

Analysis of Relaxation Processes: The Two-Node Two-Label Case

DIANNE P. O'LEARY AND SHMUEL PELEG

Abstract—Several relaxation processes are analyzed in the simple case of two nodes, each having two possible labels. It is shown that the choice of coefficients is very important. For certain values of the coefficients, some processes will have a single nontrivial convergence point regardless of the initial labeling. For other choices of the coefficients, there can be more than one possible convergence point, and different solutions can be obtained for different initial labelings. In the probabilistic approach where the coefficients are predefined in terms of joint probabilities, there are always two nontrivial convergence points for all possible coefficients. The results are also compared to the Bayesian analysis that can be obtained in this simple case of two nodes. Since certain selections of coefficients can give unacceptable results even in this simple case, it can be expected that the proper selection of coefficients will be much more important in the general case involving larger numbers of nodes and labels.

Manuscript received January 15, 1982; revised February 23, 1983. This work was supported by the National Science Foundation under Grant MCS-76-23763.

D. P. O'Leary is with the Department of Computer Science, University of Maryland, College Park, MD 20742.

S. Peleg was with the Computer Vision Laboratory, Computer Science Center, University of Maryland, College Park, MD 20742. He is now with the Department of Computer Science, The Hebrew University of Jerusalem, 91904, Jerusalem, Israel.

I. INTRODUCTION

RELAXATION processes are used for the reduction of ambiguity in problems that can be represented as graph labeling. This representation associates with each object a node whose label represents the class of the object. Labeling a node corresponds to classification of the object. Initially, every node has probability estimates for all possible labels. Relaxation processes use the probabilities at neighboring nodes to improve these estimates. Using only neighboring nodes reduces the global problem to many local problems. Solving the local problems can often give an acceptable global solution.

Let $G = (V, E)$ be a graph with $V = \{1, \dots, N\}$ its set of nodes and E its set of arcs, and let $\Lambda = \{\lambda_1, \dots, \lambda_L\}$ be a set of labels. We indicate by $P_i^{(k)}(\lambda)$ the probability estimate at the k th iteration that node i should be labeled λ . The true distribution P_i is unknown, but based on some measurements, a probability distribution $P_i^{(0)}: \Lambda \rightarrow [0, 1]$ is estimated for every node i . A relaxation operator \mathcal{R} produces an improved probability estimate from the previous

one:

$$P^{(k+1)} = \mathcal{R}(P^{(k)}, R),$$

where R is a set of predetermined coefficients. The coefficient $r_{ij}(\lambda, \lambda')$ represents the compatibility of node i being labeled λ and node j being labeled λ' .

The first updating method, called nonlinear relaxation, was suggested by Rosenfeld *et al.* [1] and was based on heuristic considerations. Peleg [2] derived an updating scheme using a probabilistic argument. Faugeras and Berthod [3], [4] suggested the optimization of a criterion on the probabilistic labelings using a maximum descent algorithm. Hummel and Zucker [5] have recently developed a measure of average local consistency, also to be optimized.

The above methods are analyzed in the following sections for their behavior in the simple case of two nodes having two possible labels. In this case $V = \{1, 2\}$, $\Lambda = \{\alpha, \beta\}$; p_α and p_β will denote the probability estimates for the first node and q_α and q_β the estimates for the second node. $\text{Pr}_{ij}(\alpha, \beta)$ is the *a priori* joint probability of nodes i and j being labeled α and β , respectively, $\text{Pr}_i(\alpha)$ is the *a priori* probability of node i being labeled α , and $\text{Pr}_j(\beta)$ is the *a priori* probability of node j being labeled β . For abbreviation we denote $\text{Pr}_1(\alpha)$ by \hat{p}_α , $\text{Pr}_2(\alpha)$ by \hat{q}_α , and $\text{Pr}_{12}(\alpha, \alpha)$ by $\hat{p}_{\alpha\alpha}$.

Bayesian analysis leads to the labeling (λ, δ) of nodes $(1, 2)$, respectively, when $A_{\lambda\delta} = p_\lambda^{(0)} q_\delta^{(0)} \text{Pr}_{12}(\lambda, \delta)$ is larger than the corresponding A value for any other possible labeling. This rule can be illustrated by a two-dimensional diagram in the $p_\alpha - q_\alpha$ plane, where each point in the unit square corresponds to a possible labeling. The Bayesian rule partitions the $p_\alpha - q_\alpha$ square into four regions. Typical cases are illustrated in Figs. 1 and 2. For any initial labeling $(p_\alpha^{(0)}, q_\alpha^{(0)})$, the most likely labeling is in the outside corner of its region. Points on the boundaries of regions correspond to ties in the A values.

Let

$$a_1 = \frac{\text{Pr}_{12}(\beta, \alpha)}{\hat{q}_\alpha}$$

$$a_2 = \frac{\text{Pr}_{12}(\beta, \beta)}{1 - \hat{q}_\alpha}$$

$$b_1 = \frac{\text{Pr}_{12}(\alpha, \beta)}{\hat{p}_\alpha}$$

$$b_2 = \frac{\text{Pr}_{12}(\beta, \beta)}{1 - \hat{p}_\alpha}$$

Then if $\hat{p}_\alpha \hat{q}_\alpha > \hat{p}_{\alpha\alpha}$, the regions are defined by the rectangle with corners $(0, 0)$ and (a_2, b_2) , the rectangle with corners $(1, 1)$ and (a_1, b_1) , and the remaining region partitioned by the curve

$$q_\alpha = \frac{p_\alpha \text{Pr}_{12}(\alpha, \beta)}{p_\alpha \text{Pr}_{12}(\alpha, \beta) + (1 - p_\alpha) \text{Pr}_{12}(\beta, \alpha)}$$

If $\hat{p}_\alpha \hat{q}_\alpha < \hat{p}_{\alpha\alpha}$, then the regions are the rectangle with corners $(0, 1)$ and (a_1, b_2) , the rectangle with corners $(1, 0)$ and (a_2, b_1) , and the remaining region partitioned by the

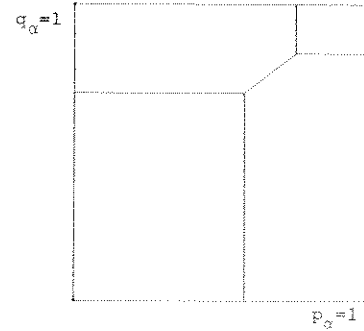


Fig. 1. Bayesian labeling when $\hat{p}_\alpha \hat{q}_\alpha > \hat{p}_{\alpha\alpha}$. This diagram was created for $\hat{p}_\alpha = 3/8$, $\hat{q}_\alpha = 2/8$, $\hat{p}_{\alpha\alpha} = 1/16$.

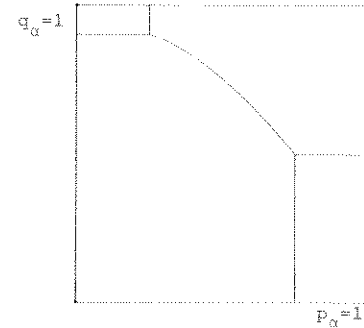


Fig. 2. Bayesian labeling when $\hat{p}_\alpha \hat{q}_\alpha < \hat{p}_{\alpha\alpha}$. This diagram was created for $\hat{p}_\alpha = 3/8$, $\hat{q}_\alpha = 2/8$, $\hat{p}_{\alpha\alpha} = 3/16$.

curve

$$q_\alpha = \frac{(1 - p_\alpha) \text{Pr}_{12}(\beta, \beta)}{(1 - p_\alpha) \text{Pr}_{12}(\beta, \beta) + p_\alpha \text{Pr}_{12}(\alpha, \alpha)}$$

If $\hat{p}_\alpha \hat{q}_\alpha = \hat{p}_{\alpha\alpha}$, then $a_1 = a_2$ and $b_1 = b_2$, and the regions are four rectangles formed by partitioning the square at the point (a_1, b_1) .

Relaxation methods give results rather different from the Bayesian result. At most two corners can be reached by relaxation, and in some cases one corner will be the only convergence point regardless of the initial labeling. This observation is discussed in the following sections.

II. PROBABILISTIC RELAXATION

In this section the probabilistic relaxation developed by Peleg [2] is analyzed. The updating of node i based only on the neighbor j is

$$P_{ij}^{(k+1)}(\alpha) = \frac{P_i^{(k)}(\alpha) \sum_{\beta \in \Lambda} P_j^{(k)}(\beta) r_{ij}(\alpha, \beta)}{\sum_{\lambda \in \Lambda} P_i^{(k)}(\lambda) \sum_{\beta \in \Lambda} P_j^{(k)}(\beta) r_{ij}(\lambda, \beta)}$$

One way to combine all estimates from all neighbors is by averaging their effects:

$$P_i^{(k+1)}(\alpha) = \frac{1}{n} \sum_{(i,j) \in E} P_{ij}^{(k+1)}(\alpha),$$

where n is the number of neighbors of node i and $(i, j) \in E$ indicates that i and j are neighbors. The coefficients

$r_{ij}(\alpha, \beta)$ are defined as

$$r_{ij}(\alpha, \beta) = \frac{\text{Pr}_{ij}(\alpha, \beta)}{\text{Pr}_i(\alpha) \text{Pr}_j(\beta)},$$

where $\text{Pr}_{ij}(\alpha, \beta)$ is the *a priori* joint probability of nodes i and j being labeled α and β , respectively, $\text{Pr}_i(\alpha)$ is the *a priori* probability of node i being labeled α , and $\text{Pr}_j(\beta)$ is the *a priori* probability of node j being labeled β .

To simplify the notation for the two-node case, we write the four coefficients as $r_{\alpha\alpha}, r_{\alpha\beta}, r_{\beta\alpha}, r_{\beta\beta}$, where

$$r_{\lambda\delta} = \frac{\text{Pr}_{12}(\lambda, \delta)}{\text{Pr}_1(\lambda) \text{Pr}_2(\delta)}.$$

The iteration process can be expressed as

$$p_\alpha^{(k+1)} = \frac{p_\alpha^{(k)}(r_{\alpha\alpha}q_\alpha^{(k)} + r_{\alpha\beta}q_\beta^{(k)})}{p_\alpha^{(k)}(r_{\alpha\alpha}q_\alpha^{(k)} + r_{\alpha\beta}q_\beta^{(k)}) + p_\beta^{(k)}(r_{\beta\alpha}q_\alpha^{(k)} + r_{\beta\beta}q_\beta^{(k)})}$$

$$p_\beta^{(k+1)} = 1 - p_\alpha^{(k+1)}$$

$$q_\alpha^{(k+1)} = \frac{q_\alpha^{(k)}(r_{\alpha\alpha}p_\alpha^{(k)} + r_{\beta\alpha}p_\beta^{(k)})}{q_\alpha^{(k)}(r_{\alpha\alpha}p_\alpha^{(k)} + r_{\beta\alpha}p_\beta^{(k)}) + q_\beta^{(k)}(r_{\alpha\beta}p_\alpha^{(k)} + r_{\beta\beta}p_\beta^{(k)})}$$

$$q_\beta^{(k+1)} = 1 - q_\alpha^{(k+1)}.$$

The compatibility coefficients have several constraints on them:

$$\text{Pr}_1(\alpha) + \text{Pr}_1(\beta) = 1$$

$$\text{Pr}_2(\alpha) + \text{Pr}_2(\beta) = 1$$

$$\text{Pr}_{12}(\alpha, \alpha) + \text{Pr}_{12}(\alpha, \beta) = \text{Pr}_1(\alpha)$$

$$\text{Pr}_{12}(\beta, \alpha) + \text{Pr}_{12}(\beta, \beta) = \text{Pr}_1(\beta)$$

$$\text{Pr}_{12}(\alpha, \alpha) + \text{Pr}_{12}(\beta, \alpha) = \text{Pr}_2(\alpha)$$

$$\text{Pr}_{12}(\alpha, \beta) + \text{Pr}_{12}(\beta, \beta) = \text{Pr}_2(\beta),$$

and all probabilities lie between zero and one. It is easy to show that this leaves three degrees of freedom in the compatibilities which, for convenience, we will take as $\text{Pr}_1(\alpha) = \hat{p}_\alpha$, $\text{Pr}_2(\alpha) = \hat{q}_\alpha$, and $\text{Pr}_{12}(\alpha, \alpha) = \hat{p}_{\alpha\alpha}$. Then the compatibilities take the form

$$\begin{pmatrix} r_{\alpha\alpha} & r_{\alpha\beta} \\ r_{\beta\alpha} & r_{\beta\beta} \end{pmatrix} = \begin{pmatrix} \frac{\hat{p}_{\alpha\alpha}}{\hat{p}_\alpha \hat{q}_\alpha} & \frac{\hat{p}_\alpha - \hat{p}_{\alpha\alpha}}{\hat{p}_\alpha(1 - \hat{q}_\alpha)} \\ \frac{\hat{q}_\alpha - \hat{p}_{\alpha\alpha}}{(1 - \hat{p}_\alpha)\hat{q}_\alpha} & \frac{\hat{p}_{\alpha\alpha} - \hat{p}_\alpha - \hat{q}_\alpha + 1}{(1 - \hat{p}_\alpha)(1 - \hat{q}_\alpha)} \end{pmatrix}$$

when $0 < \hat{p}_{\alpha\alpha} \leq \hat{p}_\alpha, \hat{q}_\alpha < 1$.

To understand the iteration process fully, let us first consider the fixed points, those points for which

$$p_\alpha^{(k+1)} = p_\alpha^{(k)} \quad q_\alpha^{(k+1)} = q_\alpha^{(k)}.$$

From these conditions, we get the relations

$$\begin{aligned} (p_\alpha^2 - p_\alpha)(r_{\alpha\alpha}q_\alpha + r_{\alpha\beta}(1 - q_\alpha)) \\ = (p_\alpha^2 - p_\alpha)(r_{\beta\alpha}q_\alpha + r_{\beta\beta}(1 - q_\alpha)) \end{aligned}$$

$$\begin{aligned} (q_\alpha^2 - q_\alpha)(r_{\alpha\alpha}p_\alpha + r_{\beta\alpha}(1 - p_\alpha)) \\ = (q_\alpha^2 - q_\alpha)(r_{\alpha\beta}p_\alpha + r_{\beta\beta}(1 - p_\alpha)). \end{aligned}$$

There are several cases.

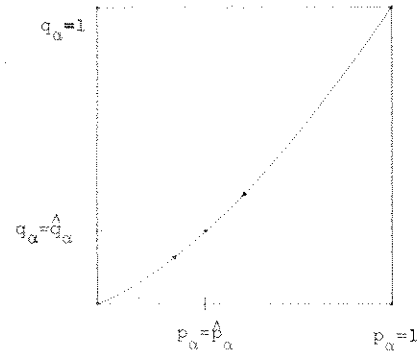


Fig. 3. Convergence of probabilistic relaxation when $\hat{p}_\alpha \hat{q}_\alpha > \hat{p}_{\alpha\alpha}$. This diagram was created for the initial probabilities of Fig. 1, giving

$$R = \begin{pmatrix} 2/3 & 10/9 \\ 6/5 & 28/30 \end{pmatrix}.$$

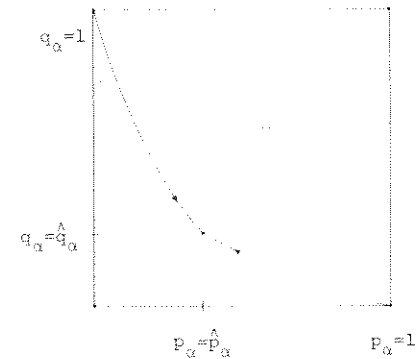


Fig. 4. Convergence of probabilistic relaxation when $\hat{p}_\alpha \hat{q}_\alpha < \hat{p}_{\alpha\alpha}$. This diagram was created for the initial probabilities of Fig. 2, giving

$$R = \begin{pmatrix} 2 & 2/3 \\ 4/10 & 6/5 \end{pmatrix}.$$

a) For any choice of $q_\alpha^{(0)}$, if $p_\alpha^{(0)} = 0$ or $p_\alpha^{(0)} = 1$, then the p components remain fixed.

b) For any choice of $p_\alpha^{(0)}$, if $q_\alpha^{(0)} = 0$ or $q_\alpha^{(0)} = 1$, then the q components remain fixed.

c) If $\hat{p}_{\alpha\alpha} = \hat{p}_\alpha \hat{q}_\alpha$, then

$$\begin{pmatrix} r_{\alpha\alpha} & r_{\alpha\beta} \\ r_{\beta\alpha} & r_{\beta\beta} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and all points are fixed points. (This corresponds to the labelings of the two nodes being independent events.)

d) If $p_\alpha^{(k)} = \hat{p}_\alpha$ and $q_\alpha^{(k)} = \hat{q}_\alpha$ then the p and q components remain fixed. This case corresponds to the "no information" labeling in Peleg [2].

No other possibilities exist. Thus, for the probabilistic labeling there are either five fixed points ($p_\alpha = 1$ or 0 , $q_\alpha = 1$ or 0 ; and $p_\alpha = \hat{p}_\alpha, q_\alpha = \hat{q}_\alpha$) or an infinite number of them (case c).

At other points, the iteration process can be understood qualitatively through the two-dimensional diagram in the $p_\alpha - q_\alpha$ plane indicating the convergence of the process for various starting points. Such diagrams are sketched in Figs. 3 and 4 for the following possible cases.

- 1) $\hat{p}_\alpha \hat{q}_\alpha > \hat{p}_{\alpha\alpha}$. In this case the diagonal elements of the compatibility matrix are dominated by the off-diagonal elements and convergence is either to the (α, β) or (β, α) labeling unless the iteration process is started on a fixed point or on a certain simple curve through $(\hat{p}_\alpha, \hat{q}_\alpha)$. See Fig. 3.
- 2) $\hat{p}_\alpha \hat{q}_\alpha = \hat{p}_{\alpha\alpha}$. Every point is fixed.
- 3) $\hat{p}_\alpha \hat{q}_\alpha < \hat{p}_{\alpha\alpha}$. This is the case converse to 1). See Fig. 4.

The answers produced by the relaxation scheme do not correspond precisely to the answers produced by the use of a Bayesian probability model.

III. NONLINEAR RELAXATION

The nonlinear updating rule was heuristically suggested by Rosenfeld *et al.* [1]. The updating factor for the estimate $P_i^{(k)}(\lambda)$ (at the k th iteration) is

$$q_i^{(k)}(\lambda) = \frac{1}{n} \sum_{(i,j) \in E} \sum_{\lambda' \in \Lambda} \tilde{r}_{ij}(\lambda, \lambda') P_j^{(k)}(\lambda'),$$

where n is the number of neighboring nodes for the node i .

The new estimate of the probability of λ at node i is defined as

$$P_i^{(k+1)}(\lambda) = \frac{P_i^{(k)}(\lambda) [1 + q_i^{(k)}(\lambda)]}{\sum_{\lambda' \in \Lambda} P_i^{(k)}(\lambda') [1 + q_i^{(k)}(\lambda')]}.$$

Thus each $P_i^{(k)}(\lambda)$ is multiplied by $[1 + q_i^{(k)}(\lambda)]$, and the values are then normalized such that $P_i^{(k+1)}$ will be a probability vector. This updating is iterated until some termination criterion is met.

The coefficients \tilde{r}_{ij} used in this scheme are in the range $[-1, 1]$, and

- a) if λ and λ' are compatible for objects a_i and a_j , respectively, then $\tilde{r}_{ij}(\lambda, \lambda') > 0$;
- b) if λ and λ' are incompatible for a_i and a_j , respectively, then $\tilde{r}_{ij}(\lambda, \lambda') < 0$;

- c) if neither labeling is constrained by the other, then $\tilde{r}_{ij}(\lambda, \lambda') = 0$;
- d) the magnitude of \tilde{r}_{ij} represents the strength of the compatibility.

Methods for computing the \tilde{r}_{ij} are suggested in [6]. Some observations on the two-node two-label case in this scheme were made by Pavlidis [7], but our analysis goes somewhat further.

For the case of two nodes and two labels, we have

$$\begin{aligned} p_\alpha^{(k+1)} &= \frac{p_\alpha^{(k)} (1 + \tilde{r}_{\alpha\alpha} q_\alpha^{(k)} + \tilde{r}_{\alpha\beta} q_\beta^{(k)})}{p_\alpha^{(k)} (1 + \tilde{r}_{\alpha\alpha} q_\alpha^{(k)} + \tilde{r}_{\alpha\beta} q_\beta^{(k)}) + p_\beta^{(k)} (1 + \tilde{r}_{\beta\alpha} q_\alpha^{(k)} + \tilde{r}_{\beta\beta} q_\beta^{(k)})} \\ &= \frac{p_\alpha^{(k)} ((1 + \tilde{r}_{\alpha\alpha}) q_\alpha^{(k)} + (1 + \tilde{r}_{\alpha\beta}) q_\beta^{(k)})}{p_\alpha^{(k)} ((1 + \tilde{r}_{\alpha\alpha}) q_\alpha^{(k)} + (1 + \tilde{r}_{\alpha\beta}) q_\beta^{(k)}) + p_\beta^{(k)} ((1 + \tilde{r}_{\beta\alpha}) q_\alpha^{(k)} + (1 + \tilde{r}_{\beta\beta}) q_\beta^{(k)})}. \end{aligned}$$

If we let $r_{\lambda\delta} = (1 + \tilde{r}_{\lambda\delta})/2$, then the iteration process takes the same form as the probabilistic scheme:

$$\begin{aligned} p_\alpha^{(k+1)} &= \frac{p_\alpha^{(k)} (r_{\alpha\alpha} q_\alpha^{(k)} + r_{\alpha\beta} q_\beta^{(k)})}{p_\alpha^{(k)} (r_{\alpha\alpha} q_\alpha^{(k)} + r_{\alpha\beta} q_\beta^{(k)}) + p_\beta^{(k)} (r_{\beta\alpha} q_\alpha^{(k)} + r_{\beta\beta} q_\beta^{(k)})} \\ p_\beta^{(k+1)} &= 1 - p_\alpha^{(k+1)} \\ q_\alpha^{(k+1)} &= \frac{q_\alpha^{(k)} (r_{\alpha\alpha} p_\alpha^{(k)} + r_{\beta\alpha} p_\beta^{(k)})}{q_\alpha^{(k)} (r_{\alpha\alpha} p_\alpha^{(k)} + r_{\beta\alpha} p_\beta^{(k)}) + q_\beta^{(k)} (r_{\alpha\beta} p_\alpha^{(k)} + r_{\beta\beta} p_\beta^{(k)})} \\ q_\beta^{(k+1)} &= 1 - q_\alpha^{(k+1)}. \end{aligned}$$

Here the r values are all positive and have been normalized to range between zero and one. The r coefficients for the probabilistic algorithm for two nodes and two labels can also be normalized in this way by dividing each of them through by a constant, and thus probabilistic relaxation is a special case of this algorithm in which there are further restrictions on the r values. Thus all the cases discussed in the previous section apply to this iteration too. If the point $(\tilde{p}_\alpha, \tilde{q}_\alpha)$,

$$\begin{aligned} \tilde{p}_\alpha &= \frac{r_{\beta\beta} - r_{\beta\alpha}}{r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta}} \\ \tilde{q}_\alpha &= \frac{r_{\beta\beta} - r_{\alpha\beta}}{r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta}} \end{aligned}$$

satisfies $0 \leq \tilde{p}_\alpha \leq 1$ and $0 \leq \tilde{q}_\alpha \leq 1$, then the previous discussion applies (and $\tilde{p}_\alpha = \hat{p}_\alpha, \tilde{q}_\alpha = \hat{q}_\alpha$). If not, then the convergence situation is similar, but no curve exists which partitions the unit square, and all initial noncorner labelings will converge to the same corner. The proof of this can be sketched as follows. Suppose $\tilde{p}_\alpha > 1$. (The cases $\tilde{p}_\alpha < 0, \tilde{q}_\alpha > 1$, and $\tilde{q}_\alpha < 0$ are similar.) Then

$$-1 < \frac{r_{\alpha\alpha} - r_{\alpha\beta}}{r_{\beta\beta} - r_{\beta\alpha}} < 0.$$

Now

Again two cases exist. If $r_{\alpha\alpha} - r_{\alpha\beta} < 0$, then q_α can be seen to be monotonically decreasing. If $r_{\alpha\alpha} - r_{\alpha\beta} > 0$ then q_α increases monotonically. By continuity of the iteration functions, then, the final convergence point is the one to which the initial point $(p, 0)$ in the first case or $(p, 1)$ in the second case converges for $0 < p < 1$.

The convergence into one corner, regardless of the initial labeling, is an undesirable result in applications of relaxation for which both the initial labeling and the coefficients

contain valid information and should affect the final labeling.

IV. OPTIMIZATION APPROACH

For the case of pairwise compatibilities, Faugeras and Berthod [3], [4] define a consistency vector q_i for every node i :

$$q_i(\lambda) = \frac{1}{n_i} \sum_{(i,j) \in E} \sum_{\lambda' \in \Lambda} r_{ij}(\lambda, \lambda') P_j(\lambda'),$$

where n_i is the number of neighbors of node i , and $q_i(\lambda)$ specifies how consistent label λ at node i is with the labelings at the neighboring nodes. Based on this they suggest a criterion to indicate both consistency and ambiguity:

$$\begin{aligned} C &= \sum_{i \in V} P_i \cdot q_i \\ &= \sum_{i \in V} \sum_{\lambda \in \Lambda} P_i(\lambda) q_i(\lambda) \\ &= \sum_{(i,j) \in E} \frac{1}{n_i} \sum_{\lambda \in \Lambda} \sum_{\lambda' \in \Lambda} P_i(\lambda) P_j(\lambda') r_{ij}(\lambda, \lambda'). \end{aligned}$$

The idea is to maximize C over all probability labelings.

Hummel and Zucker [5] derive a support function for a label λ at a node i :

$$S_i(\lambda) = \sum_{(i,j) \in E} \sum_{\lambda' \in \Lambda} r_{ij}(\lambda, \lambda') P_j(\lambda'),$$

and a local consistency measure

$$\begin{aligned} A(P) &= \sum_{i \in V} \sum_{\lambda \in \Lambda} P_i(\lambda) S_i(\lambda) \\ &= \sum_{(i,j) \in E} \sum_{\lambda \in \Lambda} \sum_{\lambda' \in \Lambda} P_i(\lambda) P_j(\lambda') r_{ij}(\lambda, \lambda'). \end{aligned}$$

Thus both optimization approaches optimize an expression of the same form, although the r_{ij} values may differ.

Hummel and Zucker also proved that when the r_{ij} s are symmetric, i.e., $r_{ij}(\lambda, \lambda') = r_{ji}(\lambda', \lambda)$ for all i, j, λ, λ' , if $A(Q)$ attains a local maximum at a point P , then P is a "consistent" labeling, which they define as

$$\sum_{\lambda \in \Lambda} P_i(\lambda) S_i(\lambda) \geq \sum_{\lambda \in \Lambda} Q_i(\lambda) S_i(\lambda), \quad i = 1, 2, \dots, n,$$

for all possible probabilistic labelings Q , where $S_i(\lambda)$ is evaluated using the labeling P . We also adopt symmetric coefficients for our analysis.

Since in the optimization approach r_{ij} is used to indicate neighborhood consistency, we assign

$$r_{ii}(\lambda, \lambda') = 0, \quad \text{for all } i, \lambda, \lambda'.$$

We then get the following expression for the target function in the two-node label case. Let

$$A = \begin{pmatrix} 0 & 0 & r_{\alpha\alpha} & r_{\alpha\beta} \\ 0 & 0 & r_{\beta\alpha} & r_{\beta\beta} \\ r_{\alpha\alpha} & r_{\beta\alpha} & 0 & 0 \\ r_{\alpha\beta} & r_{\beta\beta} & 0 & 0 \end{pmatrix}$$

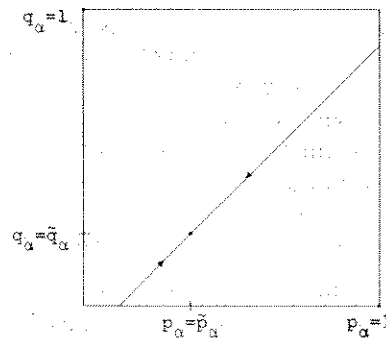


Fig. 5. Convergence of optimization method when $r_{\alpha\alpha} < r_{\alpha\beta}, r_{\beta\alpha}$ and $r_{\beta\beta} < r_{\alpha\beta}, r_{\beta\alpha}$. This diagram was created for same R as in Fig. 3.

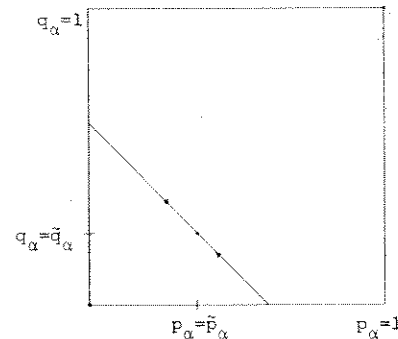


Fig. 6. Convergence of optimization method when $r_{\alpha\alpha} > r_{\alpha\beta}, r_{\beta\alpha}$ and $r_{\beta\beta} > r_{\alpha\beta}, r_{\beta\alpha}$. This diagram was created for same R as in Fig. 4.

and

$$P^T = (p_\alpha, p_\beta, q_\alpha, q_\beta).$$

Then

$$\begin{aligned} C &= A(P) = P^T A P \\ &= 2(p_\alpha(r_{\alpha\alpha}q_\alpha + r_{\alpha\beta}q_\beta) + p_\beta(r_{\beta\alpha}q_\alpha + r_{\beta\beta}q_\beta)) \\ &= 2(p_\alpha(r_{\alpha\alpha}q_\alpha + r_{\alpha\beta}q_\beta) + (1 - p_\alpha)(r_{\beta\alpha}q_\alpha + r_{\beta\beta}q_\beta)). \end{aligned}$$

Consider the behavior of this function for fixed q_α and q_β . Then C is a monotonic function of p_α , increasing with p_α if $r_{\alpha\alpha}q_\alpha + r_{\alpha\beta}q_\beta > r_{\beta\alpha}q_\alpha + r_{\beta\beta}q_\beta$, constant if equality holds and decreasing otherwise. Thus no local maxima or minima can exist for values of p_α lying strictly between zero and one. We obtain a similar argument for q_α by rearranging the expression. Thus any maximum occurs at a point where $p_\alpha = 0$ or 1 and $q_\alpha = 0$ or 1 .

Now C is a quadratic function of p_α and q_α and has a single saddlepoint $(\bar{p}_\alpha, \bar{q}_\alpha)$, found by setting the gradient equal to zero:

$$\begin{aligned} \bar{p}_\alpha &= \frac{r_{\beta\beta} - r_{\beta\alpha}}{r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta}} \\ \bar{q}_\alpha &= \frac{r_{\beta\beta} - r_{\alpha\beta}}{r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta}}. \end{aligned}$$

There are two cases:

1) If $0 \leq \bar{p}_\alpha \leq 1$ and $0 \leq \bar{q}_\alpha \leq 1$ then the function C has a saddle point inside the $p_\alpha - q_\alpha$ square. (Note that this is equivalent to $(r_{\beta\beta} - r_{\alpha\beta})(r_{\alpha\alpha} - r_{\beta\alpha}) \geq 0$ and $(r_{\beta\beta} - r_{\beta\alpha})(r_{\alpha\alpha}$

$-r_{\alpha\beta} \geq 0$.) Two local maxima exist, located at opposite corners of the square. (Similarly, two local minima are at the other two corners.) A pure steepest-ascent algorithm will converge to one of three points: one of the two maximal corners, or the inflection point $(\bar{p}_\alpha, \bar{q}_\alpha)$. The convergence regions are sketched in Figs. 5 and 6. The dividing line between the regions passes through $(\bar{p}_\alpha, \bar{q}_\alpha)$ and makes a 45° angle with the axes.

2) If $(\bar{p}_\alpha, \bar{q}_\alpha)$ does not lie within the unit square, then only one maximum and one minimum exist, and they are located at opposite corners. Thus any ascent algorithm would converge to the maximal corner independently of the initial point.

The comment at the conclusion of the last section regarding the undesirability of a single convergence point also applies here.

The criterion just considered may be modified by adding a so-called entropy term to the function C , giving a new function to be maximized:

$$\tilde{C} = C - \gamma[p_\alpha p_\beta + q_\alpha q_\beta]$$

where γ is a positive number. Let

$$\bar{p}_\alpha = \frac{-\gamma(r_{\beta\beta} - r_{\alpha\beta} + \gamma/2) + (r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta})(r_{\beta\beta} - r_{\beta\alpha} + \gamma/2)}{(r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta})^2 - \gamma^2}$$

$$\bar{q}_\alpha = \frac{-\gamma(r_{\beta\beta} - r_{\beta\alpha} + \gamma/2) + (r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta})(r_{\beta\beta} - r_{\alpha\beta} + \gamma/2)}{(r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta})^2 - \gamma^2}$$

Then the gradient of \tilde{C} is zero at (\bar{p}, \bar{q}) . For $\gamma = 0$, $\tilde{C} = C$, and the discussion above holds. For $\gamma > |r_{\alpha\alpha} - r_{\alpha\beta} - r_{\beta\alpha} + r_{\beta\beta}|$, \tilde{C} has a global minimum at $(\bar{p}_\alpha, \bar{q}_\alpha)$, rather than a saddle point. If this point is within the $p_\alpha - q_\alpha$ square, then each corner is a local maximum and a convergence point for a steepest ascent algorithm. As γ increases to infinity, the point $(\bar{p}_\alpha, \bar{q}_\alpha)$ converges to $(1/2, 1/2)$, and the four quadrants of the $p_\alpha - q_\alpha$ square define the convergence regions.

V. CONCLUDING REMARKS

The analysis of relaxation processes for the two-node two-label case shows that in some cases relaxation can yield a result that is independent of the initial labeling.

This unacceptable behavior results from certain choices of the coefficients.

The two-node two-label analysis can be useful also for understanding relaxation in more general cases and for yielding insight on how pairwise relaxation estimates should be combined in higher dimensions. Every multinode case can be reduced to a two-node two-label case by assuming independence among all nodes and labels except two. Thus, for every general relaxation case, choices of coefficients exist that give rise to single identical solutions for all initial labelings, and it is conjectured that a single solution occurs in cases other than reduction to the two-node case. It is desirable to have guidelines for choosing the coefficients in the nonlinear relaxation algorithm, as is the case in Peleg's probabilistic relaxation, but it is unclear yet whether such algebraic restrictions are necessary or verifiable in practical situations.

ACKNOWLEDGMENT

The help of Sally Atkinson in preparing this paper is gratefully acknowledged.

REFERENCES

- [1] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 420-433, June 1976.
- [2] S. Peleg, "A new probabilistic relaxation scheme," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 362-369, July 1980.
- [3] O. Faugeras and M. Berthod, "Improving consistency and reducing ambiguity in stochastic labeling: An optimization approach," *IEEE Trans. Pattern Anal. Machine Intell.*, in press.
- [4] M. Berthod and O. Faugeras, "Using context in the global recognition of a set of objects: An optimization approach," in *Proc. IFIP*, 1980.
- [5] R. Hummel and S. Zucker, "On the foundations of relaxation labeling processes," Computer Vision and Graphics Lab., McGill University, TR-80-7, July 1980.
- [6] S. Peleg and A. Rosenfeld, "Determining compatibility coefficients for curve enhancement relaxation processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 548-555, July 1978.
- [7] T. Pavlidis, *Structural Pattern Recognition*. New York: Springer, 1977.