algorithm

Of the properties of the new algorithm, the most critical, besides producing a good restoration, is the sensitivity of the solution to the values of the parameters. The user of the MAP method must select values for $\underset{\sim}{\hat{\bar{f}}}$, $\underset{\sim}{f}^{(o)}$, $\underset{\sim}{\hat{R}}_f$, $\underset{\sim}{R}_n$, $\underset{\sim}{H}$, and s. There are standard methods for determing $\underset{\sim}{H}$ for certain types of blurs and for determining the noise variance $\underset{\sim}{R}_n$. The non-linear function s should be known by studies of

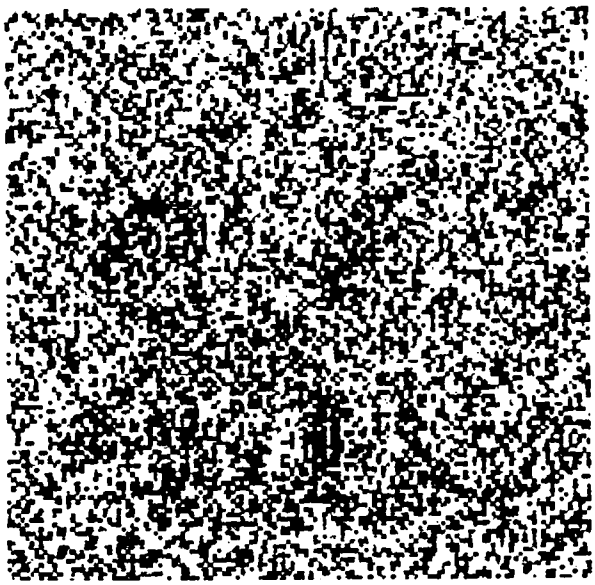

Fig. 5-a Residual of restoration in Fig. 3-d

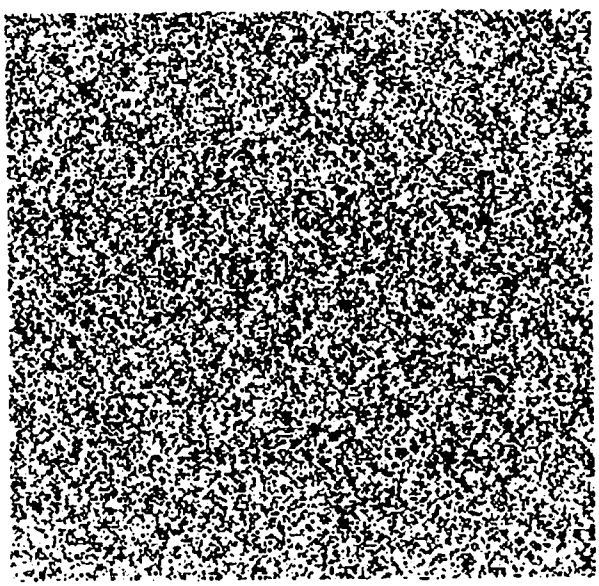

Fig. 5-b Residual of restoration in Fig. 4-d

the output medium. The parameters $\hat{\underset{\sim}{\bar{f}}}$, $\underset{\sim}{\bar{f}}^{(o)}$, and $\hat{\underset{\sim}{R}}_f$ are more arbitrary and require the user to make a decision. It is of interest to know the criticality of this decision.

3.8.1 Properties of $\hat{\underset{\sim}{\bar{f}}}$

If we consider the MAP term of Eq. (3-12) as acting as noise control, we would not expect $\hat{\underset{\sim}{\bar{f}}}$ to be particularly sensitive. To act as noise control, $\hat{\underset{\sim}{\bar{f}}}$ should act to control the large noise induced oscillations which are characteristic of $\underset{\sim}{H}^{-1}s^{-1}(\underset{\sim}{g})$. Other than these rather loose restrictions, it would appear the solution should not be sensitive to perturbations of $\hat{\bar{f}}$. This was verified by experiments where $\hat{\underset{\sim}{\bar{f}}}$'s were generated by $s^{-1}(\underset{\sim}{g})$, slight blur of $s^{-1}(\underset{\sim}{g})$, and the global average of $s^{-1}(g)$, i.e., $\frac{1}{N}\sum_{i=1}^{N} s^{-1}(g_i) = \hat{\bar{f}}_j$, $j = 1,\ldots,N$. No noticeable difference was detected in the solution and the speeds of convergence were identical. Since these three estimates for $\bar{f}$ spanned the field from good to bad and had no effect on the solution, no other estimates were tried.

3.8.2 Properties of $\hat{\underset{\sim}{R}}_f$

In the initial work on MAP, Hunt [10] used an elaborate scheme described in Chapter 2, to generate the covariance matrix $\hat{\underset{\sim}{R}}_f$. One of the first methods used to speed up the MAP computation was to eliminate the convolution required by this matrix by letting $\hat{\underset{\sim}{R}}_f$ be diagonal representing a single variance. This succeeded in reducing the computation time with no detectable effect on the solution. Later, after the new algorithm was adopted, $\hat{\underset{\sim}{R}}_f$ was made diagonal with the variance of the i^{th} element equal to $\hat{\bar{f}}_i$ as a Poisson distribution. Again, no effect was noticed on

the solution. It is noticed from examining Eq. (3-12) that the fraction $\frac{R_f}{R_n+R_f}$ is close to 1.0 if $R_f >> R_n$ which is almost always the case if s represents a photographic process. The variance of a typical pictorial snapshot is about 0.05 while the variance of typical film grain noise is about 0.0004 giving a typical signal-to-noise ratio of 21 decibels. These variances are measured in the density range of the film. However, $\underset{\sim}{R}_f$ is measured in the intensity domain. This implies a transformation from the density domain to the intensity domain by the function s^{-1}. For films, a simplified s can be defined by $s(x)=\gamma \log_{10}(x/I_o)$; hence $s^{-1}(x)=I_o 10^{x/\gamma}$, where γ is determined by the film and I_o is a base intensity level equivalent to an aperture. Clearly $\underset{\sim}{R}_f$ is proportional to I_o^2 and must be estimated according to the function s used in the model. Assuming $I_o=1$ and $\gamma=1$, then $s^{-1}(x)=10^x$; assuming further that the mean density is around 1.0, then the mean intensity is around 10.0 and $\underset{\sim}{R}_f \approx 5.0$. This value is typical of the author's experience when using such a function.

In the above discussion, $\underset{\sim}{R}_f$ was estimated from the variance of the image in the intensity domain; this image would normally be $\underset{\sim}{f}^{(o)}=s^{-1}(\underset{\sim}{g})$. Another estimate could be obtained from the variance of $\underset{\sim}{f}^{(k)}-\hat{\underset{\sim}{f}}$, k>o, for it is standard practice to set $\underset{\sim}{f}^{(o)}=\bar{\underset{\sim}{f}}$. If this is not the case, the variance of $\underset{\sim}{f}^{(o)}-\hat{\underset{\sim}{f}}$ can be used. The $\hat{\underset{\sim}{R}}_f$ estimated in this manner is in the same range as the previous estimation. This estimation method has the advantage of successively updating $\hat{\underset{\sim}{R}}_f$ each iteration and it is easily obtained in the computation of the MAP iteration.

Experiments were run using various fixed estimates of $\underset{\sim}{R}_f$. Typical results showed that the iteration scheme Eq. (3-12) failed to converge for $\underset{\sim}{R}_f < 0.001$ but converged with the same speed for all $\underset{\sim}{R}_f \geq 0.01$. No difference

was apparent to the eye in any of the converged solutions. This indicates that the main criterion for $\hat{\underset{\sim}{R}}_f$ is that it be large enough to allow the solution to move away from $\hat{\underset{\sim}{f}}$; after that condition is satisfied it need only be small enough to control gross errors, and there appears to be considerable latitude in this condition. Upon examining the iteration scheme, Eq. (3-12), it is apparent it should make little difference if $R_f \approx 1.0$ or 100.0 if $R_n \approx 0.001$. The numerical scheme introduced in Chapter 4 requires $R_f \geq 1.0$ for convergence because of an additional term.

3.8.3 Properties of $\underset{\sim}{f}^{(o)}$

In any iterative numerical method, one would assume that the starting value would be critical. Indeed, most convergence theorems in numerical analysis start with an initial estimate of the solution in a neighborhood of the true solution with certain properties. The MAP iteration method is no different. A really terrible first estimate will cause the algorithm to diverge. However, most reasonable first estimates have resulted in convergence. The speed of convergence was affected as well as the goodness of the solution. Figure 3-d shows the solution reached in six iterations with $\underset{\sim}{f}_o = s^{-1}(\underset{\sim}{g})$. Figure 6-a shows the solution reached in nine iterations with $\underset{\sim}{f}_o$ equal to the global average. The fact that both converged demonstrates the latitude of the solution. A comparison of the residual images shown in Fig. 6-b and Fig. 5-a demonstrates the varying quality of the solution. It is readily seen that the restoration in Fig. 3-d is better overall. Its residual shows fewer patterns. The restoration shown in Fig. 6-a gives better definition of the stripes on the child's cap but much poorer restoration of the mother's hair.

Fig. 6-a MAP restoration using $\underset{\sim}{f}^{(o)}$ equal to global average

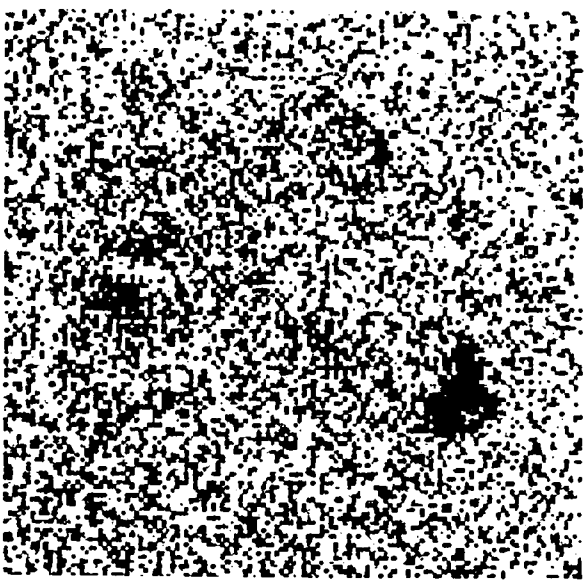

Fig. 6-b Residual of restoration in Fig. 6-a

3.9 Rate of Convergence

As with any numerical iteration scheme, there are many factors which influence the rate of convergence including, of course, the data itself. We have mentioned how the parameters $\hat{\underset{\sim}{f}}$, $\underset{\sim}{f}_0$, and $\hat{\underset{\sim}{R}}_f$ affect this rate of convergence. It is now appropriate to ask what are the effects of those parameters over which we have no control, $\underset{\sim}{H}$ and $\underset{\sim}{R}_n$. These parameters are fixed when we are given the data image $\underset{\sim}{g}$.

It is clear that as the variance of the noise increases the degradation of $\underset{\sim}{g}$ increases also. This increase of noise power makes restoration by any method more difficult and in general reduces the restorability. Thus, the worse the noise, the worse the restoration. In terms of the heuristic explanation presented earlier in this chapter, the larger noise variance results in a larger solution space. It is reasonable to guess then, that increases in the noise variance actually increases the rate of convergence. This guess has been verified by experimentation. For example, on a standard sequence of computer runs the noise variance was set at 0.0004, 0.0016, 0.0064, and 0.0256, with all other variables held constant; the numbers of iterations required for convergence were 21, 9, 4, and 2, respectively.

It is also clear that as the blur gets worse (usually measured by its second moment) the degradation of $\underset{\sim}{g}$ increases. Since the size of the solution space is determined mainly by the noise variance, the size of the point spread function has little effect on the size of the solution space. Increasing the size of the point spread function means that each point is affected by more neighboring points and that each point affects more of its neighbors. Our numerical iteration algorithm averages

the needed change of a point over those points which affect that point. Increased size in the point spread function means more points over which to average the needed change. When all other changes from other points in the image are then added to the change required by the single point, the sum is a much worse estimate of the needed change. Thus, it would seem that larger point spread functions would converge more slowly than smaller ones. This has been verified by experimentation. A sequence of computer runs was done with the image blurred by square apertures, i.e., equal weights, of sizes 3 x 3, 5 x 5, and 7 x 7 with all other variables held constant. The numbers of iterations required for convergence were 6, 21, and 43 respectively. This point will again be examined in Chapter 4 in the light of results from numerical analysis.

3.10 Timing

In computing a single iteration of the MAP algorithm given by Eq. (3-12), it is obvious that the most expensive operations are the convolutions required by the matrix multiplication of $\underset{\sim}{H}$ and $\underset{\sim}{H}^T$. The Toeplitz forms discussed in Chapter 2 allowed this computation to be speeded up by the use of the FFT, but it is still the limiting computation factor. Other computations, as multiplications for the weights and $\underset{\sim}{S}_b^{-1}$, and computations of $\underset{\sim}{S}_b^{-1}$, also add to this basic time. There is also significant time required for input/output of the several large arrays which must be stored on disk.

The time of an iteration is almost singly a function of the size of the required convolution. These convolutions were done in a size to avoid as much as possible the periodic effects of the FFT. If the image is

M x M and the point spread function is N x N, the FFT multiplication is carried out on matrices 2^k x 2^k where $2^k \geq M+N-1 > 2^{k-1}$. The timing values given in Table 1 are functions of the matrix size required for the convolution. The times represent computation done on a CDC-7600 computer.

TABLE 1
Timing for MAP Restoration

Matrix Size	Time per Map Iteration
128	8.1 seconds
256	12.8 seconds
512	88.5 seconds

CHAPTER 4

Numerical Analysis

4.1 Introduction

We have seen that the iteration scheme described in Chapter 3 works and produces good results. However reasonable the heuristic behind the scheme, it lacks a satisfying mathematical foundation. This chapter will attempt to give that necessary foundation.

4.2 Modified Picard's Method

The following method is described in Isaacson and Keller [12]. Picard's method of solving non-linear equations is derived by writing the equations in the form

$$\underset{\sim}{x} = \phi(\underset{\sim}{x}) \quad , \tag{4-1}$$

then using the iteration

$$\underset{\sim}{x}^{(k+1)} = \phi(\underset{\sim}{x}^{(k)}) \quad . \tag{4-2}$$

Unfortunately, this scheme is rarely used in actual practice because of the limitations discussed in Chapter 2. However, this scheme can be modified to yield a scheme with an improved convergence rate, as well as one which will converge when the original diverges. The new iteration is given by

$$\underset{\sim}{x}^{(k+1)} = \underset{\sim}{\theta}\phi(\underset{\sim}{x}^{(k)}) + (\underset{\sim}{I}-\underset{\sim}{\theta})\ \underset{\sim}{x}^{(k)} \quad , \tag{4-3}$$

where $\underset{\sim}{\theta}$ is a diagonal matrix given by

$$\underset{\sim}{\theta} = \begin{bmatrix} \theta_1 & & & \\ & \theta_2 & & \\ & & \ddots & \\ & & & \theta_N \end{bmatrix}, \quad \det[\underset{\sim}{\theta}] = \theta_1\theta_2\ldots\theta_N \neq 0 \quad ,$$

Clearly, if $\underset{\sim}{\theta} = \underset{\sim}{I}$, the identity matrix, the resulting scheme is the unmodified Picard's Eq. (4-2). The scalar form of Eq. (4-3) is

$$x_i^{(k+1)} = \theta_i \phi(x_i^{(k)}) + (1 - \theta_i)\, x_i^{(k)} \quad , \; i=1,2,\ldots N, \tag{4-4}$$

so computation is easily implemented.

It is stated in [12] that the new method will converge if in a neighborhood of the solution the vector function ϕ has continuous first partial derivatives, $\phi_{ij} = \dfrac{\partial \phi_i\ (\underset{\sim}{x})}{\partial x_j}$, which satisfy the conditions

$$\left| 1 - \phi_{ii}(\underset{\sim}{x}) \right| > \sum_{j \neq i} \left| \phi_{ij}(\underset{\sim}{x}) \right| \qquad i = 1,2,\ldots,N \tag{4-5}$$

for some choice of $\underset{\sim}{\theta}$. The optimum $\underset{\sim}{\theta}$ is derived as follows:

Consider the error at the k^{th} iteration

$$\underset{\sim}{e}^{(k)} = \underset{\sim}{x}^{(k)} - \underset{\sim}{\alpha} \quad , \tag{4-6}$$

where $\underset{\sim}{\alpha}$ is the solution

$$\underset{\sim}{\alpha} = \underset{\sim}{\phi}(\underset{\sim}{\alpha}) \quad .$$

Substituting Eq. (4-3) into Eq. (4-6) we get

$$\underset{\sim}{e}^{(k+1)} = \underset{\sim}{\theta}\, \underset{\sim}{\phi}(\underset{\sim}{x}^{(k)}) + (\underset{\sim}{I} - \underset{\sim}{\theta})\, \underset{\sim}{x}^{(k)} - \underset{\sim}{\alpha} \quad .$$

Rewriting $\underset{\sim}{\alpha}$ as $\underset{\sim}{\alpha} = \underset{\sim}{\theta}\,\underset{\sim}{\phi}(\underset{\sim}{\alpha}) + (\underset{\sim}{I}-\underset{\sim}{\theta})\underset{\sim}{\alpha}$ we have

$$\underset{\sim}{e}^{(k+1)} = \underset{\sim}{\theta}\left[\underset{\sim}{\phi}(\underset{\sim}{x}^{(k)}) - \underset{\sim}{\phi}(\underset{\sim}{\alpha})\right] + (\underset{\sim}{I}-\underset{\sim}{\theta})\,\underset{\sim}{e}^{(k)} \quad . \tag{4-7}$$

Using Taylor's theorem for vector valued functions we can write

$$\underset{\sim}{\phi}(\underset{\sim}{x}^{(k)}) - \underset{\sim}{\phi}(\underset{\sim}{\alpha}) = \underset{\sim}{J}_{\phi}(\underset{\sim}{\xi})\,(\underset{\sim}{x}^{(k)} - \underset{\sim}{\alpha}) \quad , \tag{4-8}$$

where $\underset{\sim}{\xi}$ is some point in the neighborhood satisfying Eq. (4-5) and $\underset{\sim}{J}_{\phi}(\underset{\sim}{\xi})$ is the matrix defined by

$$\underset{\sim}{J}_{\phi}(\underset{\sim}{\xi}) = \left(\left.\frac{\partial\,\phi_i(\underset{\sim}{x})}{\partial x_j}\right|_{\underset{\sim}{x}\,=\,\underset{\sim}{\xi}}\right) \quad .$$

Substituting Eq. (4-8) into Eq. (4-7) we have

$$\begin{aligned}\underset{\sim}{e}^{(k+1)} &= \underset{\sim}{\theta}\,\underset{\sim}{J}_{\phi}(\underset{\sim}{\xi})\,\underset{\sim}{e}^{(k)} + (\underset{\sim}{I} - \underset{\sim}{\theta})\,\underset{\sim}{e}^{(k)} \\ &= \left[\underset{\sim}{I} - \underset{\sim}{\theta} + \underset{\sim}{\theta}\,\underset{\sim}{J}_{\phi}(\underset{\sim}{\xi})\right]\underset{\sim}{e}^{(k)} = \underset{\sim}{M}_k\,\underset{\sim}{e}^{(k)} \quad . \end{aligned} \tag{4-9}$$

The method defined by Eq. (4-3) will converge if

$$\|\underset{\sim}{M}_K\| \leq q < 1 \qquad \text{for all } k \quad .$$

If we use the maximum norm

$$\|\underset{\sim}{x}\| = \underset{1\leq i\leq N}{\text{maximum}} \left\{\,|x_i|\,\right\}$$

and the induced matrix norm,

$$\| \underset{\sim}{A} \| = \underset{1 \leq i \leq N}{\text{maximum}} \left\{ \sum_{j=1}^{N} | a_{ij} | \right\} \quad ,$$

convergence is guaranteed if

$$|1 - \theta_i[1 - \phi_{ii}(\underset{\sim}{\xi})]| + |\theta_i| \sum_{j \neq i} | \phi_{ij}(\underset{\sim}{\xi})| \leq q < 1 \qquad 1 \leq i \leq N \ . \quad (4\text{-}10)$$

This condition holds because of the assumption of Eq. (4-5). Consider inequalities of the form

$$R(\theta) = |1 - \theta\alpha| + |\theta| \, b < 1 \quad , \quad b > 0 \quad .$$

If $|a| > b$ and θ is in the interval

$$0 < \theta < \frac{2}{a+b} \qquad \text{if } a>b \qquad \text{(case 1)}$$

or $$\frac{2}{a-b} < \theta < 0 \qquad \text{if } -a>b \qquad \text{(case 2)}$$

then $R(\theta) < 1$. For case 1

$$0 < \theta < \frac{2}{a+b} \qquad \text{implies}$$

$$0 < \theta a - 1 + \theta b < 1 \quad . \qquad (4\text{-}11)$$

Now consider $\theta a - 1$; if $\theta a - 1 \geq 0$ we have $|1-\theta a| = \theta a - 1$ and we are done. If $\theta a - 1 < 0$ we must have $0 < \theta a - \theta b < 1$ since $\theta b > 0$ and $a > b$. Multiplying by -1 and adding 1 we have $0 < 1 - \theta a + \theta b < 1$. Since $1 - \theta a = |1 - \theta a|$, we have proved $R(\theta) < 1$ for case 1. Case 2 follows in the same manner.

Further, since $|a|>b$ the minimum value of R is attained at $\theta^* = \frac{1}{a}$

and $$R(\theta^*) = \frac{b}{a} \quad .$$

It is also clear that $R(\theta^*-\varepsilon) < R(\theta^*+\varepsilon)$ which implies it is better to underestimate than overestimate θ.

Although we know R attains its minimum at $\theta_i = \frac{1}{|\ 1 - \phi_{ii}(\underset{\sim}{\xi})\ |}$, we do not know $\underset{\sim}{\xi}$. It is suggested by [12] that

$$\theta_i^k = \frac{1}{|\ 1 - \phi_{ii}(\underset{\sim}{x}^{(k)})|} \tag{4-11}$$

can be used, where $\underset{\sim}{x}^{(k)}$ is the solution estimate at the k^{th} iteration.

4.3 Application to the MAP Equation

Rewriting Eq. (2-17) we get the form needed in the above method.

$$\underset{\sim}{f} = \underset{\sim}{\bar{f}} + \underset{\sim}{R}_f\ \underset{\sim}{H}^T\ \underset{\sim}{S}_b \underset{\sim}{R}_n^{-1} \left[\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f})\right] = \phi(\underset{\sim}{f}). \tag{4-12}$$

The modified Picard's method yields the iteration

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^k + \underset{\sim}{\theta}\ \underset{\sim}{R}_f\ \underset{\sim}{H}^T\ \underset{\sim}{S}_b\ \underset{\sim}{R}_n^{-1} \left[\underset{\sim}{g} - s(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{f}^{(k)})\right] - \underset{\sim}{\theta}\left[\underset{\sim}{f}^{(k)} - \underset{\sim}{\bar{f}}\right]. \tag{4-13}$$

To make Eq. (4-13) compatible with Eq. (3-12), we set

$$\underset{\sim}{\theta} = \underset{\sim}{R}_n\ (\underset{\sim}{R}_n + \underset{\sim}{R}_f)^{-1}\ \underset{\sim}{S}_b^{-2} \quad , \tag{4-14}$$

and we have for $\underset{\sim}{R}_n$ and $\underset{\sim}{R}_f$ scalars

$$\underset{\sim}{f}^{(k+1)} = \underset{\sim}{f}^{k} + \frac{R_f}{R_n + R_f} \underset{\sim}{H}^T \underset{\sim}{S}_b^{-1} \left[\underset{\sim}{g} - s(\underset{\sim}{H}\underset{\sim}{f}^{(k)}) \right] - \frac{R_n S_b^{-2}}{R_n + R_f} \left[\underset{\sim}{f}^{(k)} - \underset{\sim}{\bar{f}} \right] . \qquad (4\text{-}15)$$

The only difference between Eq. (4-14) and Eq. (3-12) is the S_b^{-2} in the second term of the right hand side. We shall examine this discrepancy later.

In essence, we have heuristically derived a θ, given by Eq. (4-14). Let us now see how this compares with the optimum θ of Isaacson and Keller. From Eq. (4-12) we have

$$\phi_i(\underset{\sim}{f}) = \bar{f}_i + \frac{R_f}{R_n} \sum_{k=1}^{N} h_{ki} s'(b_k) g_k - \frac{R_f}{R_n} \sum_{k=1}^{N} h_{ki} s'(b_k) s(b_k) \quad , \qquad (4\text{-}16)$$

where

$$b_k = \sum_{\ell=1}^{N} h_{k\ell} f_\ell \quad , \qquad (4\text{-}17)$$

and the ' denotes differentiation.

$$\frac{\partial \phi_i(x)}{\partial f_j} = \phi_{ij} = \frac{R_f}{R_n} \sum_{k=1}^{N} h_{ki} s''(b_k) \frac{\partial b_k}{\partial f_j} g_k$$

$$- \frac{R_f}{R_n} \sum_{K=1}^{N} h_{ki} \left[s''(b_k) + s'(b_k)^2 \right] \frac{\partial b_k}{\partial f_j} \quad .$$

Noting that $\frac{\partial b_k}{\partial f_j} = h_{kj}$ and rearranging the summations we have

$$\phi_{ij} = \frac{R_f}{R_n} \sum_{k=1}^{N} h_{ki} h_{kj} s''(b_k) \left[g_k - s(b_k) \right] - \sum_{k=1}^{N} h_{ki} h_{kj} s'(b_k)^2 \qquad (4\text{-}18)$$

and

$$\phi_{ii} = \frac{R_f}{R_n} \sum_{k=1}^{N} h_{ki}^2 \, s''(b_k)[g_k - s(b_k)] - \sum_{k=1}^{N} h_{ki}^2 \, s'(b_k)^2 \, . \qquad (4\text{-}19)$$

While this quantity is computable, it would be quite costly; and it fails to shed much light on why our heuristic θ worked so well.

Let us now make some simplifying assumptions. If we assume the point spread function to be small compared to the size of the image, then $h_{ij} \neq 0$ only over a small region. Since $\underset{\sim}{H}$ is a blur, it is also reasonable to assume that the image $\underset{\sim}{b}$ is smooth, and we may approximate $\underset{\sim}{b}$ by a constant over the small region over which $h_{ij} \neq 0$. By replacing $s'(b_k)$ and $s''(b_k)$ by single values $s'(b_o)$ and $s''(b_o)$, respectively, and $g_k - s(b_k)$ by $g_o - s(b_o)$ we have

$$\phi_{ii} \approx \frac{R_f}{R_n} \left[s''(b_o) \, [g_o - s(b_o)] - s'(b_o)^2 \right] \sum_{k=1}^{N} h_{ki}^2 \quad . \qquad (4\text{-}20)$$

If $|g_o - s(b_o)| << 1$ and $s''(b_o)| \approx | s'(b_o)^2|$ or if $| s''(b_o)| << |s'(b_o)^2|$ and $|g_o - s(b_o)| \approx s'(b_o)^2$ we can eliminate the first term in the first brackets. To show that this is reasonable, let s represent an H and D curve for film, i.e.,

$$s(x) = \gamma \log(x) \quad .$$

Then we have

$$s'(x) = \frac{\gamma}{x} \; ;$$

$$s''(x) = -\frac{\gamma}{x^2}$$

with typical γ's ranging from 0.8 to 3.0. The term $g - s(b_o)$ is in the density domain. A typical density range for film is 0.3 to 2.0; hence it is not unreasonable to expect the difference $| g_o - s(b_o)| < 0.1$. The quantity

$g_o - s(b_o)$ is related to the noise which in most cases is small. Typical values of the standard deviation of film grain noise are $\sigma \approx 0.02 - 0.005$. Using this assumption and letting

$$c = \sum_{k=1}^{N} h_{ki}^2$$

we have

$$\phi_{ii} \approx - \frac{R_f}{R_n} C \, s'(b_o)^2 \quad . \tag{4-21}$$

Substituting Eq. (4-21) into Eq. (4-11)

$$\theta_i \approx \frac{1}{1 + \frac{R_f}{R_n} C \, s'(b_o)^2} = \frac{R_n}{R_n + R_f \, C \, s'(b_o)^2} \quad . \tag{4-22}$$

We mentioned in Chapter 3 that R_f was a parameter to be determined, thus including C in this estimate is no problem. In the case of film, R_n is the noise variance in the density domain while R_f is a signal variance in the intensity domain and we have naturally $R_n << R_f$. This last assumption gives

$$\theta_i \approx \frac{R_n}{R_f \, S'(b_o)^2} \approx \frac{R_n \, s'(b_o)^{-2}}{(R_n + R_f)} \quad . \tag{4-23}$$

It was noticed earlier that the heuristic "θ" lacked the $s'(b_o)^{-2}$ multiplying $(\underset{\sim}{f}^{(k)} - \underset{\sim}{\bar{f}})$. One reason this did not hurt our solution was that this term acts as a noise restraint; thus we are letting the solution be less restrained. A second reason leaving the $s(b_o)^{-2}$ off

did little is seen by examining the magnitudes of the terms. If $s(x) = \log(x)$ then

$$s'(x)^{-2} = x^2.$$

If our image has densities around 1.0, $x \approx 10.0$ and $x^2 \approx 100.0$. Experiments with standard film have shown $R_n \approx 0.0002$. Typical values for R_f are about 100.0. Thus, the θ of Eq. (4-22) gives

$$\theta \approx 10^{-4}$$

where in the heuristic case

$$\theta \approx 10^{-6} \quad .$$

In either case, experiments have shown little is lost by completely omitting the second term of the iteration. It does, however, guard against singular behavior or gross errors of the type permitted by the maximum likelihood estimate of noisy data.

4.4 Effects on Rate of Convergence

It is shown in Chapter 3 that larger point spread functions resulted in slower convergence rates. A reason mentioned then was that more averaging of the needed changes was being done. A second reason can be inferred from our calculation in this chapter for the optimum θ_i. As the point spread function gets larger, more h_{ij}'s are non-zero. This means that

$$\sum_{k=1}^{N} h_{ki}^2 \left\{ s''(b_k) \left[g_k - s(b_k) \right] - s'(b_k)^2 \right\}$$

can no longer be approximated very well by

$$\sum_{K=1}^{N} h_{ki}^2 \left(s''(b_o)[g_o - s(b_o)] - s'(b_o)^2 \right) \cdot$$

The estimate for θ_i given by Eq. (4-14) is no longer close to the optimal θ.

4.5 Modification of Heuristic Method

As a result of the numerical studies presented in this chapter, altering the heuristic method was found to be beneficial. It was noted that the methods differed not only in the multiplication of the MAP term but also in the order of the matrix multiplications in the first term. Tests showed that the order of multiplication of the modified Picard's method gave faster convergence (6 iterations versus 8 iterations for the picture in Fig. 3-b) than the heuristic method. This order of multiplication was used in all further restorations.

In one-dimensional simulations using the MAP restoration method the above findings were confirmed. It was also discovered that the rate of convergence was increased by using the approximation of the optimal θ given in Eq. (4-20). The summation in this equation is much easier to compute in the one dimensional case.

CHAPTER 5

Local Processing and MAP Restoration

5.1 Introduction to Local Processing

In the analysis of an image, prior to processing, one takes particular note of the varying characteristics of different regions of the entire image. If a power spectrum is to be calculated, for example, one would note the frequency characteristics of the different regions and expect to see these characteristics in the spectrum of the image. Of course, these characteristics would be diluted by averaging over the entire image. In any type of processing, there are certain properties which may define a region, even though such properties themselves may not be well defined, i.e., not mathematically defined. The mere fact that regions can be defined tells us that the defining properties may be lost when considering the image as a whole, that is, computing the defining quantities using the entire image.

Consider a simple picture of a building, trees, grass, and sky. Assume we are interested in generating a Wiener filter. This requires an estimate of the power spectrum. If we compute an estimate by using the entire image or by estimating the spectrum for smaller sections and then averaging, we will have diluted the characteristics of each area. We would expect, for example, the spectrum of the building to have most of its power concentrated in orthogonal directions; it might even have a periodic nature depending upon its structure. The sky is basically flat, and thus would have most of its power in the lower frequencies.

The grass and trees would have more power in the higher frequencies than the sky and none of the directional properties of the building. Using an averaged power spectrum might give an optimal filter for the picture taken as a whole, but such a filter would be suboptimal for each of the different areas.

It would be an advantage to consider the local properties of the image when processing the image. Local or adaptive techniques have been proposed in coding theory and bandwidth compression [4], [5]. It is natural, then, to investigate whether similar techniques could be applied to image processing in general and to the MAP problem in particular.

Local processing might be defined as any processing which attempts to use the characteristics or properties of a smaller part of an image to process that particular part. Global processing would be defined as processing that uses properties of the entire image to process the entire image. Several processing techniques have been implemented where only a part of the image was being processed at a time, usually because of computing time or space limitations. These techniques could not be classified as local processing unless the processing of each part of the image changed in some way to reflect the peculiar properties of that part.

The advantages of local processing are fairly obvious. The properties of which we wish to take advantage have not been averaged out over the entire image. Local techniques lend themselves to piecemeal computations. Local techniques can be changed drastically from one part of the image to the next.

The disadvantages of local processing are not absent. There may be extra cost in computing time because local processing requires some redundancy; for example, overlapping sections to minimize section

boundary effects. More care must be taken in noisy environments because noise cannot be averaged out as in global processing. With less data the statistics on a random signal will be poorer.

A standard technique which lends itself to local processing is density stretching. Here image densities are linearly transformed to cover the range of the display medium. Since the transformation is a function of the maximum and minimum densities, greater stretching can be obtained over smaller sections, giving greater contrast. In applying local techniques to density stretching, one must consider the disadvantages mentioned before.

Section boundaries are a problem in local processing since adjacent sections are being transformed by different linear functions. Points near the boundaries, even though originally very close in value, may be mapped onto values quite far apart depending upon the other values in their respective sections. Overlapping the sections is one way to minimize boundary effects. Overlapping is done by using the statistics of a larger section to compute the stretching parameters but actually doing the transformation on a smaller subsection. The relative sizes of the sections are dependent upon the nature of the picture and the desired effect.

Noise affects local stretching in a unique way. Where the variation of the signal is greater than the variation of the noise, the method yields the desired contrast enhancement. However, in sections where the variation of the noise is greater than that of the signal, for example, in flat regions of the picture, the noise will be expanded to the full range of the recording medium. This is not necessarily bad, for it tells one where the noise dominates the signal; but one must certainly take care in reading such a processed image.

5.2 Local Processing Applied to MAP Restoration

While local processing appears to be advantageous for almost any processing which is done globally, MAP estimation has properties which make it particularly adaptable to local processing. The computation of the iteration scheme can easily be carried out on small sections. The convergence criterion defined by Eq. (3-13) suggests that better restorations can be achieved using local methods.

It is noted that the convergence criterion

$$e(\underset{\sim}{f}^{(k)}) = \| \underset{\sim}{g} - s(\underset{\sim\sim}{H}\underset{\sim}{f}^{(k)})\|^2 \leq R_n \qquad (5\text{-}1)$$

is an average value. In the global case, this average is taken over the entire image, i.e.,

$$\| \underset{\sim}{g} - s(\underset{\sim\sim}{H}\underset{\sim}{f}^{(k)}) \|^2 = \frac{1}{N} \sum_{j=1}^{N} \left(g_j - s(b_j)\right)^2 ,$$

where $\underset{\sim}{b} = \underset{\sim\sim}{H}\underset{\sim}{f}^{(k)}$. This means that over a large area it is possible to average out very large values with very small values. If the area over which this average is taken is small, then there is less chance of large errors being offset by small ones and, thus, a more homogeneous error pattern is obtained. This implies that the smaller area or section results in better restoration. Experiments were made using varying section sizes with the result that the smaller section size produced a more pleasing reconstruction; and the residual image, defined in Chapter 3, showed significant improvement. It is only computation time limitations which will prevent us from practically applying local MAP estimation to each individual point.

It would seem that one would be more concerned with large errors in the convergence function rather than small ones. If R_n is set equal to zero, the solution would be the maximum likelihood estimate. This solution was unsatisfactory because of the domination of the restoration by the noise. If $e(\underset{\sim}{f}^{(k)})$ is too small in the presence of noise, then the solution is likely to be excessively noisy. By stopping the iteration as soon as Eq. (5-1) is satisfied, we avoid this noise problem. Local processing, by producing a more homogeneous error, prevents this noise build-up. This effect has been noted in experiments.

Certain assumptions were made in Chapter 3 to facilitate computation. One such assumption was the Toeplitz form of the matrix $\underset{\sim}{H}$, (the matrices $\underset{\sim}{R}_n$ and $\underset{\sim}{R}_f$ were also Toeplitz but were later reduced to scalars). This form permitted the use of the fast Fourier transform (FFT) to speed the computation. The advantage of the FFT is diminished when the size of the matrices is decreased. Local processing requires smaller sections and permits direct computation of the convolution. For small point spread functions, this results in faster computation times. The number of multiplications required for the convolution using the FFT for images of size N x N is $N^2[16(3\log_2(N) + 2)]$. The number of multiplications required by direct convolution for the N x N image using $N_{SEC} \times N_{SEC}$ sections and an $N_{PSF} \times N_{PSF}$ point spread function is

$$N^2\left[\frac{N^2_{PSF}\left(N_{SEC} + 2(N_{PSF} - 1)\right)^2}{N^2_{SEC}}\right] .$$

Since direct computation can be used, the Toeplitz restriction can be dropped. Since the FFT is not used, the Fourier wrap-around effect is avoided.

Local processing may save computation in another way. Since all sections of the image do not have the same content, some sections are more or less degraded than others. For example, a blurred sky is hardly noticeable whereas a blurred picket fence is quite objectionable. The convergence rate of the MAP iteration scheme is clearly affected by the content of the image. Local processing gives the scheme the opportunity to spend its time where it is most needed. It will converge rapidly on "flat" sections; it can spend many iterations on sections with more detail.

5.3 Computation of Local MAP Estimates

The quantities which are to be calculated are, of course, the same for the local and global cases. The difference is in the methods of computation. The global method relied heavily on the FFT. The local method uses direct matrix convolution in the space domain. The local method must also handle the computations in a manner designed to minimize boundary effects.

The only cause of boundary discontinuities in the MAP computation is the edge effect of the convolutions. This effect can be avoided by overlapping the sections by the area affected. Since Eq. (3-12) contains two convolutions and assuming the point spread function matrix is m x m, the amount of overlap required is 2(m-1) x 2(m-1). If the Section (we use the capital S for clarity) over which we wish to compute the MAP estimate is n x n, then we must use a working section of [n + 2(m-1)] x [n + 2(m-1)]. The necessary overlap area constitutes the major overhead in the local processing. It is clear that smaller Sections have a larger percentage of overhead computation. It is also clear that the computation

will be more inefficient for larger point spread functions.

Local processing requires less storage than global processing since it is operating on small areas. To compute the iteration given by Eq. (3-12) one matrix, m x m, is required for the point spread function; and one matrix, [n + 2(m-1)] x (n + 2(m-1)], is required for the image and results of convolution. The computation may be speeded by having additional core storage available, but it is possible to store the intermediate results on auxiliary devices.

5.4 Results of Local Processing

It is clear that local processing is a matter of degree. A large image may be sectioned into several smaller ones and processed using the same computer code that could be used to process single large images. When Section sizes get very small, it is impractical to do this and a single code which operates on a single image in a sectioned manner is used. The following results were generated using such a code. Figure 7-a is the restoration of the degraded image shown in Fig. 3-b, using 32 x 32 Sections. Figure 7-b is the residual image, g - s(Hf) associated with this restoration. Figure 8-a is the restoration of the same degraded image using 8 x 8 Sections, and Fig. 8-b is the associated residual image.

Comparison of the residuals shown in Figs. 7-b and 8-b and Fig. 5-a shows a definite benefit from processing with 8 x 8 Sections, while processing with 32 x 32 Sections shows little difference from the global restoration. It is interesting to note here that the mean square error between the three restorations and the known solution (Fig. 3-a) were all within 10^{-4} of each other in the density domain. This demonstrates

Fig. 7-a MAP restoration using 32 x 32 Sections

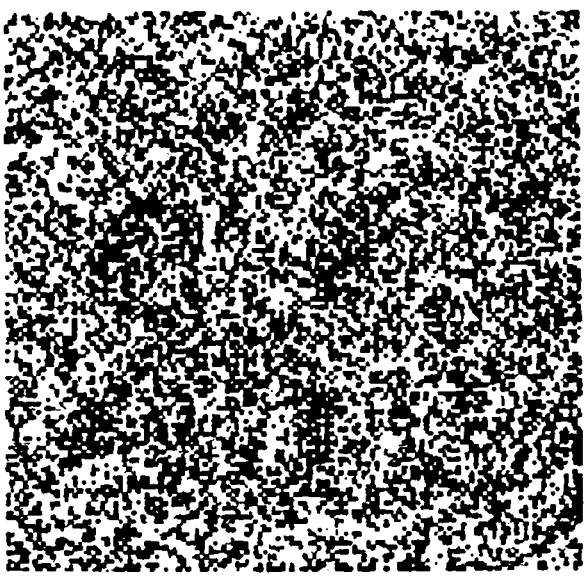

Fig. 7-b Residual of restoration in Fig. 7-a

Fig. 8-a MAP restoration using 8 x 8 Sections

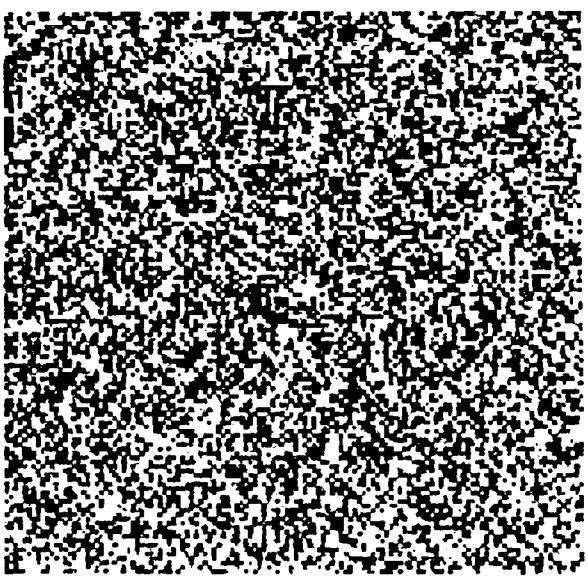

Fig. 8-b Residual of restoration in Fig. 8-a

again what is common knowledge in the image processing field, namely that mean square error is not a good criterion for judging quality.

5.5 Properties of Local Processing

From the results shown in Figs. 7 and 8, it is clear that smaller Section size results in better restorations. The question of how much better quality restoration does the smaller Section give is beyond the scope of this dissertation. Indeed, the quality of restorations is highly subjective. We can say the restoration shown in Fig. 7-a is better than that shown in Fig. 3-d only because the majority who have looked at them agree that it is. Residual image, g - s(Hf), gives a somewhat quantitative measure of goodness of restoration. We could actually measure the correlation of the residual image and use this as a quantitative measure, but there is little reason to assume this would yield much better results than mean square error. The residual image is proposed only as a qualitative guide.

While the quality of the restorations for various Section sizes is subjective, the cost of the restorations for various Section sizes is definitely quantitative. It will obviously take longer to compute a single MAP iteration for a larger Section, but there are fewer Sections and the overhead (i.e., proportion of overlap necessary to avoid edge effects) is smaller. It is also the true case that the convergence criterion, being an average, will allow larger Sections to converge faster for the same image data. It is not surprising to find that the restoration using 8 x 8 Sections took 128 seconds while the restoration using 32 x 32 Sections took 54 seconds. As noted in Chapter 3, the major time-consuming operation in

the MAP computation is the convolution. The time required for the convolution computation is a function of two variables, the Section size and the point spread function size. There are certain other computations which must be done; but as the Section size gets larger, these computations comprise a smaller percentage of the total time, and the time per iteration is nearly proportional to the number of multiplications in the convolution calculation. For larger Sections, it would be advantageous to again make the Toeplitz assumptions used in the global case and speed the computation by use of the FFT. A rough estimate of this crossover point could be obtained by plotting the number of multiplications required for each method as given in Section 5.2.

In pointing out that the cost of the convolution computation is a function of Section size and point spread function size, it is necessary to remind the reader of a distinction made earlier in this chapter. The Section we define as those points which are stored at the end of the computation; the working section includes these points together with the points over which computation must be done to avoid edge effects in the convolution. Thus, if we wish to use an 8 x 8 Section with a 7 x 7 point spread function, the computation requires a 20 x 20 working section. The convolution over such a section requires 19,600 multiplications (19,600 = 20·20·7·7). For larger point spread functions, the time per iteration is nearly proportional to the number of multiplications required for the convolution. Smaller point spread functions require less overhead and hence the auxiliary computations make up a larger proportion of the time. Table 2 shows the relationship between point spread function size and computation time for a given Section size. Table 3 is a rearrangement of the same data showing the relationship between Section size and

computation time for a given point spread function size.

TABLE 2

Relation of Point Spread Function Size and Computation Size

Section Size	Point Spread Function Size	Number of Multiplications Per Convolution	Time Per Iteration (Seconds)
32 x 32	3 x 3	11,664	.72
32 x 32	5 x 5	40,000	.91
32 x 32	7 x 7	94,864	1.25
32 x 32	9 x 9	186,624	1.71
32 x 32	11 x 11	327,184	3.09
16 x 16	3 x 3	3,600	.25
16 x 16	5 x 5	14,400	.33
16 x 16	7 x 7	38,416	.53
16 x 16	9 x 9	82,944	.88
16 x 16	11 x 11	156,816	1.70
8 x 8	3 x 3	1,296	.13
8 x 8	5 x 5	6,400	.15
8 x 8	7 x 7	19,600	.27
8 x 8	9 x 9	46,656	.47
8 x 8	11 x 11	94,864	1.00

TABLE 3

Relation of Section Size and Computation Time

Section Size	Point Spread Function Size	Number of Multiplications Per Convolution	Time Per Iteration (Seconds)
32 x 32	3 x 3	11,664	.72
16 x 16	3 x 3	3,600	.25
8 x 8	3 x 3	1,296	.13
32 x 32	5 x 5	40,000	.91
16 x 16	5 x 5	14,400	.33
8 x 8	5 x 5	6,400	.15
32 x 32	7 x 7	94,864	1.25
16 x 16	7 x 7	38,416	.53
8 x 8	7 x 7	19,600	.27
32 x 32	9 x 9	186,624	1.71
16 x 16	9 x 9	82,944	.88
8 x 8	9 x 9	46,656	.47
32 x 32	11 x 11	327,184	3.09
16 x 16	11 x 11	156,816	1.70
8 x 8	11 x 11	94,864	1.00

5.6 Implementation of Local Processing on Small Computers

An obvious question is: "Since local processing requires less storage than global processing, could it be effectively implemented using smaller computers?" The author has not attempted such a task. However, he can make recommendations as to what should be considered in such an undertaking.

While storage is not a large problem with local processing it does impose certain constraints when using small computers. When using a large general purpose computer, like the CDC-7600, N x N working sections are processed by first reading in N scan lines of the image then processing each of the required working sections in the strip of the image. If the image is large, this technique cannot be used on small machines. A simple solution would be to presection the image before starting the MAP restoration, then only the working section is required in memory at time of processing.

Computation time will remain the largest obstacle in the MAP restoration calculation, as it was on the large machine. Computations will require at least 6 to 7 significant digits or a minimum of 32 bits in floating point representation. It will be most important to have a machine which performs floating point arithmetic with hardware.

Since any image processing computation requires a large amount of input/output, the small machine used for such a task should have a good file handling system. This segment of the processing may be able to minimize some of the disadvantages in computation speed. It is standard practice to overlap input/output with computation as much as possible. Since computation time on the smaller computer is quite long, it may be

possible to "hide" most of the input/output, so that little time is spent waiting for new data.

The MAP restoration method can be implemented on small computers but the price will be high. However, the results from local processing indicate that it is most beneficial to process only that part of the image which is of interest. This reduces the size of the computation and makes the use of the small machines feasible.

CHAPTER 6

Conclusions

6.1 Summary

The purpose of this dissertation is to relate the mathematically based MAP restoration method to physical processes, which was not done in the original work [10], and to use this relation to derive an improved solution method. The MAP method was related to another image restoration method, constrained least squares estimation. A heuristic explanation was given for the behavior of the MAP solution and experimental results conformed to the predictions which were made on the basis of this explanation. From this heuristic explanation a faster solution algorithm was derived. A natural convergence criterion was proposed which was related to the noise, and the residual image derived from this criterion was found helpful in evaluating the quality of the restoration. The solution method derived from the heuristic explanation was shown to be comparable to the modified Picard's method in numerical analysis. Lastly, the solution quality was improved by using results from local processing.

6.2 Conclusions

The new MAP solution method makes MAP restoration computationally feasible but the cost is still many times the cost of more standard techniques such as Wiener filtering and power spectrum equalization. This cost, together with the results from local processing, suggests that a

reasonable approach to the restoration of an image would be first to apply one of the faster techniques. Analysis of the output from the less costly method would show if further processing were needed by the analyst. If further processing is desired, only that portion of the image which requires further processing is extracted and MAP restoration applied to that part. Analysis of the residual image will aid the analyst in determining if further processing with the MAP algorithm will be helpful. If time is very important the MAP iteration can be started with the output of another method and thus used as a refinement technique.

At present, studies are being done comparing the quality of the MAP restorations with restorations generated by other methods. Preliminary results from this study, as well as results shown in Chapter 3, indicate the MAP restorations are as good as or better than other standard methods. The noise control displayed by the MAP restorations gives it a particular advantage in cases of low signal-to-noise ratios. Further, as was mentioned in Chapter 3, since larger noise variances result in larger solution spaces, the MAP restoration algorithm requires fewer iterations for high noise cases and becomes less costly.

Local processing applied to MAP restoration proved to be beneficial in producing better image quality and in the case of small point spread functions was faster. Since only a small portion of the image is needed for computation at a single time, the method can be used on small computers with much less central memory than a CDC-7600. Computation time would be a problem on the slower minicomputers but the growing use of array processors could alleviate this problem in the near future.

While local processing can be viewed as small scale global processing, there are cases where the accuracy of local processing is needed

and it is necessary to keep a large picture intact. Such a case might arise in medical imagery where larger landmarks are needed to aid in the analysis of smaller detail. Of course, very small section sizes make extracting each section and processing it separately computationally unfeasible.

6.3 Areas for Further Research

MAP restoration has three parameters, $\underset{\sim}{\bar{f}}$, $\underset{\sim}{f}_0$, and $\underset{\sim}{R}_f$, which present the possibility of being used to apply additional a priori knowledge to finding a solution. It was pointed out in Chapter 3 that $\underset{\sim}{\hat{\bar{f}}}$ and $\underset{\sim}{\hat{R}}_f$ had little effect on the solution, hence they are not good candidates for manipulation. However, results from the same chapter showed the starting estimate $\underset{\sim}{f}_0$ to have a definite effect on both the speed of convergence and the quality of the solution. Suppose the image to be processed is text and certain letters are still illegible after the initial processing. It may be profitable to construct prototypes of the various possible letters and insert these into the $\underset{\sim}{f}_0$ and restart the MAP algorithm. The algorithm must still converge to a feasible solution, i.e., satisfy the convergence criterion; however, since the solution is dependent upon the initial estimate, by looking at the solution it should be possible to determine if the initial estimate was feasible.

While the MAP equations and derivations presented in this dissertation were done with the assumption of Gaussian statistics for both $\underset{\sim}{g}$ and $\underset{\sim}{f}$, there are other cases of interest. The MAP equation has been derived by the author using Poisson statistics for use in the case of electronic photon counting sensors. There are probably other applications unknown to the author where other statistical distributions are appropriate. Since

the fundamental MAP equation is derived in a straightforward manner from Bayes' law, there should be no problem in deriving new equations. Simple convergence criteria as well-defined as that for the Gaussian distribution are much less likely to be found for other distributions. It is likely that residual images can be found which will possess much the same properties of the residual image given for the Gaussian case.

Although the current research has resulted in a computation speed-up of more than a factor of ten, the MAP algorithm is still very slow. It is the author's opinion that little can be achieved in speed increases from new or modified iterative numerical methods. Investigation of better initial estimates and how to attain them would appear to be a more fruitful area. Since one of the advantages of MAP restoration is the absence of artifacts, using the output of a method which has produced such artifacts seems self-defeating. However, an artifact-producing restoration method might be modified to produce relatively artifact-free input for the MAP restoration. For example, a Wiener filter might be altered to do a less aggressive restoration in order to minimize artifacts. Of course, the output of such a modified method would be suboptimal for that method, but it might be much better as input to the MAP algorithm. Preprocessing of the input or postprocessing of the output of other restoration methods could achieve the same result.

Local processing has proven its worth in MAP restoration. It has not been applied to other restoration methods. Investigating which methods can benefit from local processing would yield useful information, both about the possible improvement of the methods and about the characteristics of the methods. Because of its similarity to MAP restoration, constrained least squares estimation would be a prime candidate to benefit from local

processing. The power spectrum equalization method could probably benefit by using different prototype spectra on different parts of the image. Wiener filtering presents a problem in that a noise estimate is required, and as sections get smaller this estimate becomes less accurate; this might be overcome by using prototypes similar to the case for power spectrum equalization.

In conclusion, the areas opened by MAP restoration and local processing appear to be fertile. The quality of restorations has been advanced by the use of these two methods.

ACKNOWLEDGMENTS

I thank the members of my committee for their many helpful suggestions for this dissertation. I am especially indebted to the chairman, Dr. A. H. Koschmann, for his patience and guidance.

My thanks to Dr. B. R. Hunt whose previous work on MAP restoration laid the foundation for this dissertation. I learned most of what I know about image processing from Dr. Hunt.

I wish to thank Dr. D. H. Janney of the Los Alamos Scientific Laboratory for his encouragement and for allowing me the time to complete this work.

REFERENCES

1. H. C. Andrews, "Digital image restoration: a survey," Computer Vol. 7, No. 5, pp 36-45, May 1974.

2. M. Cannon, "Blind deconvolution of spacially invariant image blurs with phase," IEEE Trans. Acous., Speech, and Sig. Proc., Vol. ASSP-24, No. 1, pp 58-63, Feb. 1976.

3. E. R. Cole, The Removal of Unknown Image Blurs by Homomorphic Filtering, Univ. Utah, Comput. Sci. Dept. Rep. UTECCSc-029, June 1973.

4. G. B. Grant and T. S. Huang, "Piecewise Fourier transformation for picture bandwidth compression," IEEE Trans. Communication Tech., Vol. COM-19, No. 2, pp 133-139, April 1971.

5. A. Habibi and P. A. Wintz, "Image coding by linear transformation and block quantization," IEEE Trans. Communication Tech., Vol. COM-19, No. 1, pp 50-63, Feb. 1971.

6. E. F. Haugh, "A structural theory for the Selwyn granularity coefficient," J. of Photographic Science, Vol. 2, pp 65-68, 1963.

7. C. W. Helstrom, "Image restoration by the method of least-squares," J. Opt. Soc. Amer., Vol. 57, pp 297-303, Mar. 1967.

8. T. S. Huang, "Some notes on film grain noise," Restoration of Atmospherically Degraded Images, Vol. 2, Woods Hole Summer Study, pp 105-109, July 1966.

9. B. R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," IEEE Trans. Comp., Vol. C-22, No. 9, pp 805-812, Sept. 1973.

10. B. R. Hunt, "Bayesian methods in digital image restoration," to be published, IEEE Trans. Computers.

11. B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for Gaussian models of images," to be published IEEE Trans. Systems, Man, and Cyber.

12. E. Isaacson and H. B. Keller, Analysis of Numerical Methods, New York: Wiley, 1966.

13. C.E.K. Mees, The Theory of the Photographic Process, New York: Macmillan, 1954.

14. J. F. Walkup and R. C. Choens, "Image processing in signal dependent noise," Opt. Engr., Vol. 13, pp 258-266, May 1974.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1976—677-343/215