A Differential Equation Approach to Maximum Entropy Image Reconstruction

X. Zhuang
E. ϕstevold
R. M. Haralick

Reprinted from: IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Vol. ASSP-35, No. 2, FEB. 1987.

A Differential Equation Approach to Maximum Entropy Image Reconstruction

XINHUA ZHUANG, EINAR ØSTEVOLD, AND ROBERT M. HARALICK, FELLOW, IEEE

Abstract—**A new algorithm is developed for solving the maximum entropy (ME) image reconstruction problem. The problem is reduced to solving a system of ordinary differential equations with appropriate initial values. The choice of initial values closely relates to the satisfaction of constraints, and we show how initial values are determined. The algorithm does not involve any optimization method. Instead of searching in the $(n + 1)$-dimensional space as required for most ME algorithms, our approach relies on solving a one-dimensional search along a well-defined and easily mastered path. Moreover, an efficient algorithm is developed to handle the search. The computer reconstruction verifies the theory.**

I. Introduction

THE success of maximum entropy methods in image reconstruction and spectrum estimation [1]–[9] has encouraged researchers in designing more efficient algorithms. In this paper, a basically new algorithm is developed for the maximum entropy reconstruction. The problem of maximum entropy reconstruction is reduced to solving a system of differential equations with easily obtained initial values. Most of the maximum entropy literature attacks the problem by some kind of optimization technique. The algorithm proposed here does not involve any optimization. It does not require search in an $(n + 1)$-dimensional space, which is required by most optimization techniques used in finding a maximum entropy solution. Here n represents the number of pixels in the image to be reconstructed. Instead, it performs a search along a path in the $(n + 1)$-dimensional space which is defined by the initial value problem of differential equations. Thus, the problem is reduced to a one-dimensional search. Since n, the number of pixels in the image, may be very large, the complexity of maximum entropy reconstruction is greatly reduced by this method. A similar idea could be used in spectrum estimation.

In Section II of this paper, the basic concepts for the maximum entropy reconstruction technique are given. This part includes the derivation of a system of differential equations defining a branch of solutions over which the one-dimensional search is performed. The section concludes with an analysis of the existence interval for λ, the search parameter. Section III deals mainly with the problem of selecting initial values so that the constraints imposed on the original formulation of the problem can be satisfied. A more comprehensive analysis of this subject is carried out in Section IV. The algorithm for solving the system of differential equations for the maximum entropy reconstruction problem is outlined in Section V. Section VI deals with the problem of adjusting a parameter in order to satisfy the constraint on the total image intensity embedded in the original problem formulation. In Section VII, experiments in image reconstruction using the algorithm outlined in Section V are described and results are presented. Conclusions follow in Section VIII.

Manuscript received September 5, 1984; revised June 16, 1986.

X. Zhuang is with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, on leave from Zhejiang Computing Technique Institute, Hangzhou, People's Republic of China.

E. Østevold is with the Norwegian Defense Research Establishment, Oslo, Norway.

R. M. Haralick is with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195.

IEEE Log Number 8610884.

II. Maximum Entropy, Stationary Point Equation, and Differential Equations

Let the required reconstructed image have pixel values represented by positive numbers $f_1, \cdots, f_n$ which are to be determined, and on which the entropy

$$H(p_1, \cdots, p_n) \triangleq -\sum_{i=1}^{n} p_i \log p_i,\ p_i \triangleq f_i \Big/ \sum_{t=1}^{n} f_t,\quad i = 1, \cdots, n \tag{1}$$

is defined. The entropy depends only on the distribution of gray levels in the image and not on the total intensity $\Sigma_{i=1}^{n} f_i$.

Let the observed image data be given by

$$d_j = \sum_{i=1}^{n} A_{ji} f_i + e_j,\quad j = 1, \cdots, m, \tag{2}$$

where the e_j's represent independent zero mean, σ_j^2 variance noise terms. We define

$$\begin{aligned} Q(f_1, \cdots, f_n) &\triangleq \tfrac{1}{2} \sum_{j=1}^{m} 1/\sigma_j^2 \left(\sum_{i=1}^{n} A_{ji} f_i - d_j \right)^2 \\ &= \tfrac{1}{2}(Af - d)^T \operatorname{diag}\,[1/\sigma_1^2, \cdots, 1/\sigma_m^2]\,(Af - d) \\ &= \tfrac{1}{2}\|Af - d\|_D^2, \end{aligned} \tag{3}$$

where $A \triangleq [A_{ji}]_{m,n}$, $f \triangleq [f_1, \cdots, f_n]^T$, $d \triangleq [d_1, \cdots, d_m]^T$, and $D \triangleq \operatorname{diag}\,[1/\sigma_1^2, \cdots, 1/\sigma_m^2]$. Typical least

squares approaches would try to determine those values $f_1, \cdots, f_n$ which minimize $Q(f_1, \cdots, f_n)$. Rather than do this, we seek those $f_1, \cdots, f_n$ which maximize entropy subject to the constraint

$$Q(f_1, \cdots, f_n) = m/2. \tag{4}$$

The motivation for this constraint comes about from the central limit theorem (see [10]) which states that with the probability one

$$\lim_{m \to \infty} 1/m \sum_{j=1}^{m} e_j^2/\sigma_j^2 = 1.$$

Thus, provided that m is large, we would expect the true value of $f_1, \cdots, f_n$ to satisfy (4). The condition $Q = m/2$ now determines the set of feasible images f which passes the given statistical test for consistency with the actual image data $\{d_1, \cdots, d_m\}$.

Although any of the feasible images is acceptable as a reconstruction, the maximum entropy criterion selects that particular f which has the least configurational information, i.e., the one where the pixel values are least separated. Hence, it can be looked upon as a smoothing criterion. Of all reconstructions which approximately fit the actual data $\{d_1, \cdots, d_m\}$, the solutions of higher entropy thus represent more "disorder," they are more "probable," "less predictable," they "assume less," and they are more "natural" according to Shannon's interpretation of entropy as an information measure.

Formally we are to maximize the entropy $H(p_1, \cdots, p_n)$ given the constraints $Q(f_1, \cdots, f_n) = m/2$ and $\Sigma_{i=1}^{n} f_i = t (t > 0)$. We introduce the second constraint for two reasons. One is that the total intensity has a status different from individual pixel values. It does not contribute to the shape of the gray tone intensity surface of the image f. For the second reason, we must look to the equality

$$H(p_1, \cdots, p_n) = H(f_1, \cdots, f_n) \Big/ \sum_i f_i - \log \sum_i f_i,$$

where $H(f_1, \cdots, f_n) \triangleq -\Sigma_i f_i \log f_i$. Introducing the second constraint $\Sigma_i f_i = t$, we obtain a linear relation between $H(f_1, \cdots, f_n)$ and $H(p_1, \cdots, p_n)$ enabling us to treat $H(f_1, \cdots, f_n)$ instead of $H(p_1, \cdots, p_n)$, the first one being much more tractable. Anyway, this constraint is not very strict, and it may be varied to obtain any required total intensity.

From the relation between $H(p_1, \cdots, p_n)$ and $H(f_1, \cdots, f_n)$, it is easily seen that the following three problems are equivalent to each other.

Problem 1:

$$\max H(p_1, \cdots, p_n)$$
$$\text{subject to } Q = m/2, \sum_i f_i = t.$$

Problem 2:

$$\max H(f_1, \cdots, f_n)$$
$$\text{subject to } Q = m/2, \sum_i f_i = t.$$

Problem 3:

$$\max \left| H(f_1, \cdots, f_n) + \mu \sum_i f_i - \lambda Q(f_1, \cdots, f_n) \right|$$
$$\text{subject to } Q = m/2, \sum_i f_i = t.$$

Instead of solving problem 3, we solve the following problem.

Problem 4:

$$\max \left| H(f_1, \cdots, f_n) + \mu \sum_i f_i - \lambda Q(f_1, \cdots, f_n) \right|$$
$$\text{subject to } Q = m/2.$$

In problem 4, the total intensity is treated like "a free boundary condition." No specific value is assigned to it in advance. In fact, problem 4 is a very reasonable formulation of the maximum entropy image reconstruction. If $f_1^*, \cdots, f_n^*$ is a solution of problem 4, then it is also a solution of problems 1, 2, and 3 with $t = \Sigma_i f_i^*$. Usually we are satisfied with the solution of problem 4. If not, we could adjust the parameter μ so that the solution of problem 4 has the required total intensity (see Section VI).

In the following, we first consider the unconstrained maximization problem related to problem 4.

Problem 5:

$$\max \left| H(f_1, \cdots, f_n) + \mu \sum_i f_i - \lambda Q(f_1, \cdots, f_n) \right|.$$

Let

$$J(f_1, \cdots, f_n; \mu, \lambda) \triangleq H(f_1, \cdots, f_n) + \mu \sum_i f_i - \lambda Q(f_1, \cdots, f_n). \tag{5}$$

Then problem 5 is to find maximal points of $J(f_1, \cdots, f_n; \mu, \lambda)$. The function $J(f; \mu, \lambda)$ is defined and continuous in the closed domain $\bar{E}$, where $E \triangleq \{(f_1, \cdots, f_n): f_i > 0, i = 1, \cdots, n\}$. When $\lambda \geq 0$, the continuous function J has a unique maximal point which is finite since J tends to minus infinity as $\|f\| \to \infty$ and is strictly concave, the latter is due to the negative definiteness of its Jacobian, $\nabla^2 J$:

$$\nabla^2 J = -\text{diag}\,[1/f_1, \cdots, 1/f_n] - \lambda A^T D A < 0. \tag{6}$$

As known, if the maximal point f^0 is an inner point, i.e., $f^0 \in E$, then the following stationary point equation holds:

$$\nabla J = -[\log f_1^0, \cdots, \log f_n^0] + (\mu - 1)h - \lambda A^T D(Af^0 - d) = 0, \tag{7}$$

where for abbreviation $h \triangleq [1, \cdots, 1]_{1 \times n}^T$.

Conversely, if ∇J equals zero at a point $f^0 \in E$ as $\lambda \geq 0$, then f^0 must be a maximal point, as easily seen from (6).

In the following, we will prove that (7) is always solvable in E as $\lambda \geq 0$. Thus, problem 5 is actually equivalent to (7) as $\lambda \geq 0$. The solution of (7) gives that unique maximal point required by problem 5 as $\lambda \geq 0$.

To prove that (7) is always (uniquely) solvable as $\lambda \geq 0$, all we need is to find such a path $f(\lambda; \mu)$ in E, defined for $0 \leq \lambda < \infty$, that the gradient ∇J is identically equal to zero along it:

$$\nabla J(f(\lambda; \mu); \mu, \lambda) \equiv 0, \quad 0 \leq \lambda < \infty. \tag{8}$$

It is clear that (8) is equivalent to the following (9) and (10):

$$\frac{d}{d\lambda} \nabla J(f(\lambda; \mu); \mu, \lambda) \equiv 0, \; 0 \leq \lambda < \infty, \tag{9}$$

$$\nabla J(f(0; \mu); \mu, 0) = 0. \tag{10}$$

It is also clear that (9) is equivalent to the following differential equations for $f(\lambda; \mu)$ (by working out the derivative $(d/d\lambda)\nabla J$):

$$\nabla^2 J(f; \mu, \lambda) \frac{df}{d\lambda} = \nabla Q(f), \quad 0 \leq \lambda < \infty, \tag{11}$$

and (10) directly gives $f(0; \mu)$ as

$$f(0; \mu) = \exp(\mu - 1)h, \tag{12}$$

where h equals $[1, \cdots, 1]^T_{1 \times n}$.

Thus, the stationary point equation is always (uniquely) solvable in E as $\lambda \geq 0$ if the following Cauchy problem of differential equations defines a solution curve in E for $0 \leq \lambda < \infty$:

$$\begin{cases} \nabla^2 J \dfrac{df}{d\lambda} = \nabla Q \\ f|_{\lambda = 0} = \exp(\mu - 1)h. \end{cases} \tag{13}$$

This is proved in Theorem 1 of the Appendix. Having shown that the unconstrained maximization, problem 5, has the form of (13), we are ready to consider the variety of computing schemes available to solve it.

To solve the maximum entropy image reconstruction problem, the only thing left is to select suitable initial conditions for the Cauchy problem (13), in order for the constraint $Q = m/2$ to be reached by the solution curve. For this we need to know more about the solution curve of (13). We discuss this in the next section.

III. Properties of Reduced System of Differential Equations

In this section we define sets giving the μ's for which a maximum entropy solution satisfying the constraint $Q = m/2$ can be found. Naturally we are interested in the behavior of $Q(f(\lambda; \mu))$ along the solution curve. It is easy to see that

$$\frac{dQ(f(\lambda; \mu))}{d\lambda} \equiv (\nabla Q)^T \frac{df}{d\lambda} \equiv (\nabla Q)^T (\nabla^2 J)^{-1} (\nabla Q). \tag{14}$$

In the following, we prove that $Q(f(\lambda; \mu))$ decreases strictly monotonically as λ increases unless the solution curve $f(\lambda; \mu)$ shrinks into a single point $\exp(\mu - 1)h$. As a matter of fact, along the solution curve there are only two possibilities. Either

$$f(\lambda; \mu) \equiv \exp(\mu - 1)h \quad \text{and} \quad \nabla Q(\exp(\mu - 1)h) = 0, \tag{15}$$

or

$$\nabla Q(f(\lambda; \mu)) \neq 0, \quad 0 \leq \lambda < \infty. \tag{16}$$

Thus, when $f(\lambda; \mu)$ does not shrink into a single point, the derivative $dQ(f(\lambda; \mu))/d\lambda$ will always be negative because of (14) and (16) and $(\nabla^2 J)^{-1} < 0$. As for (15) and (16), we can proceed as follows.

Suppose there exists a point $f^* = f(\lambda^*; \mu)$ along the solution curve such that

$$\nabla Q(f^*) = 0. \tag{17}$$

Then $f(\lambda; \mu)$ will coincide with the solution curve defined by the following Cauchy problem due to the uniqueness of the solution to

$$\begin{cases} \nabla^2 J \dfrac{df}{d\lambda} = \nabla Q, \\ f|_{\lambda = \lambda^*} = f^*. \end{cases} \tag{18}$$

However, the solution curve of (18) obviously is a single point f^* due to (17) and the initial condition in (18). Thus, it holds

$$f(\lambda; \mu) \equiv f^*, \quad 0 \leq \lambda < \infty. \tag{19}$$

Now it is not hard to see that (15) is valid since $f(0; \mu) = \exp(\mu - 1)h$. Thus, we have proved the following.

Theorem 3.1: The derivative $dQ(f(\lambda; \mu))/d\lambda$ is always negative unless $\nabla Q(\exp(\mu - 1h)) = 0$.

In order for the solution curve $f(\lambda; \mu)$ to satisfy the constraint $Q = m/2$, only those μ's need to be considered which satisfy either

$$Q(\exp(\mu - 1)h) = m/2. \tag{20}$$

or

$$Q(\exp(\mu - 1)h) > m/2 \quad \text{and} \quad \nabla Q(\exp(\mu - 1)h) \neq 0. \tag{21}$$

If (20) holds, then $\exp(\mu - 1)h$ solves problem 4. Otherwise, we need to select suitable μ's by means of (21) and then compute $f(\lambda; \mu)$ by means of (13).

Let the sets U, V, and W be defined as follows:

$$U = \{\mu: Q(\exp(\mu - 1)h) = m/2\}, \tag{22}$$

$$V = \{\mu: Q(\exp(\mu - 1)h) > m/2\}, \tag{23}$$

$$W = \{\mu: \nabla Q(\exp(\mu - 1)h) \neq 0\}. \tag{24}$$

From the previous discussion, it is clear that only those μ's which belong to U or $V \cap W$ can produce meaningful results.

IV. Computing Sets U, V, and W

In this section we will further investigate the sets U, V, and W, in order to determine initial value μ for which a

maximum entropy solution satisfying the constraint can be found.

For abbreviation we let

$$\alpha = \exp(\mu - 1)$$

$$r = \left[\sum_{i=1}^{n} A_{1i}/\sigma_1, \cdots, \sum_{i=1}^{n} A_{mi}/\sigma_m\right]^T,$$

$$s = [d_1/\sigma_1, \cdots, d_m/\sigma_m]^T. \tag{25}$$

Then a direct computation yields

$$Q(\alpha h) = a\alpha^2 + b\alpha + c, \tag{26}$$

where three coefficients a, b, c are as follows:

$$\begin{cases} a = \frac{1}{2}\|r\|^2, \\ b = -(r, s), \\ c = \frac{1}{2}\|s\|^2. \end{cases} \tag{27}$$

By using the Schwartz inequality, it is easily verified that $b^2 \leq 4ac$. We will prove that the expectation $E\{c\} \geq m/2$. Let $\Sigma_i A_{ji} f_i = g_j$. Then

$$c = \frac{1}{2}\sum_j (g_j + e_j)^2/\sigma_j^2$$

$$= \frac{1}{2}\sum_j g_j^2/\sigma_j^2 + \sum_j g_j e_j/\sigma_j^2 + \frac{1}{2}\sum_j e_j^2/\sigma_j^2, \tag{28}$$

and

$$E\{c\} = \frac{1}{2}\sum_j g_j^2/\sigma_j^2 + m/2 \geq m/2, \tag{29}$$

$$E\{c\} = m/2 \Leftrightarrow \forall j,\ g_j = 0. \tag{30}$$

Furthermore, we can prove the following theorem.

Theorem 4.1: Assume that $1/m^2 \Sigma_{j=1}^m g_j^2/\sigma_j^2 \to 0$. Then with probability one, $1/m\Sigma_j g_j e_j/\sigma_j^2 \to 0$, and $1/m(c - \frac{1}{2}\Sigma_j g_j^2/\sigma_j^2) \to \frac{1}{2}$, when $m \to \infty$.

Proof: Let $U_j = g_j e_j/\sigma_j^2$. Then the U_j's are independently distributed with zero mean and variance g_j^2/σ_j^2. Let $X_m = 1/m\ \Sigma_{j=1}^m U_j$. It is then easily seen that $E\{X_m\} = 0$, and $\lim \mathrm{Var}\{X_m\} = \lim E\{X_m^2\} = \lim (1/m^2)\ \Sigma_j g_j^2/\sigma_j^2 = 0$, $m \to \infty$, by assumption. Hence, by Chebychev's inequality [10, p. 151], $X_m \to 0$ with probability one when $m \to \infty$, which proves the first part of the theorem. Note that $1/m(c - \frac{1}{2}\Sigma_j g_j^2/\sigma_j^2) = \frac{1}{2}m\ \Sigma_j e_j^2/\sigma_j^2 + 1/m\ \Sigma_j g_j e_j/\sigma_j^2$. The term $1/m\ \Sigma_j e_j^2/\sigma_j^2$ approaches 1 ($m \to \infty$) with probability one by the central limit theorem, as mentioned before. Combining this fact with $1/m\ \Sigma_j g_j e_j/\sigma_j^2 \to 0$ ($m \to \infty$) with probability one, we have with probability one $1/m(c - \frac{1}{2}\Sigma_j g_j^2/\sigma_j^2) \to \frac{1}{2}$ ($m \to \infty$). Thus, for large m, c can rather safely be replaced by $m/2 + \frac{1}{2}\Sigma_j g_j^2/\sigma_j^2$ ($\geq m/2$). Hence, in the sequel we assume $c \geq m/2$. Q.E.D.

Using theorem 4.1 and assuming $\sigma_j^2 = \sigma^2$ for all j, we can for large m derive a simple expression for the signal-to-noise ratio (S/N):

$$(S/N) \triangleq \sum_j g_j^2 \Big/ \sum_j e_j^2$$

$$= \left(\sum_j d_j^2 - m\sigma^2\right) \Big/ (m\sigma^2) = \sum_j d_j^2 \Big/ (m\sigma^2) - 1.$$

Now using the assumption $c \geq m/2$, we are able to compute sets U, V, and W. We are interested in obtaining a nonempty U or the intersection set $V \cap W$ when $U = \phi$.

Computing the set $U = \{\mu : Q(\exp(\mu - 1)h) = m/2\}$ we have the following.

Case 1: $a > 0$, $b^2 - 4a(c - m/2) \geq 0$, $b < 0$:

$$U = \begin{cases} \{1 + \log[(-b \pm \sqrt{b^2 - 4a(c - m/2)})/(2a)]\}, \\ \quad c > m/2 \\ \{1 + \log(-b/a)\}, \quad c = m/2. \end{cases}$$

Case 2: $a > 0$, $b^2 - 4a(c - m/2) \geq 0$, $b > 0$: $V = \{(-\infty, \infty)\}$.

In all other cases, $U = \phi$, an empty set. The second case occurs very rarely. In general, U is empty or has one or two elements. This corresponds to no equientropy solution or one (or two) equientropy solution(s) for problem 4, respectively.

Computing the set $V = \{\mu : Q(\exp(\mu - 1)h) > m/2\}$ when $U = \phi$ we have the following.

Case 1: $a > 0$ and $b^2 - 4a(c - m/2) < 0$: $V = \{(-\infty, \infty)\}$.

Case 2: $a > 0$, $b^2 - 4a(c - m/2) \geq 0$, $b > 0$: $V = \{(-\infty, \infty)\}$.

Case 3: $a = 0$ ($\Rightarrow b = 0$) and $c > m/2$: $V = \{(-\infty, \infty)\}$.

Hence, $V = \{(-\infty, \infty)\}$ when $U = \phi$.

Computing the set $W = \{\mu : \nabla Q(\exp(\mu - 1)h) \neq 0\}$, we can rewrite $W = \{\mu : \exp(\mu - 1)A^T DAh \neq A^T Dd\}$. It is clear that W equals $\{(-\infty, \infty)\}$ unless the $(n \times 1)$-vector $A^T Dd$ has the same direction as the $(n \times 1)$-vector $A^T DAh$. The exceptive case occurs very rarely. Thus, in general, $W = \{(-\infty, \infty)\}$.

In summary, there are only two possibilities in general. Either U is not empty, and thus, the problem 4 has at most two equientropy solutions. Or $V \cap W = \{(-\infty, \infty)\}$, and thus, for any μ there holds

$$Q(\exp(\mu - 1)h) > \frac{m}{2}, \tag{31}$$

$$\nabla Q(\exp(\mu - 1)h) \neq 0. \tag{32}$$

As known from theorem 3.1 (32) assures

$$\frac{dQ(f(\lambda;\mu))}{d\lambda} < 0 \quad \text{as } \lambda \geq 0, \tag{33}$$

where remember: $f(\lambda;\mu)$ is the unique solution of (13).

Intuitively, from (31), we would like to choose such a μ that $Q(\exp(\mu - 1)h)$ is as close to $m/2$ as possible. When $-b/2a > 0$, we can simply choose

$$\mu = 1 + \log(-b/2a), \tag{34}$$

because of $[(d/d\alpha)\ Q(\alpha h)]_{\alpha = -b/2a} = 0$.

From (27) it is apparent that the condition, $-b/2a > 0$, is equivalent to $(r, s) > 0$. Notice that d_j and f_i are related by

$$d_j = \sum_{i=1}^{n} A_{ji} f_i + e_j, \quad j = 1, \cdots, m, \tag{35}$$

where all f_i are nonnegative, and the matrix $(A_{ji})_{n \times m}$ represents the blurring operation which occurred in the imaging processes, the latter implies $A_{ji} \geqslant 0$ in most practical cases. Thus, $(r, s) > 0$ will hold whenever the noise is not too large.

V. Algorithm for Solving Differential Equations

In this section, the algorithm for solving (13) is presented. As mentioned in the Introduction, we do a search along the solution curve by varying the parameter λ, starting with $\lambda_0 = 0$. A satisfactory solution is found when $Q(f(\lambda;\mu))/(m/2)$ is sufficiently near 1. For each λ, a large system of linear equations must be solved. This is done by making extensive use of the properties of the coefficient matrix and applying the Gauss–Seidel algorithm [11] which is an iterative technique providing fast convergence.

Suppose that the set U is empty, i.e., problem 4 has no equientropy solution, and moreover, an appropriate μ has been chosen from the intersection set $V \cap W$. From (13) we have the following discretized iterative equations:

$$\begin{cases} f^0 = \exp(\mu - 1)h, \\ k \geqslant 0: \\ (F^k + \lambda_k A^T D A)(f^{k+1} - f^k) \\ \quad = -(\lambda_{k+1} - \lambda_k) A^T D (A f^k - d), \end{cases} \tag{36}$$

where $F^k = \operatorname{diag}[1/f_1^k, \cdots, 1/f_n^k]$, $\lambda_0 = 0$, and $\lambda_{k+1} > \lambda_k$. When all $|\lambda_{k+1} - \lambda_k|$ are small enough, f^k will approximate $f(\lambda_k; \mu)$ very well. Equation (36) can be rewritten as follows:

$$\begin{cases} f^0 = \exp(\mu - 1)h, \\ k \geqslant 0: \\ (F^k + \lambda_k A^T D A) f^{k+1} \\ \quad = (2\lambda_k - \lambda_{k+1}) A^T D A F^k + h + (\lambda_{k+1} - \lambda_k) A^T D d. \end{cases} \tag{37}$$

This is a large linear system of equations. Fortunately, the coefficient matrix $(F^k + \lambda_k A^T D A)$ is positive definite and symmetric. The Gauss–Seidel iterative scheme (see [11]) is used to solve (37) efficiently.

Let P represent the matrix $(F^k + \lambda_k A^T D A)$. Let b represent the vector $(2\lambda_k - \lambda_{k+1}) A^T D A f^k + h + (\lambda_{k+1} - \lambda_k) A^T D d$. Further, let x be the $(k+1)$st estimate f^{k+1} of $f(\lambda_{k+1}; \mu)$. For $m \geqslant 1$ we have the following Gauss–Seidel iterative equations for solving $Px = b$:

$$\begin{aligned} x^{(0)} &= f^k, \\ x_i^{(m)} &= \frac{1}{P_{ii}} \left[b_i - \sum_{j<i} P_{ij} x_j^{(m)} - \sum_{j>i} P_{ij} x_j^{(m-1)} \right], \\ & \qquad i = 1, \cdots, n. \end{aligned} \tag{38}$$

It is proved (see [11]) that the convergence and the limit point are independent of the choice of $x^{(0)}$. Thus, $\lim x^{(m)} = f^{k+1}$. We choose $x^{(0)} = f^k$ since f^{k+1} should be near to f^k when $|\lambda_{k+1} - \lambda_k|$ is small. This choice of $x^{(0)}$ greatly speeds up the convergence of (38). Experiments indicate that in most cases only one iteration for each k is enough to obtain f^{k+1} from f^k. In that case, we have

$$\begin{aligned} f_i^{k+1} &= \frac{1}{P_{ii}} \left[b_i - \sum_{j<i} P_{ij} f_j^{k+1} - \sum_{j>i} P_{ij} f_j^k \right], \\ & \qquad i = 1, \cdots, n. \end{aligned} \tag{39}$$

In the sequel we let

$$\begin{cases} L = A^T D A, \\ p = A^T D d, \\ g^k = L f^k, \qquad k = 0, 1, 2, \cdots. \end{cases} \tag{40}$$

Then from (37) it follows

$$\begin{aligned} k \geqslant 0: (F^k + \lambda_k L) f^{k+1} &= (2\lambda_k - \lambda_{k+1}) g^k \\ &\quad + h + (\lambda_{k+1} - \lambda_k) p, \end{aligned} \tag{41}$$

$$\begin{aligned} k \geqslant 1: g^{k+1} &= \left(2 - \frac{\lambda_{k+1}}{\lambda_k}\right) g^k \\ &\quad + \frac{1}{\lambda_k} \left[\frac{f_1^k - f_1^{k+1}}{f_1^k}, \cdots, \frac{f_n^k - f_n^{k+1}}{f_n^k} \right]^T \\ &\quad + \left(\frac{\lambda_{k+1}}{\lambda_k} - 1\right) p. \end{aligned} \tag{42}$$

The initial values f^0, g^0, g^1 for (41)–(42) are as follows:

$$\begin{cases} f^0 = \exp(\mu - 1)h, \\ g^0 = L f^0 = \exp(\mu - 1) \left[\sum_{i=1}^{n} L_{1i}, \cdots, \sum_{i=1}^{n} L_{ni} \right]^T, \\ g^1 = L f^1. \end{cases} \tag{43}$$

Thus, (43), which involves matrix calculations, is used to find g^0, g^1, and the recursive formula (42) involving only vector calculations is used for g^k, $k \geqslant 2$.

Finally, we summarize this section by giving the following algorithm for solving the maximum entropy image reconstruction problem, which can be easily implemented.

Step 1: If $U \neq \phi$ and $\mu \in U$, then there is an equientropy solution $f^* = \exp(\mu - 1)h$. Stop.

Step 2: Choose an appropriate $\mu \in V \cap W$ such that $Q(\exp(\mu - 1)h)$ is as small as possible.

Step 3: Set $k = 0$, $\lambda_k = 0$, $f^k = \exp(\mu - 1)h$, $g^k = L f^k$, and $Q^k = Q(f^k)$.

Step 4: If $|Q^k/(m/2) - 1| \leq \epsilon$, then stop. Here ϵ is a prescribed small positive number.

Step 5: Solve (37) for f^{k+1} by using Gauss–Seidel iterative scheme (38).

Step 6: Compute $Q^{k+1} = Q(f^{k+1})$.
Step 7: If k is large enough, stop.
Step 8: If $|Q^{k+1}/(m/2) - 1| \leq \epsilon$, then stop.
Step 9: If $k \geq 1$, then compute g^{k+1} by (42), else $g^{k+1} = Lf^{k+1}$.
Step 10: If $Q^{k+1} > m/2$, then set $\delta\lambda > 0$, else set $\delta\lambda < 0$.
Step 11: Set $k = k + 1$, $\lambda_k = \lambda_{k-1} + \delta\lambda$.
Step 12: Go to step 5.

In the experiments we consider only cases where the blurring given by the matrix A can be represented as a convolution between the undegraded image and a (rectangular) mask. By making full use of the fact that just very few of the elements in the matrix L are nonzero for moderate mask sizes, it has been possible to construct a fast algorithm without resorting to Fourier transforms. If the blurring happens in the imaging processes, then the mask is usually small. For very large masks, however, the situation is different, and the possibility of time savings by use of Fourier transforms should be investigated. Equation (37) is, however, not convolution form, so implementation of Fourier transforms is not straightforward.

VI. Adjusting μ to Get the Required Total Intensity

Assume that the set U is empty and we have parameters μ_0 and λ_0 such that $\lambda_0 \geq 0$ and $|Q(f(\lambda_0; \mu_0))/(m/2) - 1| \leq \epsilon$. We are to look for a differentiable curve $\lambda = \lambda(\mu)$ passing through (μ_0, λ_0), along which the Q-values, $Q(f(\lambda(\mu); \mu))$, are constant. In this case, adjusting the parameter μ will not affect the satisfaction of the constraint $|Q/(m/2) - 1| \leq \epsilon$. Now along the required curve we should have two relations:

$$\frac{dQ(f(\lambda(\mu); \mu))}{d\mu} = 0, \tag{44}$$

$$\nabla J(f(\lambda(\mu); \mu); \mu, \lambda(\mu)) = 0, \tag{45}$$

the latter because $f(\lambda(\mu); \mu)$ is the stationary point of the function $J(f; \mu, \lambda(\mu))$ as known before.

From (45) it is easy to derive that

$$\nabla^2 J(f(\lambda(\mu); \mu); \mu, \lambda(\mu)) \frac{df(\lambda(\mu); \mu)}{d\mu} + h - \frac{d\lambda(\mu)}{d\mu} \nabla Q(f(\lambda(\mu); \mu)) = 0$$

or

$$\frac{df(\lambda(\mu); \mu)}{d\mu} = [\nabla^2 J(f(\lambda(\mu); \mu))]^{-1} \left[\frac{d\lambda(\mu)}{d\mu} \nabla Q(f(\lambda(\mu); \mu)) - h\right]. \tag{46}$$

Under the assumption $U = \phi$, we have that

$$V \cap W = \{(-\infty, \infty)\},$$

which imples that

$$\nabla Q\,(\exp(\mu - 1)h) \neq 0. \tag{47}$$

As stated before, (47) implies that

$$\nabla Q(f(\lambda(\mu); \mu)) \neq 0 \quad \text{everywhere.} \tag{48}$$

From (44) we have

$$[\nabla Q(f(\lambda(\mu); \mu))]^T \frac{df(\lambda(\mu); \mu)}{d\mu} = 0. \tag{49}$$

Combining (47)–(49) together, we obtain

$$\frac{d\lambda(\mu)}{d\mu} = \frac{[\nabla Q]^T[\nabla^2 J]^{-1}h}{[\nabla Q]^T[\nabla^2 J]^{-1}[\nabla Q]}, \tag{50}$$

where the differentiable curve passes through (μ_0, λ_0), i.e., $\lambda(\mu_0) = \lambda_0$.

Then, multiplying both sides of (46) by h^T, it follows

$$\frac{d \sum_{i=1}^{n} f_i(\lambda(\mu); \mu)}{d\mu} = h^T[\nabla^2 J]^{-1}\nabla Q + \frac{d\lambda}{d\mu} - h^T[\nabla^2 J]^{-1}h = \frac{([\nabla Q]^T[\nabla^2 J]^{-1}h)^2}{[\nabla Q]^T[\nabla^2 J]^{-1}[\nabla Q]} - h^T[\nabla^2 J]^{-1}h. \tag{51}$$

Finally, we obtain the following Cauchy problem of $(n + 2)$ differential equations for adjusting μ to get the required total intensity:

$$[\nabla^2 J] \frac{df(\lambda(\mu); \mu)}{d\mu} = \frac{d\lambda(\mu)}{d\mu} [\nabla Q] - h,$$

$$\frac{d\lambda(\mu)}{d\mu} = \frac{[\nabla Q]^T[\nabla^2 J]^{-1}h}{[\nabla Q]^T[\nabla^2 J]^{-1}[\nabla Q]},$$

$$\frac{d \sum_{i=1}^{n} f_i(\lambda(\mu); \mu)}{d\mu} = \frac{([\nabla Q]^T[\nabla^2 J]^{-1}h)^2}{[\nabla Q]^T[\nabla^2 J]^{-1}[\nabla Q]} - h^T[\nabla^2 J]^{-1}h,$$

$$f|_{\mu=\mu_0} = f(\lambda_0, \mu_0),$$

$$\lambda|_{\mu=\mu_0} = \lambda_0,$$

$$\sum_{i=1}^{n} f_i|_{\mu=\mu_0} = h^T f(\lambda_0, \mu_0). \tag{52}$$

The iterative scheme for solving (52) is as follows.
Given: $\mu_0, \lambda_0, f^0 = f(\lambda_0; \mu_0), t_0 = h^T f^0$.

$k \geq 0$:

$$\frac{\lambda_{k+1} - \lambda_k}{\mu_{k+1} - \mu_k} = \frac{[\nabla Q(f^k)]^T[\nabla^2 J(f^k; \mu_k, \lambda_k)]^{-1}h}{[\nabla Q(f^k)]^T[\nabla^2 J(f^k; \mu_k, \lambda_k)]^{-1}[\nabla Q(f^k)]},$$

$$[\nabla^2 J(f^k; \mu_k, \lambda_k)](f^{k+1} - f^k) = (\lambda_{k+1} - \lambda_k)[\nabla Q(f^k)] - (\mu_{k+1} - \mu_k)h,$$

$$\frac{t_{k+1} - t_k}{\mu_{k+1} - \mu_k} = \frac{([\nabla Q(f^k)]^T[\nabla^2 J(f^k; \mu_k, \lambda_k)]^{-1}h)^2}{[\nabla Q(f^k)]^T[\nabla^2 J(f^k; \mu_k, \lambda_k)]^{-1}[\nabla Q(f^k)]} - h^T[\nabla^2 J(f^k \mu_k, \lambda_k)]^{-1}h. \tag{53}$$

Fig. 1. Source image.

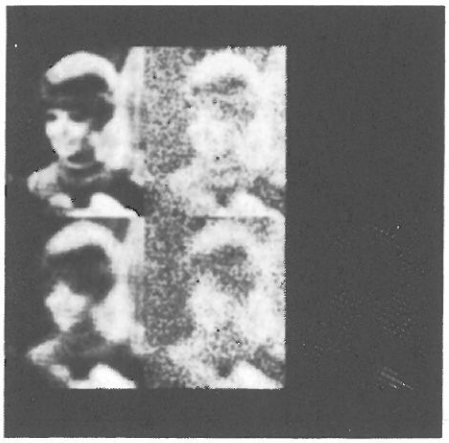

Fig. 2. Input images: 1–5.

VII. Experiments

Three source images were used in the experiments. The first one (Fig. 1) is a picture of a girl. The size is 128 × 128 pixels, and the resolution in gray level is 8 bits. Five different degraded images were produced from this image. These images, which are input images for the algorithm, will be referred to as images 1–5. Images 1 and 2 were created by first convolving the source image with a 5 × 5 mask with all elements equal to unity. Random noise with zero mean and standard deviations of 4 and 40, respectively, was added to the blurred image. This corresponds to the signal-to-noise ratio (S/N) of 1000 and 10, respectively, for the resulting images. To get a stronger blurring, this source image was convolved with a 7 × 7 mask with the center 5 × 5 elements all equal to zero and the 24 border elements equal to unity. Noise with zero mean and standard deviations of 4 and 40, respectively, was added to the blurred image to produce images 3 and 4. To test the ability of the algorithm to reconstruct an image when we have incomplete measurements, image 5 was created. It is identical to image 1, but the gray level is known for only every tenth pixel. This is, for convenience, done in a regular manner producing the line pattern which can be seen in Fig. 2 where the five input images are shown. Data for these images are listed in Table I.

The second source image [Fig. 3(a)] contains text and has size 480 columns by 100 lines. It was blurred by a 7 × 7 Gaussian filter with standard deviation of 5. Noise with standard deviation of 4 was added to produce the input image shown in Fig. 3(b). The third source image [Fig. 4(a)] is a 128 × 128 checkerboard image. The corresponding input image was produced by the same degradation process as was applied to the text image.

As stated by (36), the initial guess for the solution vector f is $f^0 = \alpha h$, where $\alpha = \exp(\mu - 1)$ and $h = [1, \cdots, 1]^T_{n\times 1}$. For fastest possible convergence, α is chosen so that $Q(\alpha)$ is smallest possible (provided $U = \phi$), that is, we choose α so that $dQ(\alpha, \cdots, \alpha)/d\alpha = 0$ [see (34)].

For each reconstruction, the algorithm was run until the resulting Q was sufficiently near $Q = m/2$ (m is the number of data points, i.e., pixels). The number of iterations for each λ and step size in λ, $\delta\lambda = |\lambda_{k+1} - \lambda_k|$ were determined empirically, and the chosen values represent a reasonable, but in no way the best, choice. The optimal choice of $\delta\lambda$ also varies strongly with noise level. The basic policy we use to control the step size in our exper-

TABLE I
Input Images Created from Girl Source Image

Image	Size of Mask	Standard Deviation of Noise	(S/N)	Sample Interval
Image 1	5 × 5	4.0	1000	1
Image 2	5 × 5	40.0	10	1
Image 3	7 × 7	4.0	1000	1
Image 4	7 × 7	40.0	10	1
Image 5	5 × 5	5.0	1000	10

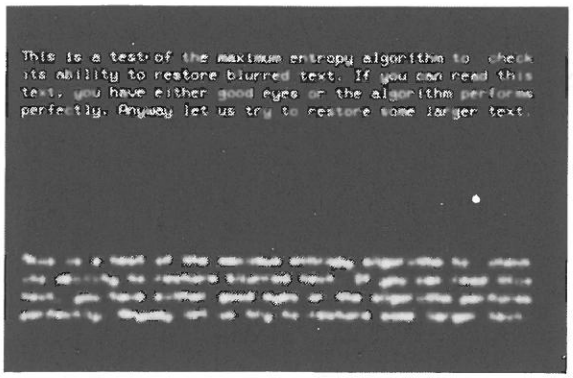

Fig. 3. Text image (a) Source image. (b) Degraded input image.

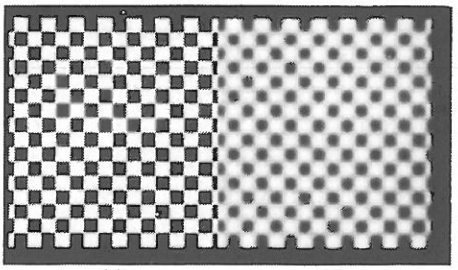

Fig. 4. Checkerboard image. (a) Source image. (b) Degraded input image.

TABLE II
DATA FOR RECONSTRUCTION OF GIRL IMAGE

Image	K	$\delta\lambda_1$	$\delta\lambda_2$	λ-Final	Q-Initial	Q-Final	$m/2$	ϵ	CPU
Image 1	1	0.2	0.2	0.2	769041	52588			1.36
Image 1	6	0.2	0.4	1.4	769041	10340		0.262	10.15
Image 1	12	0.2	0.4	3.8	769041	8381	8192	0.023	20.10
Image 2	1	4.0	4.0	4.0	15970	13539			1.36
Image 2	5	4.0	4.0	20.0	15970	8875		0.083	8.40
Image 2	9	4.0	4.0	36.0	15970	8379	8192	0.023	15.05
Image 3	1	0.2	0.3	0.2	643424	73208			2.50
Image 3	6	0.2	0.4	1.6	643424	11845		0.446	18.10
Image 3	10	0.25	0.5	4.0	643424	8423	8192	0.028	30.15
Image 4	1	4.0	4.0	4.0	14715	12814			2.50
Image 4	4	4.0	4.0	16.0	14715	9473		0.156	12.40
Image 4	8	4.0	4.0	32.0	14715	8412	8192	0.027	24.00
Image 5	1	0.5	0.5	0.5	77151	47400			1.45
Image 5	8	0.5	1.0	5.5	77151	3777		3.606	14.30
Image 5	15	0.5	1.0	12.5	77151	825	820	0.006	26.20

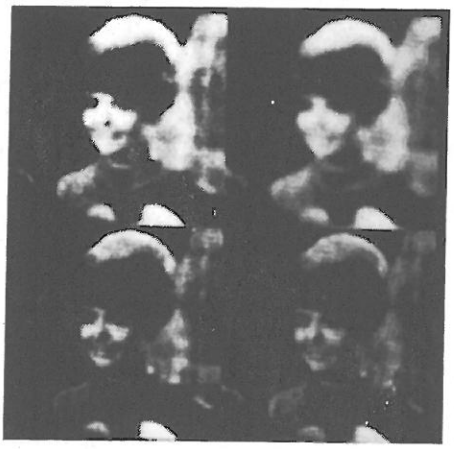

Fig. 5. Results for image 1. Image 1 (upper left), 1 step in λ (upper right), 6 steps (lower left), and final result after 12 steps (lower right).

Fig. 6. Results for image 2. Image 2 (upper left), 1 step in λ (upper right), 5 steps (lower left) and final result after 9 steps (lower right).

iments is: if the relative reduction in Q in one step is below a user specified limit, the step size is doubled for the next step. This can save a lot of time compared to keeping $\delta\lambda$ constant for a whole search. Indeed, in some cases, it is quite necessary in order to be able to reach a solution in a reasonable time. One Gauss–Seidel iteration was enough to obtain f^{k+1} from f^k with satisfactory accuracy as mentioned before.

Table II gives an overview of some of the experiments run for the five input images generated from the girl source image. K is the number of steps. $\delta\lambda_1$ is initial step size in λ, and $\delta\lambda_2$ is the final step size. Q-initial is the initial Q-value computed from the flat (constant) image which is the starting image. m represents the number of observed data. $\epsilon \triangleq |\ Q\text{-final} - (m/2)|/(m/2.)$ CPU is the CPU time in minutes and seconds.

The corresponding images are shown in Figs. 5–9 for images 1–5. The upper left image is the input image, upper right is the result after one step in λ, lower left is the intermediate result, and at the lower right the final result is shown.

Table III shows some of the experimental data for the text image and checkerboard image. The corresponding

Fig. 7. Results for image 3. Image 3 (upper left), 1 step in λ (upper right), 6 steps (lower left), and final result after 10 steps (lower right).

reconstructed images are shown in Figs. 10 and 11, respectively.

The overall results indicate a very good performance in reconstruction of blurred images even in cases where we have very few measurements. The performance in reconstruction of extremely noisy images seems to be a little

TABLE III
RECONSTRUCTION DATA FOR TEXT AND CHECKERBOARD IMAGE

Image	K	$\delta\lambda_1$	$\delta\lambda_2$	λ-Final	Q-Initial	Q-Final	$m/2$	ϵ	CPU
Text	22	0.50	256.0	768.5	827968	24574	24000	0.024	3.13.00
Check	6	0.25	0.5	2.0	1935043	50248		5.133	18.00
Check	25	0.25	32768.0	98308.0	1935043	9297	8192	0.135	1.15.00

Fig. 8. Results for image 4. Image 4 (upper left), 1 step in λ (upper right), 4 steps (lower left), and final result after 8 steps (lower right).

Fig. 9. Results for image 5. Image 5 (upper left), 1 step in λ (upper right), 8 steps (lower left), and final result after 15 steps (lower right).

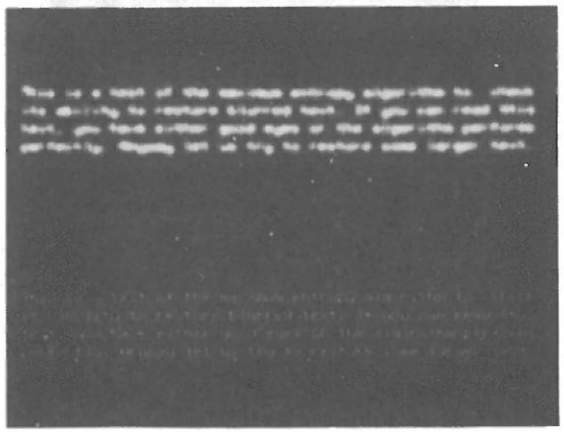

Fig. 10. Reconstruction of text image. (a) Input image. (b) Final result after 22 steps in λ.

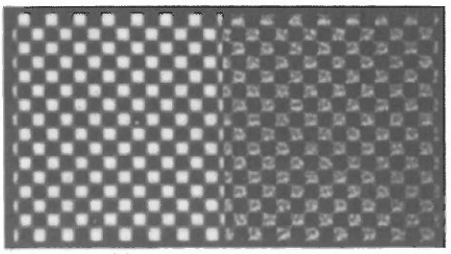

Fig. 11. Reconstruction of checkerboard image. (a) After 6 steps in λ. (b) Final result after 25 steps.

poorer. The experiments indicate that the algorithm runs better in deblurring than in noise removal. In fact, if no noise was present, it is possible (in principle at least) to obtain a perfect reconstruction (which in this case is a pure deconvolution or deblurring). In practice, however, noise is always present, and even very small amounts of noise could cause the reconstructed data to fluctuate some. The present algorithm, however, applies noise statistics [see (4)], which probably is the best way (statistically speaking) of dealing with the noise.

Figs. 5–9 indicate that in many cases it is not necessary to search all the way until Q reaches Q-final. The final results are not very much different from the intermediate results also displayed in the same figures. Thus, accepting the intermediate results as the final reconstruction, a lot of CPU time can be saved. As seen, the final results in Table II have their ϵ-values less than 0.018 and the intermediate results have their ϵ-values larger than 0.083. Therefore, letting $\epsilon = 0.1$ in the algorithm might give a reasonable stopping condition.

Table II reveals the importance of varying the step size $\delta\lambda$. The initial step size cannot be increased significantly from its present values, so a constant step size would in many cases make practical applications impossible.

The checkerboard image was the only case where the result did not undisputedly improve with the number of iterations. The average contrast between black and white squares did increase, but pixel values for the two regions are overlapping after 25 iterations (see Fig. 11), whereas this was not the case for just 6 iterations with a final λ of only 2.0 (shown in the same figure).

The experiments were run on a VAX-11/780 computer. Data for the experiments as well as CPU time consumed are listed in Tables II and III. Although the general problem complexity is the square of the number of pixels, the computational complexity can in many cases (for instance, when the degradation matrix has a simple convolution form) be reduced to being linear in the number of

pixels. This is confirmed through experiments on a 256 × 256 image. CPU time is, for many applications, crucial. Taking into account the size of the problem, the CPU times listed in Tables II and III are very satisfactory. It is, however, evident that this time can be reduced even more if a better technique for finding an optimal or near optimal $\delta\lambda$ for each step in the search along the path is developed.

VIII. Conclusions

The theoretical analysis in previous sections indicates that the maximum entropy image reconstruction is determined by a 1-D path, which is defined by a system of differential equations, i.e., (13), with easily obtained initial condition $(-b/2a)h$, where $a = \frac{1}{2}\|r\|^2$, $b = -(r, s)$, $h = [1, \cdots, 1]^T_{1\times n}$, and r, s are given in (25). Along the path, the Q-value is monotonically decreasing to reach at the satisfaction of the statistical constraint equation, i.e., $Q = m/2$.

The experiments run so far indicate that choosing $\epsilon = 0.1$ as the stopping condition of the algorithm, i.e., $|Q(f^k) - (m/2)|/(m/2) < 0.1$, gives quality reconstructions and thus is appropriate. This suggests accepting intermediate results as the final reconstruction, as mentioned before. As a result, a lot of CPU time is saved; and actual CPU time is less than or at least compatible with other techniques (e.g., see [8] which uses optimization technique quite efficiently and considerately). To save more CPU time, a more efficient procedure is needed to control the step size in each iteration. Although the step-size problem has been widely studied in the area of numerical analysis (e.g., see [12]), it seems that some modifications are necessary in order to apply some existing results in our algorithm. We are now working on the problem.

A final assessment of the algorithm must wait for this work to be fulfilled. However, the experiments have verified the theory and indicated that the algorithm presented in this paper is a very prospective tool in image reconstruction. In the most general case, this reconstruction will require either an enormous storage capacity (the matrix $L = A^TDA$ has n^2 elements, where n is the number of pixels in the image), or a substantial amount of CPU time (if all elements of L have to be computed every time they are used). However, in many practical applications, simplifications are possible and the reconstruction can be done at a reasonable cost in CPU time without requiring excessive storage capacity.

Appendix

Theorem 1: A unique solution curve $f(\lambda; \mu)$ is defined in E for $0 \le \lambda < \infty$ by the following Cauchy problem of differential equations:

$$\begin{cases} \nabla^2 J \cdot \dfrac{df}{d\lambda} = \nabla Q, \\ f|_{\lambda=0} = \exp(\mu - 1)h. \end{cases} \tag{A-1}$$

Proof: Since the coefficient matrix $\nabla^2 J$ is negative definite and both $\nabla^2 J$ and ∇Q are analytic in E, (A-1) does define a unique solution curve $f(\lambda; \mu)$ in E for $0 \le \lambda < \alpha$, where α is some positive number or plus infinity. Thus, all we need now is to prove that the solution curve could be extended whenever $\alpha < \infty$.

Suppose $\alpha < \infty$. We are to prove that first of all the solution curve $f(\lambda; \mu)$ tends to a point $f^* \epsilon E$ as λ approaches α from below, and the limit point f^* satisfies the stationary point equation (7), and then the solution curve $f(\lambda; \mu)$ is extended beyond α by the solution curve of the following Cauchy problem:

$$\begin{cases} \nabla^2 J \cdot \dfrac{df}{d\lambda} = \nabla Q, \\ f|_{\lambda=0} = f^*. \end{cases} \tag{A-2}$$

Along the solution curve $f(\lambda; \mu)$, we have

$$\nabla J(f(\lambda; \mu); \mu, \lambda) \equiv 0, \quad 0 \le \lambda < \alpha. \tag{A-3}$$

Premultiplying (A-3) with $f^T(\lambda; \mu)$ yields

$$\begin{aligned} f^T \nabla J \equiv & -\sum_i f_i \log f_i + (\mu - 1) \sum_i f_i \\ & - \lambda f^T (A^T DA) f + \lambda f^T A^T Dd \equiv 0, \\ & 0 \le \lambda < \alpha, \end{aligned} \tag{A-4}$$

where, for abbreviation, the arguments in $f(\lambda; \mu)$ are dropped. Equation (A-4) indicates that the norm $\|f(\lambda; \mu)\|$ must be bounded as $\lambda \to \alpha - 0$. Otherwise, $f^T \nabla J$ would go to minus infinity because of $f^T(A^TDA)f \ge 0$ and $-\Sigma f_i \log f_i + (\mu - 1) \Sigma f_i + \lambda f^T A^T Dd \to -\infty$. The boundedness $\|f(\lambda; \mu)\|$ assures that each component $f_i(\lambda; \mu)$ has a positive lower bound as $\lambda \to \alpha - 0$. Otherwise, the corresponding component of ∇J would go to ∞, as easily seen from (7). Therefore, not only the entire solution curve, but also its limit point(s) as $\lambda \to \alpha - 0$ belong to E. Now it is clear that each limit point will satisfy (7), and hence, coincide with that unique maximal point of J at α. If we denote the unique limit point by f^*, then (A-2) does define a unique solution curve in E for some small interval $(\alpha - \delta, \alpha + \delta)$ with $\alpha - \delta \ge 0$ which coincides with the maximal point path of J as $\lambda \epsilon (\alpha - \delta, \alpha + \delta)$. Hence, the solution curve given by (A-2) coincides with $f(\lambda; \mu)$ as $\lambda \epsilon (\alpha - \delta, \alpha)$ and extends $f(\lambda; \mu)$ beyond α. Q.E.D.

Acknowledgment

We are very grateful to the reviewer's valuable comments.

References

[1] J. P. Burg, "Maximum entropy spectral analysis" (thesis), in *Proc. 37th Meet. Soc. Exploration Geophysicists*, 1967, Stanford, CA, 1975.

[2] S. F. Gull and B. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, p. 686, 1978.

[3] S. F. Gull, "The maximum entropy algorithm applied to image enhancement," *Proc. IEEE*, vol. 5, p. 170, 1980.

[4] E. T. Jaynes, "On the rationale of maximum entropy methods," *Proc. IEEE*, vol. 70, Sept. 1982.
[5] S. Kay and S. L. Marple, Jr., in *Rec. IEEE ICASSP*, 1979, pp. 151–154.
[6] J. H. McClellan, "Multidimensional spectral estimation," *Proc. IEEE*, vol. 80, pp. 1029–1039, Sept. 1982.
[7] A. Papoulis, "Maximum entropy and spectral estimation: A review," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1176–1186, Dec. 1981.
[8] S. F. Burch, S. F. Gull, and J. Skilling, "Image restoration by a powerful maximum entropy method," *Comput. Vis., Graph., Image Processing*, vol. 23, pp. 113–128, 1983.
[9] S. J. Wernecke and L. R. D'Addario, "Maximum entropy image reconstruction," *IEEE Trans. Comput.*, vol. C-26, Apr. 1977.
[10] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. II*, 2nd ed. New York: Wiley, 1971.
[11] J. N. Franklin, *Matrix Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1968.
[12] G. Dahlquist and A. Björck, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

Xinhua Zhuang graduated from the Peking University, Peking, China, in 1963, after a four-year undergraduate program and a two-year graduate program in mathematics.

Before 1983 he served as a Senior Research Engineer at the Computing Technique Institute, Hangzhou, China. He was a Visiting Scholar of Electrical Engineering at the Virginia Polytechnic Institute and State University, Blacksburg, from 1983 to 1984; a Visiting Scientist of Electrical and Computer Engineering at the University of Michigan, Ann Arbor, granted by Machine Vision International, Ann Arbor, MI, from 1984 to 1985. He was selected as a Consultant at the Advisory Group for Aerospace Research and Development, NATO, in 1985. Currently he is a Visiting Research Professor of the Coordinated Science Laboratory and a Visiting Professor of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. His professional interests lie in applied mathematics, image processing, computer and robotic vision, artificial intelligence, and computer architecture. He is a contributor (with E. Østevold and R. M. Haralick) to the book *Image Recovery: Theory and Application* (H. Stark, Ed.); a contributor (with R. M. Haralick and H. Joo) and editor (with R. M. Haralick) of the book *Consistent Labeling Problems in Pattern Recognition*; and an author (with T. S. Huang and R. M. Haralick) of the planned book entitled *Image Time Sequence Motion Analysis*.

Einar Østevold was born in Bergen, Norway, in 1952. He received the Master's degree in physics from the Norwegian Institute of Technology, Trondheim, in 1975.

Since 1977 he has worked in the field of image processing at the Norwegian Defense Research Establishment (NDRE). Most of his work has been related to image segmentation and object classification. Beginning in August 1983 he spent one year as a Visiting Scientist at the Spatial Data Analysis Lab at the Virginia Polytechnic Institute and State University, Blacksburg, where he did research on maximum entropy techniques in image restoration. His interests include image sequence analysis and target tracking.

Mr. Østevold is a member of the Norwegian Society for Image Analysis and Pattern Recognition (NOBIM).

Robert M. Haralick (S'62–S'67–M'69–SM'76–F'84) was born in Brooklyn, NY, on September 30, 1943. He received the B.A. degree in mathematics from the University of Kansas in 1964, the B.S. degree in electrical engineering in 1966, the M.S. degree in electrical engineering in 1967, and the Ph.D. degree in 1969.

He has worked with Autonetics and IBM. In 1965 he worked for the Center for Research, University of Kansas, as a Research Engineer, and in 1969, when he completed the Ph.D. degree at the University of Kansas, he joined the Faculty of the Department of Electrical Engineering there where he last served as a Professor from 1975 to 1978. In 1979 he joined the Faculty of the Department of Electrical Engineering at Virginia Polytechnic Institute and State University, Blacksburg, where he was a Professor and Director of the Spatial Data Analysis Laboratory. From 1984 to 1986 he was the Vice President of Research at Machine Vision International, Ann Arbor, MI. He now occupies the Boeing chaired Professorship in the Department of Electrical Engineering at the University of Washington, Seattle. He has done research in pattern recognition, multiimage processing, remote sensing, texture analysis, data compression, clustering, artificial intelligence, and general systems theory, and has published over 180 papers. He is responsible for the development of GIPSY (General Image Processing System), a multiimage processing package which runs on a minicomputer system.

Dr. Haralick is a member of the Association for Computer Machinery, Sigma Xi, the Pattern Recognition Society, and the Society for General Systems Research.