A Neural Net Algorithm for Maximum Entropy Image Reconstruction

Xinhua Zhuang

Electrical and Computer Engineering
University of Missouri-Columbia
Columbia, Missouri 65211
U.S.A.

Robert M. Haralick

Department of Electrical Engineering, FT-10
University of Washington
Seattle, Washington 98195
U.S.A.

Abstract

This paper theoretically solves an open problem of whether or not there exists a solution to the maximum entropy image reconstruction problem. It also proves the convergence of the previously proposed maximum entropy image reconstruction (MEIR) algorithm (see Zhuang, Ostrov, and Haralick, 1987) and derives its fast convergence.

All in all, we will eventually find out that what the MEIR algorithm accomplishes is exactly that which a neural net algorithm does when governed by an energy function, which measures the degree of constraint satisfaction.

1. Introduction

This paper represents a continuation of the work reported in an earlier published paper (see Zhuang et. al., 1987). The first contribution is to prove the existence of the solution to the maximum entropy image reconstruction (MEIR) problem, an open problem (see Wu, 1988). The second one is to show the fast convergence of the MEIR algorithm proposed in Zhuang et. al. The third one is to prove that the MEIR algorithm is equivalent to a neural net algorithm whose governing energy function measures the degree of constraint satisfaction.

The paper is organized as follows. In Section 2, we briefly reformulate the MEIR problem for the reader's convenience. In Section 3, we prove the existence and uniqueness of the solution to the MEIR problem. In the same section, we establish the equivalence between the MEIR algorithm and a neural net algorithm. The governing energy function will be shown to be monotonically decreasing and fastly approaching zero. In Section 4, we give both digital implementation and a possible block scheme for electronic implementation of the neural net algorithm. The final section contains our conclusions.

As is clear from the derivations below, the problem statement, the theoretical results, and the neural net reconstruction algorithm all can be easily shifted to the multidimensional signal reconstruction case without any essential changes.

2. MEIR Problem Statement

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

\[
H(p_1, ..., p_n) = - \sum p_i \log(p_i), \quad \text{with } p_i = \frac{f_i}{\sum_i f_i}
\]  

is defined. The entropy depends only on the distribution of gray levels in the image and not on the total intensity

\[
\sum_i f_i.
\]

Let the observed image data be given by

\[
d_j - L(f) + e_j, \quad \text{with } L(f) = \sum_i A_{ij} f_i, \quad j=1, ..., m.
\]

where the \( e_j \)'s represent independent zero mean, \( \sigma_j^2 \) variance noise terms. We assume \( \sigma_j^2 \) is known. We define

\[
Q(f_1, ..., f_n) = \frac{1}{2m} \sum_j (L(f) - d_j)^2 = \frac{1}{\sigma_j^2}
\]

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

\[
Q(f_1, ..., f_n) = \frac{1}{2}
\]

The motivation for this constraint comes about from the central limit theorem (Feller, 1971) which states that with probability one.
\[
\lim_{m} \frac{1}{m} \sum_{i} \frac{e_i^2}{\sigma_i^2} = 1.
\]

Thus, provided \(m\) is large, we would expect the true value of \(f_1, ..., f_n\) to satisfy (4). The condition (4) now determines the set of feasible images each of which satisfies the given statistical test for consistency with the actual image data \((d_1, ..., d_n)\).

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. Formally, we maximize the entropy \(H(p_1, ..., p_n)\) given the constraints (4) and \(\Sigma f_i = F\). As argued in Zhuang et. al., there are two reasons for introducing the second constraint. 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 \(f\); the second reason is concerned with the computational efficiency. By introducing the second constraint, we obtain a linear relation between \(H(f_1, ..., f_n)\) and \(H(p_1, ..., p_n)\), i.e.,

\[
H(p) = \frac{H(0)}{\Sigma f} - \log \Sigma f,
\]

enabling us to treat \(H(f_1, ..., f_n)\) which is more tractable than \(H(p_1, ..., p_n)\). In any case, this constraint is not very strict, and it may be varied to obtain any required total intensity.

From the relation between \(H(p)\) and \(H(f)\), it is easily seen that the following three problems are equivalent to each other.

**Problem 1:**

\[
\begin{align*}
\{ & \text{max } H(p_1, ..., p_n) \\
& \text{subject to } Q = \frac{1}{2} \Sigma f_i = F
\end{align*}
\]

**Problem 2:**

\[
\begin{align*}
\{ & \text{max } H(f_1, ..., f_n) \\
& \text{subject to } Q = \frac{1}{2} \Sigma f_i = F
\end{align*}
\]

**Problem 3:**

\[
\begin{align*}
\{ & \text{max } J(f; \mu, t) = [H(f_1, ..., f_n) + \mu \Sigma f_i Q(f_1, ..., f_n)] \\
& \text{subject to } Q = \frac{1}{2} \Sigma f_i = F
\end{align*}
\]

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

**Problem 4:**

\[
\begin{align*}
\{ & \text{max } J(f; \mu, t) \\
& \text{subject to } O(f) = \frac{1}{2}
\end{align*}
\]

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 problem. It is essentially the same as used by Gull and Daniell, 1978, and Gull, 1980. If \(f_1, ..., f_n\) is a solution of problem 4, then it is also a solution of problems 1, 2, and 3 with \(F = \Sigma f\). Usually, we are satisfied with the solution of problem 4. If not, we could adjust the parameter \(\mu\) so that the solution of problem 4 has the required total intensity. As usual, we first consider the unconstrained maximization problem related to problem 4.

**Problem 5:**

\[
\{ \text{max } J(f; \mu, t) \}
\]

Problem 5 is to find maximal points of \(J(f; \mu, t)\). The function \(J(f; \mu, t)\) is defined and continuous in the closed domain \(\{f : f_i \geq 0, i = 1, ..., n\}\). When \(t \geq 0\), 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, \(\nabla J\):

\[
\nabla^2 J = -\text{diag} \left[ \frac{1}{f_1}, ..., \frac{1}{f_n} \right] - t \nabla^2 Q < 0
\]

(Notice that \(\nabla^2 Q \geq 0\))

It was proved that in Zhuang et. al. that for each fixed \(\mu\) and each fixed \(t \geq 0\), the function \(J(f; \mu, t)\) reaches its unique maximal point (denoted as \(f(t; \mu)\)) internally, i.e., \(f_i(t; \mu) > 0, i = 1, ..., n\), and \(f(t; \mu)\) \((t \geq 0)\) coincides with the solution curve determined by the stationary point equation, i.e., \(\nabla J(f(t; \mu, t)) = 0 \quad (t \geq 0)\), or equivalently by the following initial value problem of differential equations:

\[
\begin{align*}
\frac{df}{dt} - \nabla Q, & \quad t > 0 \\
\quad f(t; \mu), & \quad \exp(\mu-1)(1, ..., 1)_{1 \times n}
\end{align*}
\]

This initial value problem comprises the basis for our MEIR algorithm.

It was also proved that there are only two possibilities in general. Either problem 4 has at most two equientropy solutions or for any \(\mu\) and \(t \geq 0\) along the solution curve defined by (7) there hold
Without loss of generality, we always assume (8) in the remaining part of the paper.

3. Existence, Uniqueness, and Convergence

To prove the existence and uniqueness of solution to the MEIR problem formulated as (5), we will first show that, along the solution curve \( f(t; \mu) (t \geq 0) \) defined by (7), \( Q(f(t; \mu)) \) approaches zero as \( t \to \infty \). In fact, we can prove even more, that is

\[
Q(f(t; \mu)) = O\left(\frac{1}{t}\right)
\]  
(9)

To simplify the proof, we make use of the assumption that the set \( S = \{ f : Q(f) = 0 \} \) with each \( f \) being positive \( \} \) is nonempty. Let \( f^0 \) be an arbitrary point in \( S \). By the definition of \( f(t; \mu) \), it follows,

\[
J(f(t; \mu); \mu, \nu)
\geq J(f^0; \mu, \nu)
\geq -\sum f_i \log f_i + \mu \sum f_i^0 - tQ(f^0)
\geq -\sum f_i \log f_i + \mu \sum f_i^0 \text{ since } Q(f^0) = 0
\]

\[
\text{const. independent of } t
\]  
(10)

Thus,

\[-\sum f_i(t; \mu) \log f_i(t; \mu) + \mu \sum f_i(t; \mu)
\geq tQ(f(t; \mu)) + \text{const.}
\geq \text{const.}
\geq -\infty
\]  
(11)

Since the left side of (11) tends to \( -\infty \) as some \( f_i \to \infty \), we conclude that there must exist \( 0 < B \leq \infty \) so that for each \( i, i = 1, \ldots, n \) and \( f_i(t; \mu) \geq \),

\[0 < f_i(t; \mu) \leq B
\]  
(12)

Finally, we obtain

\[Q(f) \leq \frac{1}{t} \left\{ \mu \sum f_i \sum f_i \log f_i - \text{const.} \right\} \text{ by (10)}
\leq \frac{1}{t} O(1) \text{ by (12)}
\]  
(13)

which straightforwardly leads to (9).

Secondly, we point out that, along the solution curve \( f(t; \mu) \), the constraint function \( Q(f(t; \mu)) \) is strictly monotonically decreasing. This is because

\[\forall f^0 \leq 0, \forall Q > 0 \text{ by (8)}
\]

\[
\frac{dQ(f(t; \mu))}{dt} = (\nabla Q)(\nabla Q)^{-1}(\nabla Q) < 0
\]  
(14)

Thus, along the solution curve \( f(t; \mu) (t \geq 0) \), the constraint function \( Q(f(t; \mu)) \) will strictly monotonically decrease from \( Q(f(0; \mu)) \) to zero as just as the time reciprocal. Because of assumption (8), i.e., \( Q(f(t; \mu)) > 1/2 \), there must exist a unique time \( t^* \) at which \( Q(f(t; \mu)) \) reaches 1/2. In other words, problem 4 does have a unique solution that will be found by tracing along the one-dimensional path \( f(t; \mu) \) until reaching \( t^* \).

Now we are entitled to say that the initial value problem of differential equations defines a neural net algorithm whose governing energy function is characterized by the constraint function \( Q \).

Under the control of the energy function, the artificial neural net will evolve towards the global state \( f(t^*; \mu) \) that solves the MEIR problem.

The above results are true with an arbitrary choice of \( \mu \). In Zhaih et al., the \( \mu \) value which minimizes \( Q(f(0; \mu)) \) was chosen in order to speed up the convergence.

4. Neural Net Algorithm Implementation

There are a number of ways to implement the neural net algorithm or to solve the initial value problem, i.e., (7). One possible digital implementation of (7) that was proposed and experimentally verified in Zhaih et al. (1987) is as follows,

\[
\begin{cases}
  f^0 - \exp(\mu-1)(t_i - 1)_{1 \leq i \leq n} \\
  k \geq 0; \\
  \tau_f f^k(t^{k+1} - f^k) - (t_{k+1} - t_k)\nabla Q^k
\end{cases}
\]  
(15)

where \( t_{k+1} > t_k \), \( \tau_f^k \) is \( \tau_f \) valued at \( f^k \), and \( \nabla Q^k \) is \( \nabla Q \) valued at \( f^k \), respectively. When all \( (t_{k+1} - t_k) \) are small enough, \( f^k \) will approximate \( f(t^*; \mu) \) very well. Equation (15) can be rewritten as follows,

\[
\begin{cases}
  f^0 - \exp(\mu-1)(t_i - 1)_{1 \leq i \leq n} \\
  k \geq 0; \\
  \tau_f^k f^k(t^{k+1} - f^k) - (t_{k+1} - t_k)\nabla Q^k + \tau_f^k f^k
\end{cases}
\]  
(16)

This is a large linear system of equations. Fortunately, the coefficient matrix \( \tau_f^k \) is negative definite and symmetric. The Gauss-Seidel iterative scheme (Franklin 1968) is used to solve (16) efficiently. As argued in Zhaih et al. (1987), only one iteration for each \( k \) is enough to obtain \( f^{k+1} \) from \( f^k \) while using Gauss-Seidel iterative scheme. Let \( P \) represent the matrix \( \tau_f^k \). Let \( b \) represent the vector \( (t_{k+1} - t_k) \).
Further, let \( x \) be the \((k+1)\)th estimate \( f^{k+1} \) of \( f(k+1; \mu) \). Then, the digital implementation of the neural net algorithm calculates \( f^{k+1} \) in the following way,

\[
f_t^{k+1} = \frac{1}{P_t} \left[ b_t - \sum_{j=1}^{P_t} P_t f^{k+1}_j - \sum_{j=1}^{P_t} P_t f^{k}_j \right], \quad t = 1, \ldots, n \tag{17}
\]

Although the neural net algorithm can be implemented in discretized way, it can also be implemented as an electronic neural net algorithm. To see it clearly, we write down the initial value problem in an equivalent form as follows,

\[
\begin{align*}
\frac{df}{dt} &= -\text{diag}(f), \quad f_n \frac{d}{dt} \left( \nabla Q(f) \right) \\
f(0; \mu) &= \exp(\mu - 1)(1, \ldots, 1)_{1 \times n}
\end{align*}
\tag{18}
\]

Notice that

\[
Q = \frac{1}{2m} \sum_{j=1}^{m} \left( L_j(f) - d_j \right)^2 \\
= \frac{1}{2m} (Af - d)^t \text{diag} \left[ \frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_m^2} \right] (Af - d) \\
\Delta Q = \frac{1}{m} A^t \text{diag} \left[ \frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_m^2} \right] A \\
\Delta^2 Q = \frac{1}{m} A^t \text{diag} \left[ \frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_m^2} \right] A
\tag{19}
\]

where \( A = A_{(\mu, \text{max})} \), \( d = (d_1, \ldots, d_m) \). Equation (18) can then be rewritten as follows,

\[
\begin{align*}
\frac{df}{dt} &= -\text{diag}(f), \quad f_n \frac{d}{dt} \left( \nabla Q(f) \right) \\
&= -\frac{1}{m} \text{diag}(f), \quad f_n A^t \text{diag} \left[ \frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_m^2} \right] d \\
&= -\frac{1}{m} \text{diag}(f), \quad f_n A^t \text{diag} \left[ \frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_m^2} \right] A \frac{d}{dt} f \\
f(0; \mu) &= \exp(\mu - 1)(1, \ldots, 1)_{1 \times n}
\end{align*}
\tag{19}
\]

A possible block scheme to implement (19) is shown in Figure 1.

5. Conclusion

In this paper, we proved the existence and uniqueness of the solution to the MEIR problem and derived the fast convergence of the previously proposed MEIR algorithm that was finally proved a neural net algorithm. We also proposed a possible block scheme for electronic implementation of the neural net algorithm. Including robustness in the algorithmic development within the general maximum entropy image reconstruction framework is a problem we are currently working on. For detailed exposition of the maximum entropy principle in image recovery and related references please refer to Stark (1987).

References


