# A STUDY OF OPTIMISATION APPROACHES TO PROBABILISTIC RELAXATION LABELLING ON A 3 NODE 2 LABEL PROBLEM.

**J. Kittler, J. Illingworth and V. Mallesh**

Department of Electronics and Electrical Engineering
University of Surrey, Guildford, GU2 5XH. U.K.

### Abstract

We present a comparative study of several optimisation approaches used in probabilistic relaxation labelling. For test purposes, we propose a 3 node 2 label variant of Rosenfelds famous triangle labelling problem. Experiments comparing Rosenfelds original probability updating rule with the theoretically better justified projected gradient method of Hummel et al show that the latter is to preferred.

### 1. Introduction

One of the fundamental tasks in image understanding is to label (name) the constituent entities of an image. At the lowest levels of image processing these entities may be the image pixels and labels of interest may be for instance *edge* and *non edge* while at higher levels of representation we may be dealing with groups of segments or geometric objects which we wish to assign into plausible object categories such as *house, car etc.* The primary source of information for making decisions about labels are measurements that can be extracted for each entity or object from the image. However due to corruption by noise and other image degradations, the labelling based purely on object properties captured by the measurements is likely to be ambiguous and may be inconsistent.

Relaxation techniques are a class of iterative labelling algorithms which in addition to object properties take into account any constraining relationships that must hold between individual image objects. A number of researchers have demonstrated that the use of this source of information, referred to as *image context*, considerably improves labelling performance.

In this paper we concentrate on labelling algorithms which belong to the family of probabilistic relaxation schemes. At each iteration of the relaxation process the distribution of label probabilities over the label set associated with each image object is updated. For each object the redistribution of the total probability mass depends on the support lent for the individual labels by its context providing neighbourhood.

The probabilistic relaxation algorithm introduced in the seminal paper of Rosenfeld (Rosenfeld Hummel and Zucker 1976) and its variants (Kittler and Illingworth 1985) are heuristic on at least two counts; first, in the manner that the contextual support of neighbouring objects is quantified and second, in the way that the label probability distributions are updated. Recent research has attempted to replace any heuristic elements with results derived from a proper mathematical formulation of the labelling problem. For instance, it has been demonstrated that probabilistic relaxation can be cast as an optimisation problem. Any probability updating scheme may be considered as a specific optimum seeking algorithm aiming to find a solution to the labelling problem.

Despite this unifying viewpoint, various probability relaxation algorithms do exhibit different properties. The purpose of this paper is to compare different schemes experimentally with the following criteria in mind:

- rate of convergence
- computational complexity
- global versus local nature of computation
- bias.

Similar studies have already been reported in the literature but their scope has been limited to the simplest type of the support ( arithmetic average) which leads to objective functions of a relatively simple form. In the present study a more complex evidence combining formula developed in a proper statistical framework (Kittler 1986) is used as a support function. Criteria of labelling consistency that might be optimised then involve high order product terms which substantially affect the nature of the optimisation problem and therefore the behaviour of updating algorithms.

The second departure from previous studies is in the use of computer simulated data which allows an objective rather than a subjective assessment of the algorithm bias. For this comparative study we have tried to select test problems complex enough to fully exercise the

AVC 1987 doi:10.5244/C.1.42

structure of the evidence combining formula yet simple enough to remain amenable to formal analysis.

The paper is organised as follows. In Section 2 the test problem is introduced and the corresponding relaxation labelling task formulated. In Section 3 the probability updating schemes to be experimentally studied will be reviewed. Section 4 constitutes an analysis of the three node two label problem in which the possible solution points reachable by the optimisation process are evaluated. Experimental results obtained are presented in Section 5 while Section 6 is devoted to their discussion.

## 2. Problem Definition

Let us represent the system of image objects by a network of nodes $a_i, i = 1, ..., N$ where a branch between a pair of nodes signifies a node interaction. Each node $a_i$ assumes a unique label $\theta_i$ from a label set $\Lambda = \{\lambda_1, ..., \lambda_m\}$. At each stage a probabilistic relaxation algorithm attempts to quantify which of the labels in $\Lambda$ is most appropriate for node $a_i, i = 1, ..., N$ by associating a probability value $P_i(\theta_i = \lambda_j)$ with the $j^{th}$ label alternative. The initial probability assignments are based on object properties extracted from the image.

A generally accepted criterion of labelling is the criterion of average local consistency $F$ defined as

$$F = \sum_{i=1}^{N} \sum_{j=1}^{m} P_i(\theta_i = \lambda_j) Q_{ij} \tag{1}$$

where $Q_{ij}$ is the contextual support for label $\lambda_j$ at object $a_i$ provided by the interacting objects.

The test problem on which the properties of various probabilistic relaxation algorithms which optimise function (1) will be studied is a network of three nodes each having an identical label set $\Lambda$. The network is illustrated in figure 1. We will consider the simplest conceivable label set consisting of only two labels $\Lambda = \{1, 2\}$. This can be regarded as a special case of the famous four label triangle problem introduced by Rosenfeld (Rosenfeld, Hummel and Zucker 1976).

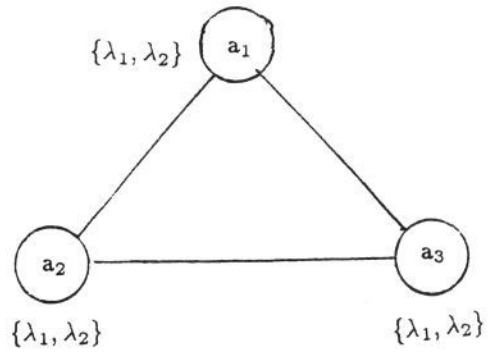


Figure 1.
3 node 2 label network

The contextual support will be computed using the evidence combining formula advocated in (Kittler 1986). Accordingly

$$Q_{ij} = \sum_{\substack{\theta_1 \\ 1 \neq i}} \sum_{\substack{\theta_2 \\ 2 \neq i}} \sum_{\substack{\theta_3 \\ 3 \neq i}} \{ \prod_{\substack{l=1 \\ l \neq i}}^{3} \frac{P_l(\theta_l)}{P(\theta_l)} \} \frac{P(\theta_l, \forall l \neq i, \theta_i = \lambda_j)}{P(\theta_i = \lambda_j)} \tag{2}$$

where $P(\theta_l)$ is the a priori probability that node $a_l$ has label $\theta_l$ and $P(\theta_1, \theta_2, \theta_3)$ is the joint probability of occurrence of a particular combination of node labels in the network. The joint probabilities constitute our world model and their meaning corresponds to the compatibility coefficients of the original Rosenfeld, Hummel and Zucker relaxation labelling scheme. Naturally these probabilities must satisfy

$$\sum_{\theta_1} \sum_{\theta_2} \sum_{\theta_3} P(\theta_1, \theta_2, \theta_3) = 1 \tag{3}$$

and

$$P(\theta_i) = \sum_{\theta_l} P(\theta_i = \lambda_j, \theta_l, \forall l \neq i) \tag{4}$$

## 3. Probability updating methods

A key element of any probabilistic relaxation scheme is the probability updating method. This defines how estimates of label probabilities $P_i(\theta_i = \lambda_j)$ are affected by the constraints expressed within the contextual support functions, $Q_{ij}(\theta_i = \lambda_j)$.

Within the framework of standard optimisation theory, the probabilistic labelling problem involves searching through the m×N dimensional space composed of all label probability assignments, to find local maxima of the average local consistency function, F, defined in (1). Unfortunately, in anything other than the simplest problems the label probability space becomes so large that it is impractible to search it exhaustively and it is usually too complex to find solutions analytically. It is therefore necessary to adopt iterative guided search schemes to find the optima of the problem functional. The basic idea of these methods is to repeatedly use knowledge or appropriate assumptions about the behaviour of the functional to choose a direction of probability update which will result in an increase of the functional. Suppose estimates of label probabilities at iteration k are represented as a vector of probabilities $P^k$. At iteration k+1 the new probability vector $P^{k+1}$ will be given by

$$P^{k+1} = P^k + \alpha u^k \tag{5}$$

where $u^k$ is an update vector or update direction and $\alpha$, the stepsize, is a scalar which determines how far to

move in update direction $u^k$. Guided search methods consist of choosing $u^k$ and $\alpha$ so that

$$F(P^{k+1}) > F(P^k) \tag{6}$$

subject to the probability normalisation and positivity conditions expressed in (3) and (4). In most situations the choice of both $u^k$ and $\alpha$ is not unique and the selection of a particular algorithm for these may affect both the final local maxima found and the speed of convergence to that maximum. In this paper we will consider, for the 3 node labelling problem previously introduced, two ways of calculating $u^k$ viz

- A projected steepest ascent method presented in (Mohammed, Hummel and Zucker 1983).
- The RHZ update rule: the heuristic rule of (Rosenfeld, Hummel and Zucker. 1976)

The next two subsections will consider each of these update directions in greater detail and is then followed by discussion of a suitable choice of stepsize.

### 3.1 Projected steepest ascent

A choice of update direction which leads to an increase of the functional,F, is the direction of the gradient of the F, $\nabla$F. This is the direction of most rapid local change of the functional and therefore the method which uses it is called the method of steepest ascent. As steepest ascent always chooses a direction in which F increases it is guaranteed to converge to a least a local maximum of F. For our 3 node problems it can be easily shown that the gradient is given by

$$\nabla F(P) = 3Q \tag{7}$$

However the direction of steepest ascent cannot be used directly for the update direction because it will lead to violation of probability normalisation conditions. The condition $\sum_j P_i(\theta_i = \lambda_j)$=1 means that updates must lie in the diagonal hyperplane of each subspace defined by the label probabilities of a single object node. This condition can be imposed by projecting the gradient direction, $\nabla$F, into these hyperplanes using a projection operator, $\mathcal{P}$, defined by

$$\mathcal{P} = I - \frac{nn^T}{|n|^2} \tag{8}$$

where I is the unit diagonal matrix and the vector $n^T$=[1,...,1] defines the diagonal constraining hyperplanes.

A further refinement of the projected gradient method is necessary when a component of the P vector i.e. $P_i(\theta_i = \lambda_j)$ becomes zero because projection of the full gradient may then lead to an infeasible negative update. In the algorithm of Hummel et al this is solved by discarding the gradient component corresponding to the infeasible update and then projecting the reduced gradient into the reduced subspace. This process is repeated until all proposed updates are feasible or all components have been removed.

### 3.2 The RHZ update rule

In their original probabilistic relaxation labelling paper Rosenfeld, Hummel and Zucker heuristically derived an update rule which when expressed in the same form as (5) yields

$$u_i^k(\theta_i = \lambda_j) =$$

$$\frac{P_i^k(\theta_i = \lambda_j)(Q_{ij}^k(\theta_i = \lambda_j) - \sum_l P_i^k(\theta_i = \lambda_l)Q_{il}^k(\theta_i = \lambda_l))}{\sum_l P_i^k(\theta_i = \lambda_l)Q_{il}^k(\theta_i = \lambda_l)} \tag{9}$$

and stepsize $\alpha = 1$. Although it was not originally motivated within the framework of the optimisation approach it has been shown by Lloyd (Lloyd 1983) that this $u^k$ is a good update direction for the average local consistency criterion if $\nabla$F is equal to or a multiple of the total support Q. This result can be easily demonstrated by showing that $\nabla F.u^k \geq 0$. Lloyd showed this was the case for a total support function calculated as the arithmetic average of pairwise supports but it is equally true for the network of figure 1 and support function of equation (2).

### 3.3 Stepsize calculation

Once a suitable update direction $u^k$ has been chosen it is necessary to decide how far to move in this direction to achieve maximal increase in F. In the case of our 3 node problem the change in F as a function of distance ,$\phi$, away from the current labelling can be calculated. This yields a complicated cubic function of $\phi$. The maxima of this function, $\phi'$, can be found by solving the quadratic which results from equating its derivatives to zero. The value of $\alpha$ must also always be chosen so that updating never results in generating negative probabilities. Therefore the largest value which $\alpha$ can assume is a value d given by

$$d = min\left\{ \frac{-P_i(\theta_i = \lambda_j)}{u_i(\theta_i = \lambda_j)} \;\middle|\; u_i(\theta_i = \lambda_j) < 0 \right\} \quad \forall i, \forall j \tag{10}$$

In summary, the optimal value for $\alpha$ is given by

$$\alpha = min(\phi', d) \tag{11}$$

## 4. Network Analysis

Before presenting the experimental results we shall analyse the 3-node 2-label network with the aim of identifying all the possible solution points that may be

reached by iterative optimisation. Let us denote by $r_{ijk}$ the following ratio of label probabilities

$$r_{ijk} = \frac{P(\theta_1 = i, \theta_2 = j, \theta_3 = k)}{P(\theta_1 = i)P(\theta_2 = j)P(\theta_3 = k)} \tag{12}$$

For the sake of simplicity we shall also introduce a shorthand notation

$$\begin{aligned} p_i &= P_1(\theta_1 = i) \\ t_j &= P_2(\theta_2 = j) \\ s_k &= P_3(\theta_3 = k) \end{aligned} \tag{13}$$

Substituting from (12) and (13) into (1) and (2) and noting that

$$\begin{aligned} p_1 &= 1 - p_2 \\ t_1 &= 1 - t_2 \\ s_1 &= 1 - s_2 \end{aligned} \tag{14}$$

we can express the average local consistency function as

$$\begin{aligned} F = 3\{&p_1[t_1 s_1 r_{111} + t_1(1-s_1)r_{112} \\ &+(1-t_1)s_1 r_{121} + (1-t_1)(1-s_1)r_{122}]+ \\ &+(1-p_1)[t_1 s_1 r_{211} + t_1(1-s_1)r_{212} \\ &+(1-t_1)s_1 r_{221} + (1-t_1)(1-s_1)r_{222}]\} \end{aligned} \tag{15}$$

Combining the corresponding terms the objective function can be rewritten as

$$\begin{aligned} F = 3\{&p_1 t_1 s_1 k_1 + p_1 t_1 k_2 + p_1 s_1 k_3 + t_1 s_1 k_4 \\ &+p_1 k_5 + t_1 k_6 + s_1 k_7 + k_8\} \end{aligned} \tag{16}$$

where

$$\begin{aligned} k_1 &= r_{111} - r_{112} - r_{121} + r_{122} - r_{211} + r_{212} + r_{221} - r_{222} \\ k_2 &= r_{112} - r_{122} - r_{212} + r_{222} \\ k_3 &= r_{121} - r_{122} - r_{221} + r_{222} \\ k_4 &= r_{211} - r_{212} - r_{221} + r_{222} \\ k_5 &= r_{122} - r_{222} \\ k_6 &= r_{212} - r_{222} \\ k_7 &= r_{221} - r_{222} \\ k_8 &= r_{222} \end{aligned} \tag{17}$$

Differentiating (16) with respect to $p_1, t_1$ and $s_1$ and setting the first partial derivatives to zero, the conditions for stationary points of $F$ are found to be

$$\begin{aligned} t_1 s_1 k_1 + t_1 k_2 + s_1 k_3 + k_5 &= 0 \\ p_1 s_1 k_1 + p_1 k_2 + s_1 k_4 + k_6 &= 0 \\ p_1 t_1 k_1 + p_1 k_3 + t_1 k_4 + k_7 &= 0 \end{aligned} \tag{18}$$

The system of equations has two solutions but each corresponds to a saddle point. This follows from the fact that the second term of the Taylor expansion of function $F$ at these points does not have a constant sign in the vicinity of the stationary points. Any maxima or minima of $F$ must therefore lie on the boundary of the feasible region. As the feasible region in the present case is a 3-D cube illustrated in figure 2, the maxima and minima will lie on the faces of the cube.

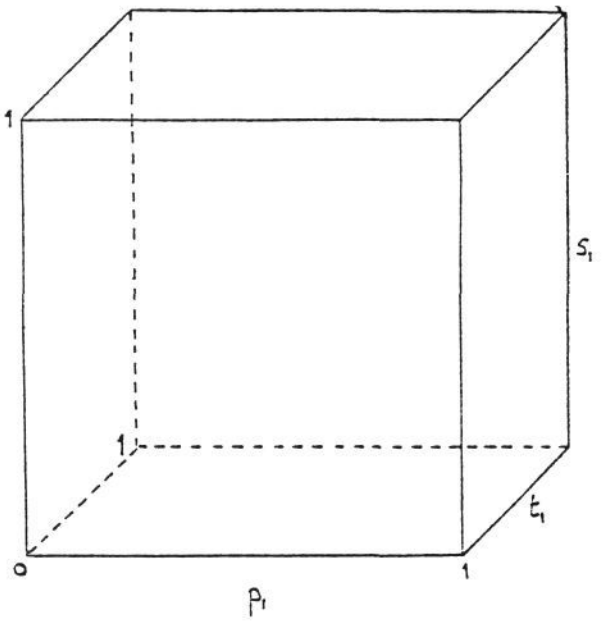


Figure 2.
Feasible labelling region.

Analysing the properties of the objective function at the respective faces in a similar manner, it transpires that each face contains one stationary point which is also a saddle point. For instance, for $t_1 = 0$ function $F$ reduces to

$$F = 3[p_1 s_1 k_3 + p_1 k_5 + s_1 k_7 + k_8] \tag{19}$$

with the saddle point at

$$p_1 = \frac{-k_7}{k_3}$$

and

$$s_1 = \frac{-k_5}{k_3} \tag{20}$$

It thus follows that maxima and minima must lie on the cube edges. It is then a simple matter to show that the optima must coincide with some of the cube vertices.

Starting from an arbitrary point in the feasible region any algorithm optimising function $F$ in (16) should reach one of the following solutions:

- saddle point defined by (18)

- saddle point on one of the faces of the cube
- local optimum at one of the cube vertices.

## 4. Experimental results

For our experimental investigations we propose a reduced version of the famous triangle labelling problem introduced in the original paper of Rosenfeld et al. This problem requires a 3D interpretation to be given to a 2D image in which 3 straight lines form a triangle. Each node of the labelling network corresponds to a line in the image and in the original statement of the problem each line can have one of 4 labels arising from 4 distinct physical situations i.e. two surfaces meet to form either a concave, $-$, or a convex, $+$, dihedral angle or one surface occludes a second surface, $\rightarrow$, or, $\leftarrow$. If the lines are to represent a planar triangular object in relation to a transparent background plane then only 8 from the $4^3$=64 possible unambiguous labelling combinations are physically realisable. These labellings form a world model and deductions from experimental measurements must be compatible with this world model.

The triangle problem naturally reduces to a two label problem when it is noted that the 8 physically possible labellings group into two equal size but disjoint sets. One set involves only the $\rightarrow$ and $-$ labels while the second set of labellings contains only $\leftarrow$ and $+$ labels. Thus for our reduced triangle problem we consider a label set $\Lambda = \{\rightarrow,-\}$ and a model in which 4 unambiguous labelling combinations are physically possible. This yields prior joint probabilities of :

$$P(\rightarrow,\rightarrow,\rightarrow) = P(\rightarrow,\rightarrow,-) = P(-,\rightarrow,\rightarrow) = P(\rightarrow,-,\rightarrow) = \frac{1}{4}$$
$$P(-,-,-) = P(-,-,\rightarrow) = P(\rightarrow,-,-) = P(\rightarrow,-,-) = 0 \quad (21)$$

and, for all nodes, prior singlet probabilities of

$$P(\rightarrow) = \frac{3}{4} \quad P(-) = \frac{1}{4} \quad (22)$$

The only non zero $r_{ijk}$ are

$$r_{\rightarrow\rightarrow\rightarrow} = \frac{16}{27}, \quad r_{\rightarrow\rightarrow-} = r_{-\rightarrow\rightarrow} = r_{\rightarrow-\rightarrow} = \frac{48}{27} \quad (23)$$

and the functional F reduces to

$$F = 3\{p_1t_1s_1k_1 + p_1t_1k_2 + p_1s_1k_3 + t_1s_1k_4\} \quad (24)$$

with $k_1 = r_{\rightarrow\rightarrow\rightarrow} - 3r_{\rightarrow\rightarrow-} = \frac{-128}{27}$, $k_2 = r_{\rightarrow\rightarrow-} = \frac{48}{27}$, $k_3 = r_{\rightarrow-\rightarrow} = \frac{48}{27}$, $k_4 = r_{-\rightarrow\rightarrow} = \frac{48}{27}$.

From these values and the analysis of Section 4 we can find solutions for all classes of stationary point. Firstly, the solution to equations corresponding to (18) yields global saddle points at $(p_1,t_1,s_1)$ values of $(\frac{3}{4},\frac{3}{4},\frac{3}{4})$ and (0,0,0). The first of these saddle points corresponds to the no information point of the relaxation labelling process i.e. these probabilities are the same as prior expectations. Secondly, analysis of the function F on faces of the cubic, feasible probability region shows that saddle points exist on the top, rightside and backside faces. These constrained saddle points are at $(p_1,t_1,s_1)$ values of (0.6,0.6,1.0), (1.0,0.6,0.6) and (0.6,1.0,0.6) respectively. Finally the function values and labellings at the vertices of the probability cube are given by :

| $(p_1,t_1,s_1)$ | labelling | F |
|---|---|---|
| (0,0,0) | $(-,-,-)$ | 0 |
| (0,0,1) | $(-,-,\rightarrow)$ | 0 |
| (0,1,0) | $(-,\rightarrow,-)$ | 0 |
| (0,1,1) | $(-,\rightarrow,\rightarrow)$ | $\frac{48}{9}$ |
| (1,0,0) | $(\rightarrow,-,-)$ | 0 |
| (1,0,1) | $(\rightarrow,-,\rightarrow)$ | $\frac{48}{9}$ |
| (1,1,0) | $(\rightarrow,\rightarrow,-)$ | $\frac{48}{9}$ |
| (1,1,1) | $(\rightarrow,\rightarrow,\rightarrow)$ | $\frac{16}{9}$ |

The three vertices with $F = \frac{48}{9}$ are local maxima and vertex (0,0,0) is a global stationary point.

Table 1 shows results obtained with three optimisation schemes:

- RHZ1: the RHZ update direction with unit stepsize.
- RHZ2: the RHZ update direction with an optimal stepsize.
- HZM : the projected gradient algorithm of Hummel et al with an optimal stepsize.

Each row of Table 1 contains the initial values of $(p_1,t_1,s_1)$, the same values once the relaxation process has converged, values of the local average consistency, F, and the number, $N_c$, of iterations to achieve convergence. Table 1 illustrates results as a function of different starting positions, $(p_1,t_1,s_1)$, within the feasible labelling cube. Cases 1 to 4 are examples of the relaxation process when it is started well within the volume of the labelling cube. It is seen that different initial probabilities lead to convergence to different stationary points. The stationary point found is not always an unambiguous labelling point i.e. case 1 converges to the global stationary point in the cube interior while case 3 moves to a constrained saddle point on one of the cube faces. Hummel and Zucker argue that when deep inside the cube the RHZ update direction should approximate the HZM projected gradient direction and in cases 1 to 4 it is seen that from the same initial point the three methods converge to the same stationary point. Figure 3 illustrates a typical case, in the $(p_1,t_1)$ plane, of the path taken by the three different relaxation methods. One of the most striking features of cases 1 to 4 is that

| Case # | Initial $(p_1, t_1, s_1)$ | RHZ1 | RHZ2 | HZM |
|---|---|---|---|---|
| 1 | 0.50 0.50 0.50<br>F=2.22 | 0.75 0.75 0.75<br>F=3.00 $N_c$=6 | 0.75 0.75 0.75<br>F=3.00 $N_c$=2 | 0.75 0.75 0.75<br>F=3.00 $N_c$=2 |
| 2 | 0.50 0.40 0.50<br>F=2.04 | 1.00 0.00 1.00<br>F=5.33 $N_c$=17 | 1.00 0.00 1.00<br>F=5.33 $N_c$=4 | 1.00 0.00 1.00<br>F=5.33 $N_c$=4 |
| 3 | 0.30 0.30 0.50<br>F=1.44 | 0.60 0.60 1.00<br>F=3.20 $N_c$=20 | 0.60 0.60 1.00<br>F=3.20 $N_c$=4 | 0.60 0.60 1.00<br>F=3.20 $N_c$=4 |
| 4 | 0.20 0.30 0.50<br>F=1.23 | 0.00 1.00 1.00<br>F=5.33 $N_c$=15 | 0.00 1.00 1.00<br>F=5.33 $N_c$=5 | 0.00 1.00 1.00<br>F=5.33 $N_c$=5 |
| 5 | 0.00 0.00 0.90<br>F=0.00 | 0.00 0.00 0.90<br>F=0.00 $N_c$=1 | 0.00 0.00 0.90<br>F=0.00 $N_c$=1 | 0.60 0.60 1.00<br>F=3.20 $N_c$=1 |
| 6 | 0.00 0.00 0.10<br>F=0.00 | 0.00 0.00 0.10<br>F=0.00 $N_c$=1 | 0.00 0.00 0.10<br>F=0.00 $N_c$=1 | 1.00 1.00 0.00<br>F=5.3 $N_c$=2 |
| 7 | 0.01 0.01 0.10<br>F=0.01 | 0.60 0.60 1.00<br>F=3.20 $N_c$=15 | 0.60 0.60 1.00<br>F=3.20 $N_c$=3 | 1.00 1.00 0.00<br>F=5.33 $N_c$=3 |
| 8 | 0.75 0.75 0.75<br>F=3.00 | 0.75 0.75 0.75<br>F=3.00 $N_c$=1 | 0.75 0.75 0.75<br>F=3.00 $N_c$=1 | 0.75 0.75 0.75<br>F=3.00 $N_c$=1 |
| 9 | 0.00 0.00 0.00<br>F=0.00 | 0.00 0.00 0.00<br>F=0.00 $N_c$=1 | 0.00 0.00 0.00<br>F=0.00 $N_c$=1 | 0.00 0.00 0.00<br>F=0.00 $N_c$=1 |
| 10 | 1.00 1.00 1.00<br>F=1.78 | 1.00 1.00 1.00<br>F=1.78 $N_c$=1 | 1.00 1.00 1.00<br>F=1.78 $N_c$=1 | 0.75 0.75 0.75<br>F=3.00 $N_c$=1 |

Table 1. Comparison of optimisation methods: RHZ1, RHZ2 and HZM.
F is the value of local average consistency
$N_c$ is the number of iterations to convergence

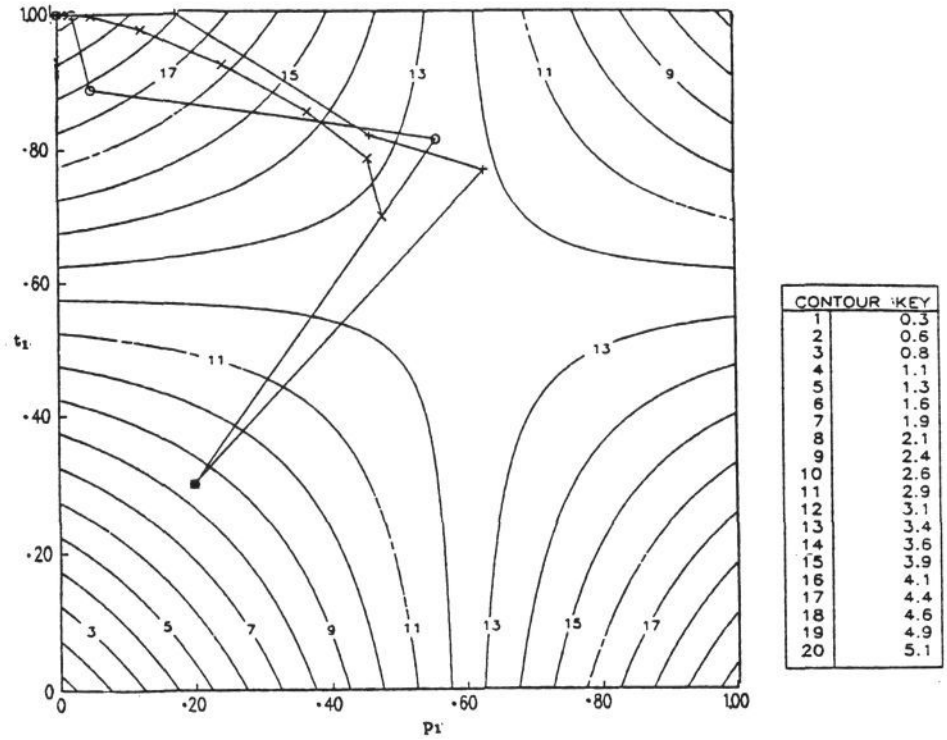


Figure 3. Contour plot of average local consistency function on top face of labelling cube indicating the convergence of the RHZ1 ×, RHZ2 + and HZM ∘ methods from initial point (0.2,0.3,0.5).

RHZ2 and HZM converge at about the same rate but the number of iterations needed for RHZ1 is 3-5 times greater.

Cases 5 to 7 illustrate the behaviour of the method when the initial probabilities lie close to or actually on one of the cube edges. If the edge is part of a face which contains a maximum or saddle point then the relaxation process will converge to one of these points. However the 3 edges emanating from the (0,0,0) vertex are not part of such faces. Cases 5 and 6 show two important things. Firstly, the RHZ update method is unable to increase a probability component that has become zero. In cases 5 and 6 the RHZ relaxation processes do not move from the initial probability points even though these points are not stationary points of the average local consistency function. Secondly, the HZM process does converge to a valid stationary point but the convergence point is a function of where along the edge the process is started. For the edge connecting the (0,0,0) and (0,0,1) vertices the HZM process converged to either the saddle point on the top face or the (1,1,0) vertex. The former point was reached only when $s_1$ exceeded the prior probability $P(\rightarrow)$. Case 7 shows the effect of moving away from the edge by a small amount. The RHZ processes are no longer tied to the edge and they do now converge to a stationary point of the local consistency function. However it is seen that the point of convergence is not necessarily the same as that found by the HZM process.

Cases 8 and 9 show the effect of initiating the relaxation process from the global saddle points. As expected, all three relaxation processes remain at these points.

Finally case 10 illustrates the effect of starting at the constrained minimum vertex (1,1,1). Once again we see the inability of the RHZ processes to move from zero probability points, in this case $p_2$,$t_2$ and $s_2$ are zero. However the HZM method does move away from the minimum and finds the more consistent global saddle point in the interior of the labelling cube.

The comparative times for each of the processes was determined by allowing the relaxation calculation to proceed at the stationary points even after convergence checking had indicated that it had reached these points. The procedures were coded in Fortran on a serial machine, a $\mu$VAX-2 computer, and times for 1000 iterations of the RHZ1, RHZ2 and HZM processes were 9.6, 19.8 and 22.1 seconds respectively. In a serial implementation the calculations consist of 3 principle components : time $\tau_u$ to loop over each node and its neighbours to calculate an update direction, time $\tau_\alpha$ to loop over each node and its neighbours to calculate an optimal stepsize and time $\tau_n$ to update the probability value of each pixel in turn. As each node has 2 neighbours the times $\tau_u$ and $\tau_\alpha$ should dominate over $\tau_n$ and the calculation time should be the sum of these two approximately equal components. Inspection of the RHZ1 and RHZ2 timings confirms this. The RHZ2 method is only about 10% faster than the HZM projected gradient method. Although the calculation of an optimal stepsize doubles the computation time per iteration of the relaxation process the RHZ2 and HZM methods should still be quicker than the RHZ1 method as they require only $\frac{1}{3}$ to $\frac{1}{5}$ the number of iterations. These conclusions will of course be dramatically altered if a parallel architecture appropriate for local neighbourhood calculations is used. In this case the dominating time will become the global calculation of $\tau_\alpha$. However schemes for hierarchical networks do exist which could well be used to overcome this bottleneck.

## 6. Discussion and Conclusions

In this section we shall consider the results reported in Section 5 in the light of the criteria set out in the introduction. Some of these, such as rate of convergence and computational complexity can readily be translated into quantitative measures. The notion of bias however needs further elaboration.

As predicted by the analysis outlined in Section 4 the average local consistency function for the triangle problem has several saddle points and local maxima. In the first instance an algorithm must be deemed to be biased if it does not reach one of these stationary points. However, reaching just any one of these stationary points cannot be considered as a satisfactory solution to the optimisation problem. Each stationary point is a reflection of particular evidence conveyed by the prior world knowledge and initial label probabilities based on measurements. Vice versa a piece of evidence cannot be mapped onto more than one of the potential solutions.

By the same token we must dismiss as a desirable solution the global maximum of $F$, if any, unless the supporting evidence justifies it. Such an objective would lead to the same solution regardless of the initial probability assignment which would be inconsistent with the role various information sources clearly play in image interpretation. It follows that in any situation the optimal solution must relate to the amount of evidence available. It is therefore reasonable to identify this solution as the stationary point that would be reached from the initial probability assignment by a steepest ascent search method.

With the optimum solution specified, we can now return to the experimental results of Section 5. By definition, the solutions obtained by the HZM method are unbiased. The results of the RHZ algorithms are consistent with the HZM method in the majority of cases. However, there are initial conditions for which the RHZ

algorithm will not even reach a stationary point (see Table 1, cases 5 and 6). Typically this may happen when the initial probability assignment for some labels corresponds to a hard labelling (zero - one probabilities). Whenever the initial probabilities are nonzero, the RHZ algorithm appears to find a stationary point but this may differ from the HZM solution as seen in case 7 of Table 1.The RHZ algorithm may therefore lead to a biased solution.

It is interesting to note that the hard labelling obtained by maximum selection would lead to physically meaningful objects. This observation is quite encouraging as it suggests that the combination of probabilistic relaxation followed by maximum selection operation could yield a globally consistent labelling. In contrast maximum selection applied to the initial label probability assignment would frequently result in physically impossible objects. Probabilistic relaxation appears then to achieve labelling consistency where it did not exist before or reinforce it in cases where objects would be labelled consistently based just on measurements.

Note also that the global saddle point corresponds to the *no information* point. Hence if the initial label probabilities are equal to the a priori probabilities as in case 8 of Table 1, no updating takes place. This confirms the unbiasedness property of the support function (2) demonstrated in (Kittler 1986).

Whenever the optimal step size is used regardless of whether in conjunction with the RHZ or HZM scheme the rates of convergence are high. The computation of the optimal step size however requires an additional effort. The overall computational complexity of an algorithm is therefore best expressed as a product of the number of iterations and the computational time per iteration. The results reported in Section 5 suggest that for a serial implementation the optimal step calculation is well worth the extra effort required.

The main drawback of step size calculation is that it converts intrinsically local algorithms amenable to parallel implementation to global ones. Parallel processing is particularly important for low level vision where relaxation labelling algorithms may operate on individual pixels in the image. However the global computation of step size can be carried out on a parallel architecture quite efficiently if it has a pyramid structure.

In summary, the RHZ relaxation process has been shown to provide biased results when optimising average local consistency. The HZM method is to be preferred to the RHZ methods as it is theoretically better justified and seems, at least on the current problem, to be no more computationally demanding than the RHZ methods.

## References


[1] Rosenfeld A., Hummel R.A. and Zucker S.W., "Scene labeling by relaxation operations." IEEE Trans. Systems Man and Cybernetics. 6:6 (1976) 420-433.

[2] Kittler J. and Illingworth J., "Relaxation labelling algorithms - a review." Image and Vision Computing. 3:4 (1985) 206-216.

[3] Lloyd S.A., "An optimisation approach to relaxation labeling algorithms." Image and Vision Computing. 1:2 (1983) 85-91.

[4] Hummel R.A. and Zucker S.W., "On the foundations of relaxation labeling processes." IEEE Trans. Pattern Analysis and Machine Intelligence. 5:3 (1983) 267-287.

[5] Mohammed J.L., Hummel R.A. and Zucker S.W., "A gradient projection algorithm for relaxation methods." IEEE Trans. Pattern Analysis and Machine Intelligence. 5:3 (1983) 330-332.

[6] Kittler J., "Compatibility and support functions in probabilistic relaxation." In Proc. $8^{th}$ International Joint Conference on Pattern Recognition. Paris. (1986) 186-189.