Journal of Mathematical Imaging and Vision 7, 309–323 (1997)

The Dynamics of Nonlinear Relaxation Labeling Processes

MARCELLO PELILLO
Dipartimento di Matematica Applicata e Informatica, Università "Ca'Foscari" di Venezia, Via Torino 155, 30173 Venezia Mestre, Italy
pelillo@dsi.unive.it

Abstract. We present some new results which definitively explain the behavior of the classical, heuristic nonlinear relaxation labeling algorithm of Rosenfeld, Hummel, and Zucker in terms of the Hummel-Zucker consistency theory and dynamical systems theory. In particular, it is shown that, when a certain symmetry condition is met, the algorithm possesses a Liapunov function which turns out to be (the negative of) a well-known consistency measure. This follows almost immediately from a powerful result of Baum and Eagon developed in the context of Markov chain theory. Moreover, it is seen that most of the essential dynamical properties of the algorithm are retained when the symmetry restriction is relaxed. These properties are also shown to naturally generalize to higher-order relaxation schemes. Some applications and implications of the presented results are finally outlined.

Keywords: relaxation labeling processes, consistency, growth transformations, Liapunov functions, stability

1. Introduction

In 1976, Rosenfeld, Hummel and Zucker [1] introduced a class of parallel iterative procedures which have now become a standard technique in the pattern recognition and machine vision domains. These algorithms, generally known as *relaxation labeling processes*, attempt to exploit contextual information in order to provide "consistent" solutions in classification (or labeling) problems where noise and uncertainty can affect the accuracy of classical non-contextual pattern recognition algorithms. In their influential paper, Rosenfeld and his co-workers developed a number of relaxation schemes and, after pointing out their intrinsic limitations, eventually introduced a nonlinear algorithm solely on the basis of *ad-hoc* heuristic arguments. Unfortunately, they were not able to relate their model to a precise and satisfactory definition of consistency, and this lack of rigorous justification has always been indicated as one of the most serious drawbacks of the model. Nevertheless, a variety of papers have later reported on the practical usefulness of the algorithm (see, e.g., [2] for a review and an extensive bibliography).

The recognized difficulties with the original heuristic algorithm motivated several investigators to provide formal definitions of consistency and develop alternative relaxation schemes. Essentially, two radically distinct approaches were pursued. On the one hand, some have tried to set the labeling problem within a probabilistic framework, exploiting the well-known tools of Bayesian analysis. The most notable example of this approach is that of Peleg [3], and the more recent works of Kittler and Hancock [4], and Christmas et al. [5] are in the same vein. As noted in [6], however, even if Bayesian analysis provides much insight into the understanding of relaxation labeling algorithms, the approach is capable of accounting for at most one iteration of the process, and to understand its dynamical properties one has necessarily to resort to some approximation. Recently, however, the behavior of these "probabilistic" relaxation labeling procedures has begun to be clarified thanks to the work of Stoddart et al. [7] who have uncovered certain interesting dynamical properties of the Hancock-Kittler and Christmas-Kittler-Petrou probabilistic schemes.

A different standpoint consists of explicitly defining some quantitative measure of consistency to be maximized, and then restating the labeling problem as one of optimization. This appears to be advantageous in that the progress at each iteration can be quantitatively

assessed, and the convergence properties of the algorithm are therefore easily derivable. Early attempts in this direction were made by Ullman [8], and later by Faugeras and Berthod [9]. Illingworth and Kittler [10] provide a relatively recent account of these methods. It is a remarkable fact that most of these developers (of either side) have devoted great care to demonstrate that the original nonlinear relaxation formulas can be regarded as an approximation of their schemes.

Undoubtedly, however, a landmark contribution in the theory of relaxation labeling processes was given by Hummel and Zucker [6] who developed a general theory of consistency that naturally extends the ideas of classical (discrete) constraint satisfaction. Based on this theory, they developed an alternative relaxation procedure which turns out to possess desirable convergence properties under unrestricted circumstances. In addition, like many of their predecessors, they showed that the standard nonlinear scheme *does* approximate their algorithm, and this led them to explain its success in terms of their theory. That of Hummel and Zucker is now generally agreed to be the standard theory of consistency in labeling problems. According to the distinction drawn by Kittler and Hancock [4], the Hummel-Zucker consistency theory aims at providing an *object centered* interpretation, i.e., one in which our attention is confined to obtaining a label assignment to a single object at a time. This contrasts with the *message centered* approach, where we seek instead a joint labeling configuration that best explains the available set of measurements for all objects. The latter standpoint has recently motivated Hancock and Kittler [11] to formulate the labeling problem as one of *maximum a posteriori probability* (MAP) estimation problem, and develop a powerful discrete relaxation algorithm.

Over the past years, a number of investigators have attempted to formally establish a direct relationship between the original relaxation scheme and some well-founded theory of consistency. Lloyd [12], being apparently unaware of the work of Hummel and Zucker, showed that the nonlinear scheme can be regarded as a suboptimal way of maximizing a certain consistency function, which is in fact the one proposed in [6], under some restrictive assumptions concerning the compatibility relations. She showed this by simply rewriting the relaxation formulas so as to show that the algorithm moves towards a gradient direction. The "suboptimality" stems from the fact that the process makes use of a fixed step size (equal to unity). As is well known, gradient methods can ensure an increase in the objective function only when infinitesimal steps are taken, so that Lloyd's analysis does not really provide much insight into the dynamical behavior of the algorithm. An interesting paper about the properties of nonlinear relaxation is one of Elfving and Eklundh [13] who presented, in particular, a local convergence theorem which in our opinion deserves special attention because of its potential use in certain novel applications of the algorithm (see the discussion in Section 8). More recently, the problem of establishing a correspondence between nonlinear relaxation and Hummel and Zucker's theory of consistency has been tackled by Levy [14] but, unfortunately, his analysis is entirely based on very strict conditions upon the compatibility relations which makes it uninteresting for practical applications.

The principal contribution of this paper is to definitively show how, despite its heuristic and simple-minded character, the nonlinear relaxation algorithm is in fact intimately related to the theory of consistency developed by Hummel and Zucker. Specifically, based on a powerful result of use in the theory of probabilistic functions of Markov processes, it is shown that under a certain symmetry restriction Rosenfeld et al.'s relaxation scheme monotonically increases the well-known consistency function derived in [6]. This greatly extends earlier results of Lloyd [12] and Levy [14]. In terms of dynamical systems theory it can be stated therefore that the nonlinear relaxation labeling algorithm possesses a Liapunov (or energy) function which drives the process towards the nearest consistent solution. In addition, it is demonstrated that most of the essential dynamical properties of the algorithm continue to hold even if the symmetry requirement is relaxed, and the model is therefore proven to accomplish something useful under unrestricted circumstances. A generalization to relaxation schemes that incorporate high-order compatibility relations is also discussed. Although these properties are traditionally associated with relaxation labeling and the term "relaxation" is indeed often used as a synonymous with energy-minimization, we stress the fact that in the case of the original heuristic formulas, such an association has been more an "act of faith" than the actual result of theoretical analysis.

This paper is organized as follows. In Section 2 we introduce the nonlinear relaxation algorithm as defined by Rosenfeld et al. [1]. Section 3 briefly reviews the theory of consistency developed by Hummel and Zucker [6], and presents some results that will be helpful in

the subsequent part of the paper. In Section 4 we introduce the so-called Baum-Eagon inequality, and in Section 5 we apply this result to analyze the dynamical behavior of relaxation in the symmetric case. In Section 6, we relax the symmetry restriction and show how the connection with the theory of consistency still holds. Section 7 extends these results to the case of high-order compatibility relations. Section 8 discusses certain implications of the presented results, and then concludes the paper.

2. Nonlinear Relaxation Labeling

The labeling problem involves a set of objects $\mathbf{B} = \{b_1, \ldots, b_n\}$ and a set of possible labels $\mathbf{\Lambda} = \{1, \ldots, m\}$.[1] The purpose is to label each object of $\mathbf{B}$ with one label of $\mathbf{\Lambda}$. To accomplish this, two sources of information are exploited. The first one relies on *local* measurements which capture the salient features of each object viewed in isolation; classical pattern recognition techniques can be practically employed to carry out this task. The second source of information, instead, accounts for possible interactions among nearby labels and, in fact, incorporates all the contextual knowledge about the problem at hand. This is quantitatively expressed by means of a real-valued four-dimensional matrix of compatibility coefficients $R = \{r_{ij}(\lambda, \mu)\}$. The coefficient $r_{ij}(\lambda, \mu)$ measures the strength of compatibility between the hypotheses "b_i has label λ" and "b_j has label μ:" high values correspond to compatibility and low values correspond to incompatibility. In our discussion, the compatibilities are assumed to be nonnegative, i.e., $r_{ij}(\lambda, \mu) \geq 0$, but this seems not to be a severe limitation because, as will be seen later, all the interesting concepts involved here exhibit a certain "linear invariance" property. In this paper, moreover, we will not concern ourselves with the crucial problem of how to derive the compatibility coefficients. Suffice to say that they can be either determined on the basis of statistical grounds [1, 15, 16] or, according to a more recent standpoint, adaptively learned over a sample of training data [17–19].

The initial local measurements are assumed to provide, for each object $b_i \in \mathbf{B}$, an m-dimensional vector $\bar{p}_i^{(0)} = (p_i^{(0)}(1), \ldots, p_i^{(0)}(m))^T$ (where "T" denotes the usual transpose operation), such that $p_i^{(0)}(\lambda) \geq 0$, $i = 1 \ldots n$, $\lambda \in \mathbf{\Lambda}$, and $\sum_\lambda p_i^{(0)}(\lambda) = 1$, $i = 1 \ldots n$. Each $p_i^{(0)}(\lambda)$ can be regarded as the initial, noncontextual degree of confidence of the hypothesis "b_i is labeled with label λ." By simply concatenating $\bar{p}_1^{(0)}, \bar{p}_2^{(0)}, \ldots, \bar{p}_n^{(0)}$ we obtain a weighted labeling assignment for the objects of $\mathbf{B}$ that will be denoted by $\bar{p}^{(0)} \in \mathbb{R}^{nm}$. A relaxation labeling process takes as input the initial labeling assignment $\bar{p}^{(0)}$ and iteratively updates it taking into account the compatibility model R.

At this point, we introduce the space of weighted labeling assignments:

$$\mathbb{K} = \left\{ \bar{p} \in \mathbb{R}^{nm} \mid p_i(\lambda) \geq 0,\ i = 1 \ldots n,\ \lambda \in \mathbf{\Lambda} \text{ and } \sum_{\lambda=1}^{m} p_i(\lambda) = 1,\ i = 1 \ldots n \right\}$$

which is a linear convex set of $\mathbb{R}^{nm}$. Every vertex of $\mathbb{K}$ represents an *unambiguous* labeling assignment, that is one which assigns exactly one label to each object. The set of these labelings will be denoted by $\mathbb{K}^*$:

$$\mathbb{K}^* = \{\bar{p} \in \mathbb{K} \mid p_i(\lambda) = 0 \text{ or } 1,\ i = 1 \ldots n,\ \lambda \in \mathbf{\Lambda}\}.$$

Moreover, a labeling $\bar{p}$ in the interior of $\mathbb{K}$ (i.e., $0 < p_i(\lambda) < 1$, for all i and λ) will be called *strictly ambiguous*.

Now, let $\bar{p} \in \mathbb{K}$ be any labeling assignment. To develop a relaxation algorithm that updates $\bar{p}$ in accordance with the compatibility model, we need to define, for each object $b_i \in \mathbf{B}$ and each label $\lambda \in \mathbf{\Lambda}$, what is called a *support* function. This should quantify the degree of agreement between the hypothesis that b_i is labeled with λ, whose confidence is expressed by $p_i(\lambda)$, and the context. This measure is commonly defined as follows (see, e.g., [4, 20, 21] for alternative definitions):

$$q_i(\lambda; \bar{p}) = \sum_{j=1}^{n} \sum_{\mu=1}^{m} r_{ij}(\lambda, \mu) p_j(\mu). \tag{1}$$

Putting together the $q_i(\lambda; \bar{p})$'s, as for the $p_i(\lambda)$'s, we obtain an nm-dimensional support vector that will be denoted by $\bar{q}(\bar{p})$.[2] Support factors have an obvious interpretation: $q_i(\lambda)$ is high when high-confidence neighboring labels are "compatible" with λ on b_i; conversely, it is low when high-confidence neighboring labels are "incompatible" with λ. Furthermore, notice that low-confidence nearby labels have little or no influence on the support measure, and this is what one should expect. The above discussion suggests a way to properly adjust the labeling $\bar{p}$: increase $p_i(\lambda)$ when $q_i(\lambda)$ is high and decrease it when $q_i(\lambda)$ is low. This naturally leads

to the following updating rule

$$p_i(\lambda) := p_i(\lambda)q_i(\lambda) \Big/ \sum_{\mu=1}^{m} p_i(\mu)q_i(\mu) \qquad (2)$$

where the denominator serves simply to ensure that the updated vectors are still in $\mathbb{IK}$. Formulas (1) and (2) define the original nonlinear relaxation operator of Rosenfeld et al. [1] which was in fact originally motivated by making recourse to the simple-minded, heuristic arguments just developed.

The relaxation algorithm will be best viewed as a continuous mapping $\mathcal{T}$ of the assignment space onto itself. It starts out with $\bar{p}^{(0)}$ and iteratively produces a sequence of points $\bar{p}^{(0)}, \bar{p}^{(1)}, \bar{p}^{(2)}, \ldots \in \mathbb{IK}$, where $\bar{p}^{(t+1)} = \mathcal{T}(\bar{p}^{(t)})$, $t \geq 0$. The process continues until (at least in theory) a fixed, or equilibrium, point is reached, which means that $\mathcal{T}(\bar{p}^{(t)}) = \bar{p}^{(t)}$, for some t. It can be easily shown that a labeling $\bar{p}$ is an equilibrium point for $\mathcal{T}$ if and only if the following relation holds [22]:

$$q_i(\lambda) = c_i \text{ whenever } p_i(\lambda) > 0, \quad i = 1 \ldots n, \ \lambda \in \Lambda \qquad (3)$$

for some nonnegative constants $c_1, \ldots, c_n$ (note that unambiguous labelings are therefore equilibrium points for $\mathcal{T}$; the converse, of course, need not be true).

One of the first questions that must be raised about the relaxation scheme discussed here, relates to the well-definedness of the iterates defined in Eq. (2). We now provide a practical criterion to test whether a given initial labeling $\bar{p}^{(0)}$ will give rise to a well-defined sequence $\{\bar{p}^{(0)}, \bar{p}^{(1)}, \bar{p}^{(2)}, \ldots\}$. First, however, we need an auxiliary concept. We say that a square matrix $A = (a_{ij})$ is *zero-symmetric* if $a_{ij} = 0$ implies $a_{ji} = 0$, for all i and j. Note that both symmetric and positive matrices are trivially zero-symmetric. The next result asserts that the sequence of points produced by the nonlinear relaxation mapping is well-defined for virtually all cases of practical interest.

Proposition 2.1. *Suppose that the compatibility matrix R is zero-symmetric, and let $\bar{p}^{(0)} \in \mathbb{IK}$ be a labeling for which $\mathcal{T}(\bar{p}^{(0)})$ is defined. Then, so will be $\mathcal{T}(\bar{p}^{(t)})$ for all $t \geq 1$.*

Proof: Suppose to the contrary that, for some $t_1 \geq 1$, $\mathcal{T}(\bar{p}^{(t_1)})$ is undefined. This amounts to stating that there exists some i such that $\sum_\mu p_i^{(t_1)}(\mu)q_i^{(t_1)}(\mu) = 0$ and this implies that for every μ we have either $p_i^{(t_1)}(\mu) = 0$ or $q_i^{(t_1)}(\mu) = 0$. By the hypothesis, instead, we have $\sum_\mu p_i^{(0)}(\mu)q_i^{(0)}(\mu) > 0$ which means that there exists at least one α for which $p_i^{(0)}(\alpha) > 0$ and $q_i^{(0)}(\alpha) > 0$. This second condition, in turn, implies that there exist j and β such that $p_j^{(0)}(\beta) > 0$ and $r_{ij}(\alpha, \beta) > 0$. On the other hand, we have either $p_i^{(t_1)}(\alpha) = 0$ or $q_i^{(t_1)}(\alpha) = 0$ and, since $p_i^{(0)}(\alpha) > 0$, there exists a $t_2 \leq t_1$ such that $p_i^{(t_2)}(\alpha) > 0$ and $q_i^{(t_2)}(\alpha) = 0$. Again, this last condition implies $p_j^{(t_2)}(\beta) = 0$ (being $r_{ij}(\alpha, \beta) > 0$), which means that, for some $t_3 < t_2$, $q_j^{(t_3)}(\beta) = 0$, because $p_j^{(0)}(\beta) > 0$. Finally, since $r_{ji}(\beta, \alpha) > 0$ for the zero-symmetry assumption, this would imply that $p_i^{(t_3)}(\alpha) = 0$ and consequently $p_i^{(t_2)}(\alpha) = 0$, being $t_2 > t_3$. This contradiction proves the proposition. □

We note that the conditions of the above proposition are very likely to hold in practical applications. In fact, should $\mathcal{T}(\bar{p}^{(0)})$ not be defined it would be so because for some object, say i, $p_i^{(0)}(\lambda) > 0$ implies $q_i^{(0)}(\lambda) = 0$. This would mean that the initial local measurements are extremely inaccurate and, in particular, that they strongly disagree with the constraint model. Fortunately, in nearly all practical applications of interest this is never the case. We add also that zero-symmetric compatibility matrices are very commonly encountered in practice (e.g., symmetric or positive matrices), and even if this were not the case, it would be straightforward to make a non-zero-symmetric matrix become zero-symmetric (it would suffice to replace all those zero-components for which the zero-symmetry condition fails with a small positive constant). In view of the above considerations, we will no longer be concerned here with this problem, and it will be reasonably assumed that the relaxation iterates are always well-defined.

3. Consistency and its Properties

In this section, we briefly review Hummel and Zucker's theory of constraint satisfaction [6] which commences by providing a general definition of consistency. By analogy with the unambiguous case, which is more easily understood, a weighted labeling assignment $\bar{p} \in \mathbb{IK}$ is said to be *consistent* if

$$\sum_{\lambda=1}^{m} p_i(\lambda)q_i(\lambda; \bar{p}) \geq \sum_{\lambda=1}^{m} v_i(\lambda)q_i(\lambda; \bar{p}), \quad i = 1 \ldots n \qquad (4)$$

for all $\bar{v} \in \mathbb{IK}$. Furthermore, if strict inequalities hold in (4), for all $\bar{v} \neq \bar{p}$, then $\bar{p}$ is said to be *strictly consistent*. It can be seen that a necessary condition for $\bar{p}$ to be strictly consistent is that it is an unambiguous one, that is $\bar{p} \in \mathbb{IK}^*$. Consistency is also usefully characterized by the following condition: $(\bar{v} - \bar{p}) \cdot \bar{q}(\bar{p}) \leq 0$ for all $\bar{v} \in \mathbb{IK}$, where "$\cdot$" denotes the standard inner product operator.

Given a labeling $\bar{p} \in \mathbb{IK}$, the tangent set at $\bar{p}$, denoted by $T_{\bar{p}}$, is defined as the set of possible directions along which one can move an infinitesimal amount away from $\bar{p}$, while remaining in $\mathbb{IK}$. It turns out that the tangent set at $\bar{p}$ is given by:

$$T_{\bar{p}} = \left\{ \bar{d} \in \mathbb{R}^{nm} \;\middle|\; \sum_{\lambda=1}^{m} d_i(\lambda) = 0, \quad i = 1 \ldots n, \right.$$
$$\left. p_i(\lambda) = 0 \Rightarrow d_i(\lambda) \geq 0, \quad i = 1 \ldots n, \ \lambda \in \mathbf{\Lambda} \right\}.$$

Owing to the convexity of $\mathbb{IK}$, all the tangent vectors at $\bar{p}$ are of the form $\gamma(\bar{v} - \bar{p})$, for some $\gamma \geq 0$ and $\bar{v} \in \mathbb{IK}$. Accordingly, consistency is equivalent to the condition $\bar{d} \cdot \bar{q}(\bar{p}) \leq 0$, for all $\bar{d} \in T_{\bar{p}}$.

A further useful characterization of consistent labelings is given in the next theorem, which also provides, unlike the preceding ones, an operational criterion to test the consistency of a given labeling[3].

Theorem 3.1. *A labeling $\bar{p} \in \mathbb{IK}$ is consistent if and only if for all $i = 1 \ldots n$ the following conditions hold*:

1) $q_i(\lambda) = c_i$, *whenever* $p_i(\lambda) > 0$
2) $q_i(\lambda) \leq c_i$, *whenever* $p_i(\lambda) = 0$

for some nonnegative constants $c_1, \ldots, c_n$.

Proof: Suppose that $\bar{p}$ is consistent. For all $i = 1 \ldots n$, put $c_i = \max \{q_i(\lambda; \bar{p}) : \lambda \in \mathbf{\Lambda}\}$ and let $\lambda_i \in \mathbf{\Lambda}$ be a label such that $q_i(\lambda_i) = c_i$. Then, trivially, condition (2) is satisfied. Suppose now by contradiction that $p_j(\mu) > 0$ and $q_j(\mu) < c_j$ for some j and μ, and consider the following vector $\bar{v} \in \mathbb{IK}$:

$$v_i(\lambda) = \begin{cases} p_i(\lambda), & \text{for } i \neq j, \\ 1, & \text{for } i = j \text{ and } \lambda = \lambda_j, \\ 0, & \text{for } i = j \text{ and } \lambda \neq \lambda_j. \end{cases}$$

We have

$$\sum_{\lambda=1}^{m} v_j(\lambda) q_j(\lambda; \bar{p}) = q_j(\lambda_j; \bar{p}) = c_j$$

and, since $p_j(\mu) > 0$ and $q_j(\mu) < c_j$,

$$\sum_{\lambda=1}^{m} p_j(\lambda) q_j(\lambda; \bar{p}) < c_j = \sum_{\lambda=1}^{m} v_j(\lambda) q_j(\lambda; \bar{p}).$$

This contradicts the hypothesis that $\bar{p}$ is consistent and proves the first part of the theorem.

Now, suppose that conditions 1) and 2) are satisfied and let $\bar{v} \in \mathbb{IK}$. For any $i = 1 \ldots n$ we have

$$\sum_{\lambda=1}^{m} p_i(\lambda) q_i(\lambda; \bar{p}) = \sum_{\substack{\lambda=1 \\ p_i(\lambda) > 0}}^{m} p_i(\lambda) c_i = c_i$$

and from $q_i(\lambda; \bar{p}) \leq c_i$ we obtain

$$\sum_{\lambda=1}^{m} v_i(\lambda) q_i(\lambda; \bar{p}) \leq \sum_{\lambda=1}^{m} v_i(\lambda) c_i = c_i$$
$$= \sum_{\lambda=1}^{m} p_i(\lambda) q_i(\lambda; \bar{p})$$

which proves that $\bar{p}$ is consistent. □

From the preceding theorem and characterization (3) of fixed points, the next corollary follows immediately, which establishes a first connection between nonlinear relaxation labeling and Hummel and Zucker's theory of consistency.

Corollary 3.2. *Let $\bar{p} \in \mathbb{IK}$ be consistent. Then $\bar{p}$ is a fixed point for the nonlinear relaxation operator $\mathcal{T}$. Moreover, if $\bar{p}$ is strictly ambiguous the converse also holds.*

In [6], Hummel and Zucker introduced the *average local consistency*, defined as

$$A(\bar{p}) = \sum_{i=1}^{n} \sum_{\lambda=1}^{m} p_i(\lambda) q_i(\lambda) \tag{5}$$

and proved the following fundamental result.

Theorem 3.3 (Hummel-Zucker, [6]). *Suppose that the compatibility matrix R is symmetric (i.e., $r_{ij}(\lambda, \mu) = r_{ji}(\mu, \lambda)$ for all i, j, λ, μ). Then any local maximum $\bar{p} \in \mathbb{IK}$ of A is consistent.*

Basically, this follows immediately from the fact that, when R is symmetric, we have

$$\frac{\partial A(\bar{p})}{\partial p_i(\lambda)} = 2 \sum_{j=1}^{n} \sum_{\mu=1}^{m} r_{ij}(\lambda, \mu) p_j(\mu) = 2 q_i(\lambda), \tag{6}$$

where $\partial A(\bar{p})/\partial p_i(\lambda)$ denotes the partial derivative of A with respect to $p_i(\lambda)$, evaluated at $\bar{p}$. In other words, $\nabla A(\bar{p}) = 2\bar{q}$, $\nabla A(\bar{p})$ being the gradient of A at $\bar{p}$. Note that, in general, the converse of Theorem 3.3 need not be true since, to prove this, second-order derivative information would be required. However, the next proposition asserts that, by demanding that $\bar{p}$ be strictly consistent, this *does* happen.

Proposition 3.4. *Let $\bar{e} \in \mathbb{IK}^*$ be strictly consistent, and suppose that R is symmetric. Then $\bar{e}$ is a strict local maximum of the average local consistency A.*

Proof: Suppose to the contrary that $\bar{e}$ is not a strict local maximum of A. Then, we can construct a sequence of points $\{\bar{v}_k\}$, $\bar{v}_k \in \mathbb{IK}$, such that $\bar{v}_k \to \bar{e}$ as $k \to \infty$, and $A(\bar{e}) \leq A(\bar{v}_k)$ for all k. Each $\bar{v}_k$ can be written in the form $\bar{v}_k = \bar{e} + \delta_k \bar{z}_k$, where $\delta_k > 0$ and the $\bar{z}_k$'s are tangent vectors at $\bar{e}$ with Euclidean norm $\|\bar{z}_k\| = 1$, for all k. Clearly, we have $\delta_k \to 0$ and since the sequence $\{\bar{z}_k\}$ is bounded, it must have some convergent subsequence converging to some $\bar{z}$. We may assume, without loss of generality, that the sequence $\{\bar{z}_k\}$ is itself convergent to $\bar{z}$. Notice that, since each $\bar{z}_k$ is in $T_{\bar{e}}$, $\bar{z}$ must be in $T_{\bar{e}}$ too. Now, applying Taylor's theorem, we can write

$$A(\bar{e}) - A(\bar{v}_k) = -\nabla A(\bar{e} + \theta_k \delta_k \bar{z}_k) \cdot \delta_k \bar{z}_k, \quad \text{all } k$$

for some $\theta_k \in [0, 1]$. Hence, $\nabla A(\bar{e}+\theta_k\delta_k\bar{z}_k)\cdot\delta_k\bar{z}_k \geq 0$ which implies $\nabla A(\bar{e} + \theta_k\delta_k\bar{z}_k) \cdot \bar{z}_k \geq 0$ for all k. Now, by letting $k \to \infty$, we have $\nabla A(\bar{e}) \cdot \bar{z} \geq 0$ and, by the symmetry of R, $\nabla A(\bar{e}) = 2\bar{q}(\bar{e})$. Thus we have found a vector $\bar{z} \in T_{\bar{e}}$ such that $\bar{z} \cdot \bar{q}(\bar{e}) \geq 0$. This contradicts the hypothesis that $\bar{e}$ is strictly consistent and proves the proposition. □

To conclude this section, we show that the concept of consistency is invariant under certain linear transformations of the compatibility strengths. Let R be a compatibility matrix and let $\mathcal{C}(R)$ denote the set of consistent labelings with respect to R. From [6] we know that $\mathcal{C}(R) \neq \emptyset$. Let α and β be arbitrary constants, with $\alpha > 0$, and construct the matrix R' as follows: $r'_{ij}(\lambda, \mu) = \alpha r_{ij}(\lambda, \mu) + \beta$. Then, it is straightforward to see that $\mathcal{C}(R) = \mathcal{C}(R')$. This follows immediately from the fact that $q'_i(\lambda) \equiv \sum_j \sum_\mu r'_{ij}(\lambda, \mu) p_j(\mu) = \alpha q_i(\lambda) + n\beta$. Thus, if the compatibility matrix R contains negative values, put $r'_{ij}(\lambda, \mu) = r_{ij}(\lambda, \mu) - \kappa$, where κ is the smallest negative value of R: we have that the transformed matrix R' is nonnegative, and $\mathcal{C}(R) = \mathcal{C}(R')$. This justifies therefore our restriction to nonnegative compatibilities.

4. Baum-Eagon's Inequality and Extensions

In the late 1960s, Baum and Eagon [24] introduced a class of nonlinear transformations in probability domains and proved a fundamental result which turns out to have a surprising relation with the Hummel-Zucker relaxation labeling theory. Their result generalizes an earlier one by Blakley [25] who discovered similar properties for certain homogeneous quadratic transformations. (Indeed, our analysis of nonlinear relaxation labeling could have been partly based upon Blakley's work but, because of its greater generality, the Baum-Eagon result was preferred.) Let $\mathbb{IK}$ denote the assignment space defined in Section 2 and, as usual, let $x_i(\lambda)$ represent the (i, λ) component of the vector $\bar{x} \in \mathbb{IK}$. The next theorem introduces what is known as the Baum-Eagon inequality.

Theorem 4.1 (Baum-Eagon [24]). *Let $P(\bar{x})$ be a homogeneous polynomial in the variables $\{x_i(\lambda)\}$ with nonnegative coefficients, and let $\bar{x}$ be a point of the domain $\mathbb{IK}$. Define the mapping $\mathcal{M}$ as follows:*

$$(\mathcal{M}(\bar{x}))_i(\lambda) = x_i(\lambda) \frac{\partial P(\bar{x})}{\partial x_i(\lambda)} \Bigg/ \sum_{\mu=1}^{m} x_i(\mu) \frac{\partial P(\bar{x})}{\partial x_i(\mu)} \,. \tag{7}$$

Then $P(\mathcal{M}(\bar{x})) > P(\bar{x})$, unless $\mathcal{M}(\bar{x}) = \bar{x}$.

In Eq. (7), the notation $(\mathcal{M}(\bar{x}))_i(\lambda)$ stands for the (i, λ) component of the vector $\mathcal{M}(\bar{x})$. A continuous mapping σ for which $f(\sigma(\bar{x})) \geq f(\bar{x})$ (f being an arbitrary real-valued function) is called a *growth transformation* for σ. Indeed, in a subsequent paper, Baum and Sell [26] proved much more. They showed that Theorem 4.1 still holds in the case of arbitrary (nonhomogeneous) polynomials with nonnegative coefficients, and further extended the result by proving that $\mathcal{M}$ increases P *homotopically*, which means that

$$P(\eta\mathcal{M}(\bar{x}) + (1-\eta)\bar{x}) \geq P(\bar{x}), \quad 0 \leq \eta \leq 1 \tag{8}$$

with equality if and only if $\mathcal{M}(\bar{x}) = \bar{x}$. In words, this means that not only is $P(\bar{x})$ smaller than $P(\mathcal{M}(\bar{x}))$, but $P(\bar{x})$ is also less than the value of P at any point lying on the segment joining $\bar{x}$ to $\mathcal{M}(\bar{x})$. However, they

showed through an example that $P(\eta\mathcal{M}(\bar{x})+(1-\eta)\bar{x})$ may fail to be monotone in η. In addition, in [26] Baum and Sell provide an analysis of the asymptotic behavior of growth transformations in the vicinity of local extrema.

As noted in [26], the mapping $\mathcal{M}$ defined in Theorem 4.1 makes use of first derivatives only and yet is able to take finite steps while increasing P. This contrasts sharply with classical gradient methods, for which an increase in the objective function is guaranteed only when infinitesimal steps are taken, and determining the optimal step size entails computing higher-order derivatives. We add that performing gradient ascent in $\mathbb{IK}$ requires some projection operator to ensure that the constraints not be violated, and this causes some problems for points lying on the boundary [9, 27]. In (7), instead, a computationally simple row normalization is required.

The Baum-Eagon inequality provides an effective iterative means for maximizing polynomial functions in probability domains, and in fact it has served as the basis for various statistical estimation techniques developed within the theory of probabilistic functions of Markov chains [28]. More recently, its usefulness in the field of speech recognition has been proven [29]. Further properties of the transformation $\mathcal{M}$ are investigated in [30, 31] and, recently, Gopalakrishnan et al. [32] introduced a new class of growth transformations for rational functions which have interesting applications to certain statistical estimation problems of interest in the speech recognition domain. Finally, in [7] a continuous-time version of the Baum-Eagon theorem is presented.

5. Nonlinear Relaxation Processes are Growth Transformations

This section provides some results concerning the nonlinear relaxation operator $\mathcal{T}$ which hold when the compatibility matrix R happens to be symmetric. This appears not to be too strict a condition because most of the compatibility coefficients proposed in the literature are symmetric (e.g., [1, 3, 15]), and symmetric compatibilities can also be easily learned from training data [17]. We begin by observing that the average local consistency defined in Section 3 is a homogeneous quadratic polynomial in the variables $\{p_i(\lambda)\}$ with nonnegative coefficients (the $r_{ij}(\lambda,\mu)$'s). Moreover, recall that when R is symmetric the gradient of A at $\bar{p}$ equals $2\bar{q}(\bar{p})$. Thus, by simply applying Theorem 4.1, the following fundamental result is easily proven.

Theorem 5.1. *The nonlinear relaxation operator $\mathcal{T}$ is a growth transformation for the average local consistency A, provided that compatibility coefficients are nonnegative and symmetric.*

More explicitly, the preceding theorem asserts that the nonlinear relaxation scheme strictly increases the average local consistency on each iteration, i.e.,

$$A(\bar{p}^{(t+1)}) > A(\bar{p}^{(t)}), \quad t = 0, 1, \ldots \tag{9}$$

until a fixed point is reached. Even more interestingly, from (8) we can assert that $A(\bar{p}^{(t)})$ is also smaller than the value of A at each labeling assignment lying on the segment joining $\bar{p}^{(t)}$ to $\bar{p}^{(t+1)}$, for each time step $t \geq 0$.

Another way of looking at Theorem 5.1 is from the standpoint of dynamical systems theory [33, 34]. The nonlinear relaxation operator $\mathcal{T}$ defines in fact a discrete dynamical system and it is therefore of particular interest to study how it behaves in the vicinity of its equilibrium points. In the theory of dynamical systems this is formalized by the concept of stability. An equilibrium point $\bar{x}$ is said to be *stable* if, whenever started sufficiently close to $\bar{x}$, the system will remain near to $\bar{x}$ for all future times. A stronger property, which is even more desirable, is that the equilibrium point $\bar{x}$ be *asymptotically stable*, meaning that $\bar{x}$ is stable and in addition is a *local attractor*, i.e., when initiated close to $\bar{x}$, the system tends towards $\bar{x}$ as time increases. One of the most fundamental tools for establishing the stability of a given equilibrium point is known as the Liapunov's direct method. It involves seeking a so-called *Liapunov* function, i.e., a continuous real-valued function defined in state space which is nonincreasing along a trajectory. Of particular interest are *strict* Liapunov functions which are, instead, strictly decreasing on nonconstant trajectories.

Now, from the preceding theorem we can assert that $-A$ is a strict Liapunov function for the nonlinear operator $\mathcal{T}$. This, in conjunction with the fact that every strictly consistent labeling is a local minimum of $-A$ (see Proposition 3.4), proves the following theorem which is an obvious consequence of Liapunov's theorem of asymptotically stability [33, 34].

Theorem 5.2. *Let $\bar{e} \in \mathbb{IK}^*$ be strictly consistent and suppose that the compatibility matrix R is nonnegative*

and symmetric. Then $\bar{e}$ is an asymptotically stable equilibrium point for the nonlinear relaxation scheme $\mathcal{T}$ and, consequently, is a local attractor.

As in Section 3, it is simple to show that the restriction to nonnegative compatibilities does not actually affect the generality of our discussion. In fact, suppose that compatibility coefficients are allowed to be negative. Consider the scaled compatibility matrix R' as defined at the end of Section 3, and let A' be the corresponding average local consistency. It is easily verified that $A'(\bar{v}) > A'(\bar{p})$ if and only if $A(\bar{v}) > A(\bar{p})$, for all $\bar{v}$, $\bar{p} \in \mathrm{I\!K}$. This means that if we have a growth transformation $\mathcal{M}$ for A', then $\mathcal{M}$ will be also a growth transformation for A, and *vice versa*. Put another way, the classes of growth transformations for A and A' coincide. Therefore, when negative compatibilities are given, we can use the scaled matrix R' instead of R; doing so, the relaxation operator $\mathcal{T}$ is also guaranteed to monotonically increase the average local consistency with respect to the original matrix R.

Before concluding this section, we observe that many relaxation labeling formulas proposed in the literature have exactly the same form as that specified in Eq. (2), but with different support functions. Some notable examples are presented in [3–5, 20]. The Baum-Eagon inequality may therefore be useful to explain the dynamical behavior of these algorithms, provided that the support vector $\bar{q}$ is proven to be proportional to the gradient of some polynomial consistency function. We mention that by using a "truncated" version of the Baum-Eagon theorem Stoddart et al. [7] have recently shed light on the behavior of the Kittler-Hancock [4] and the Christmas-Kittler-Petrou [5] probabilistic relaxation algorithms. Finally, we note that the Baum-Eagon inequality may also be useful in defining novel relaxation formulas once the labeling problem is cast as one of maximizing a polynomial consistency function over $\mathrm{I\!K}$ (e.g., [9]).

6. Nonlinear Relaxation with Arbitrary Compatibilities

In the preceding section, we have restricted ourselves to the case of symmetric compatibility coefficients and have shown how, under this circumstance, the heuristic nonlinear relaxation scheme is closely related to the theory of consistency of Hummel and Zucker. However, although symmetric compatibilities can easily be derived and asymmetric matrices can always be symmetrized (i.e., by considering $R + R^T$), it would be desirable for a relaxation process to work also when no restriction on the compatibility matrix is imposed [6, 10]. This is especially true when the relaxation algorithm is viewed as a plausible model of how biological systems perform visual computation [35].

Despite the fact that in the asymmetric case no functional to be maximized can be found, in this section it is shown that the nonlinear relaxation algorithm still performs useful computations in this case, and its connection with the theory of consistency continues to hold. This appears to be interesting as the inability of nonlinear relaxation to handle unrestricted compatibilities has been one of the strongest arguments against its usefulness [10]. We note, however, that certain results concerning the behavior of the algorithm in the case of arbitrary coefficients have long been established; particularly, we refer to a theorem by Elfving and Eklundh [13], an extension of which will be presented later in this section.

Let $\bar{p}$ be a non-consistent labeling, and consider the problem of updating $\bar{p}$ in agreement with the compatibility model. As Hummel and Zucker themselves pointed out [6], a reasonable strategy for doing this is to take a step in the same direction as $\bar{q}$; such a direction clearly exists for $\bar{p}$ is not consistent. Accordingly, the updating problem can be formally stated as follows:

$$\text{given } \bar{p} \in \mathrm{I\!K}, \text{ find } \bar{d} \in T_{\bar{p}} \text{ such that } \bar{d} \cdot \bar{q} \geq 0. \quad (10)$$

This kind of problems arises frequently in mathematical programming, where the role of the support vector $\bar{q}$ is indeed played by the gradient of some differentiable objective function. In particular, problems like (10) occur in the context of the so-called *methods of feasible directions*, and are solved by appropriate direction generator algorithms (see Zoutendijk [36] as the standard reference). In the problem we are dealing with, however, no objective function exists but the motivations that lead to (10) are quite similar.

Hummel and Zucker's approach to solving the updating problem consists of seeking, among all vectors satisfying (10), that which maximizes the inner product with $\bar{q}$ or, in other words, which makes the smallest angle with $\bar{q}$. They developed an updating rule, discussed in greater detail in [27], that when applied to interior points performs a simple orthogonal projection of $\bar{q}$ into the tangent plane $T_{\bar{p}}$. Unfortunately, in the case of boundary points (i.e., having at least one component equal to zero) the situation is much more

complicated and the procedure becomes computationally expensive. This led recently Parent and Zucker [37] to develop an alternative, more efficient updating rule which is also related to the previous one. According to Zoutendijk's distinction of direction generator methods ([36], Chap. 12), Hummel and Zucker's approach to solving the updating problem can be classified as an *optimization* method, but this is by no means the only possible strategy, and other approaches can well be pursued.

The next theorem asserts that the nonlinear relaxation algorithm *does* solve problem (10), and thus can be regarded as an alternative to the optimization approach developed in [6, 27]. It is based on a straightforward modification of a result of Levinson et al. [29] (see also [12] for a related derivation).

Theorem 6.1. *Let $\bar{p} \in \mathbb{IK}$ and $\bar{q}$ be a labeling and the associated support vector, respectively. Let $\mathcal{T}$ denote the nonlinear relaxation mapping defined in* (2). *Then* $(\mathcal{T}(\bar{p}) - \bar{p}) \cdot \bar{q} > 0$, *unless* $\mathcal{T}(\bar{p}) = \bar{p}$.

Proof: Put $k_i = \sum_\mu p_i(\mu) q_i(\mu)$, for all $i = 1 \ldots n$. From the definition of $\mathcal{T}$ we have

$$(\mathcal{T}(\bar{p}))_i(\lambda) \geq p_i(\lambda) \text{ if and only if } q_i(\lambda) \geq k_i,$$

therefore $(\mathcal{T}(\bar{p}))_i(\lambda) - p_i(\lambda)$ and $q_i(\lambda) - k_i$ have the same sign, and this implies that

$$\sum_{i=1}^{n} \sum_{\lambda=1}^{m} [(\mathcal{T}(\bar{p}))_i(\lambda) - p_i(\lambda)][q_i(\lambda) - k_i] \geq 0, \qquad (11)$$

from which we obtain $(\mathcal{T}(\bar{p}) - \bar{p}) \cdot \bar{q} \geq 0$. Since $\bar{p}$ is not a fixed point for $\mathcal{T}$, it must be true that $(\mathcal{T}(\bar{p}))_i(\lambda) > p_i(\lambda)$, for some i and λ for which $p_i(\lambda) > 0$, and this holds if and only if $q_i(\lambda) > k_i$. Thus, in sum (11) there exists at least one summand greater than zero, and this implies that $(\mathcal{T}(\bar{p}) - \bar{p}) \cdot \bar{q} > 0$, which proves the theorem. □

From the preceding theorem, we can therefore assert that the nonlinear relaxation scheme $\mathcal{T}$ belongs to the class of *feasibility* direction generator methods [36, p. 306] in that it does not concern itself with the problem of determining the "best" direction according to some criterion. In contrast with the Hummel-Zucker operator [6], it aims at solving problem (10) without any quest for optimality, and this represents the most important conceptual difference between the two algorithms. Note also that, unlike the Hummel-Zucker operator, the algorithm is able to determine the step size automatically.

One important question that has not been (purposely) raised until now, involves the relation between the stopping points of relaxation and the consistent labelings. As we have seen in Section 3, any consistent labeling is a stopping point for the nonlinear scheme, and the converse holds in very restricted circumstances, indeed of little practical interest. In fact, it may well happen that some fixed point $\bar{p}$ will not be consistent, and this follows essentially from the fact that the nonlinear scheme (2) cannot leave a face or edge of $\mathbb{IK}$. However, if we demand that the initial labeling $\bar{p}^{(0)}$ be strictly ambiguous (see Section 2) then this comes true. Before we show this, however, an auxiliary result is needed.

Lemma 6.2. *Let $\{\bar{p}^{(t)}\}$ be the sequence of points produced by the nonlinear relaxation operator $\mathcal{T}$. Suppose that, for some $t_1 \geq 0$, we have $q_i^{(t_1)}(\lambda) = 0$. Then, $q_i^{(t)}(\lambda) = 0$ for all $t \geq t_1$.*

Proof: If $q_i^{(t_1)}(\lambda) = 0$, then $\sum_{j,\mu} r_{ij}(\lambda, \mu) p_j^{(t_1)}(\mu) = 0$. Since all the quantities involved in the sum are nonnegative, we will have either $r_{ij}(\lambda, \mu) = 0$ or $p_j^{(t_1)}(\mu) = 0$, for all $j = 1 \ldots n$ and $\mu \in \mathbf{\Lambda}$. In the latter case, by the definition of $\mathcal{T}$ we will have $p_j^{(t)}(\mu) = 0$ for all $t \geq t_1$; therefore, for all $t \geq t_1$, $q_i^{(t)}(\lambda) = 0$. □

Theorem 6.3. *Let $\bar{p}^{(0)}$ be a strictly ambiguous labeling, and suppose that the sequence $\{\bar{p}^{(t)}\}$ produced by the nonlinear relaxation process $\mathcal{T}$ converges to the fixed point $\bar{p}^* \in \mathbb{IK}$. Then $\bar{p}^*$ is consistent.*

Proof: From Theorem 3.1, we have simply to show that

1) $q_i^*(\lambda) = c_i, \quad$ if $p_i^*(\lambda) > 0$
2) $q_i^*(\lambda) \leq c_i, \quad$ if $p_i^*(\lambda) = 0$

for some nonnegative constants $c_1, \ldots, c_n$, where $q_i^*(\lambda) \equiv q_i(\lambda; \bar{p}^*)$. Since $\bar{p}^*$ is a fixed point for the relaxation operator $\mathcal{T}$, condition 1) is fulfilled. Suppose now that $p_i^*(\lambda) = 0$. Since $\bar{p}^{(0)}$ is strictly ambiguous there exists a $t_1 \geq 1$ such that $p_i^{(t_1)}(\lambda) = 0$ and $p_i^{(t_1-1)}(\lambda) \neq 0$, and this implies that $q_i^{(t_1-1)}(\lambda) = 0$. From Lemma 6.2 we have $q_i^{(t)}(\lambda) = 0 \leq c_i$, for all $t \geq t_1 - 1$, and so will therefore be $q_i^*(\lambda)$. This proves the proposition. □

Probably, the preceding result holds true under still less restrictive hypotheses, but the condition that $\bar{p}^{(0)}$

not be on the boundary is not as strict as it would seem since in many practical applications of interest the initial local measurements *do* provide strictly ambiguous labelings. Even if this were not the case, any "boundary" initial labeling could be easily replaced with some nearby interior point, without distorting the information produced by the local measurements. It is interesting to notice that other updating rules require that the initial labeling assignments be strictly ambiguous [37].

To conclude our analysis in the general case of arbitrary compatibilities we need, as in the previous section, some local convergence result that asserts something about the behavior of nonlinear relaxation in the vicinity of consistent labelings. Interestingly enough, such a result has been available since the early 1980s thanks to the work of Elfving and Eklundh [13], but it seems to us that its importance as well as its potential applications have not been fully recognized (cf. Section 8). Here we present an extension of that result.

Theorem 6.4. *Let $\bar{e} \in \mathbb{IK}^*$ be a strictly consistent labeling. Then $\bar{e}$ is an asymptotically stable equilibrium point for the nonlinear relaxation scheme $\mathcal{T}$.*

Proof: To prove the theorem we have to show that $\bar{e}$ is a stable equilibrium point, and also a local attractor for $\mathcal{T}$. The latter condition was earlier proven in [13, Theorem 10] by showing that the spectral radius of the Jacobian of $\mathcal{T}$ evaluated at any strictly consistent labeling is less than 1. The fact that $\bar{e}$ is a local attractor follows therefore immediately from a well-known result by Ostrowski [38, Theorem 22.1] (we do not reproduce the proof here and refer to the original paper for technical details). It remains thus to see that $\bar{e}$ is stable. Formally, this is expressed by the following condition: for any $\epsilon > 0$ there exists a $\delta > 0$ such that $\|\bar{p} - \bar{e}\| < \delta$ implies $\|\mathcal{T}^t(\bar{p}) - \bar{e}\| < \epsilon$ for all $t \geq 0$, where $\mathcal{T}^t$ denotes the tth iterate of $\mathcal{T}$, i.e., $\mathcal{T}^0(\bar{p}) = \bar{p}$ and $\mathcal{T}^t(\bar{p}) = \mathcal{T}(\mathcal{T}^{t-1}(\bar{p}))$ for $t \geq 1$. First of all, since $\bar{e}$ is a local attractor for $\mathcal{T}$, there must exist a $\delta' > 0$ such that $\lim_{t\to\infty} \mathcal{T}^t(\bar{p}) = \bar{e}$ whenever $\|\bar{p} - \bar{e}\| < \delta'$. Now, let $\epsilon > 0$ be an arbitrary positive constant. There exists a nonnegative integer t_0 such that for all $t > t_0$ we have $\|\mathcal{T}^t(\bar{p}) - \bar{e}\| < \epsilon$ whenever $\|\bar{p} - \bar{e}\| < \delta'$. Furthermore, since $\mathcal{T}$ is continuous, so is $\mathcal{T}^t$ for all $t \geq 0$. This means that, for any choice of t, there exists a $\delta_t > 0$ such that $\|\bar{p} - \bar{e}\| < \delta_t$ implies $\|\mathcal{T}^t(\bar{p}) - \mathcal{T}^t(\bar{e})\| = \|\mathcal{T}^t(\bar{p}) - \bar{e}\| < \epsilon$ (recall that $\bar{e}$ is a fixed point for $\mathcal{T}$). Therefore, by setting $\delta = \min\{\delta', \delta_1, \ldots, \delta_{t_0}\}$ the condition for the stability of $\bar{e}$ follows immediately, thereby proving the theorem. □

This is the analog to the fundamental local convergence result of Hummel and Zucker [6, Theorem 9.1]. Note that, unlike Theorem 5.2 no restriction on the structure of the compatibility matrix is imposed here. We conclude by mentioning that in [13] it is also shown that the rate of convergence of $\mathcal{T}$ is linear.

7. Higher-Order Relaxation Schemes

One of the working assumptions upon which our entire analysis of nonlinear relaxation labeling has rested, is that contextual relations be expressed in terms of pairwise compatibility coefficients, neglecting in fact possible higher-order interactions. The extent to which this approximation is valid depends of course on the complexity of the problem at hand, but a vast body of computational experience clearly demonstrates that second-order relations usually suffice to model satisfactorily real-world constraints. In principle, higher-order interactions could well be employed in practical applications, but this would cause computational problems due to the combinatorial growth of calculations and storage requirements involved, and sometimes the improvement in performance does not compensate for the increase in computational cost [39].

Despite the above considerations, however, we still find it interesting to study higher-order relaxation processes. This is partly because in certain applications the use of high-order correlations has really shown to be advantageous (see, e.g., [40–42]). But, even more interestingly, higher-order interactions have recently proven to greatly enhance the performance of neural models in such tasks as learning and generalization [43], and have shown to significantly increase the storage capacity of associative memories [44]. Additionally, such "multiplicative" interactions exhibit an intriguing biological plausibility [45].

In this section, it is shown that all the properties of "second-order" nonlinear relaxation discussed previously naturally generalize to the case of higher-order compatibility relations. For clarity of discussion we begin by considering third-order compatibilities. In this case, the compatibility matrix becomes a three-dimensional matrix (each dimension having nm entries), and compatibility coefficients express the degree of agreement of triples of object-label configurations. Specifically, the coefficient $r_{ijk}(\lambda, \mu, \nu)$ quantifies the

compatibility among labels λ, μ and ν on objects b_i, b_j and b_k, respectively. Accordingly, our linear support function becomes

$$q_i(\lambda) = \sum_{j,\mu} \sum_{k,\nu} r_{ijk}(\lambda, \mu, \nu) p_j(\mu) p_k(\nu) \qquad (12)$$

and the third-order nonlinear relaxation operator $\mathcal{T}$ has precisely the same form as in Eq. (2) but, now, with this new meaning of the $q_i(\lambda)$'s. Moreover, the definition of consistency as well as all its properties hold true in this case, with the appropriate definition of $\bar{q}$. In particular, the average local consistency A turns out to be a cubic homogeneous polynomial.

For third-order compatibilities the symmetry restriction is slightly more complicated than the corresponding second-order counterpart. It states that

$$\begin{aligned} r_{ijk}(\lambda, \mu, \nu) + r_{kij}(\nu, \lambda, \mu) + r_{jki}(\mu, \nu, \lambda) \\ = 3 r_{ijk}(\lambda, \mu, \nu) \end{aligned} \qquad (13)$$

for all $i, j, k = 1 \ldots n$ and $\lambda, \mu, \nu \in \mathbf{\Lambda}$. In this case, we have that $\nabla A(\bar{p}) = 3\bar{q}$ and, trivially, the Baum-Eagon inequality can still be applied here. This implies that third-order nonlinear relaxation processes continue to be growth transformations for the average local consistency A or, put another way, $-A$ is still a Liapunov function for $\mathcal{T}$. Therefore, all the results concerning the behavior of the relaxation algorithm in the symmetric case hold true in this case (Section 5). Likewise, it is readily seen that the properties of Section 6 (arbitrary compatibilities) continue to hold.

The above discussion extends to the more general case of Kth-order compatibility relations in a straightforward fashion. In this case, the appropriate support function is

$$\begin{aligned} q_i(\lambda) = \sum_{i_2,\lambda_2} \sum_{i_3,\lambda_3} \cdots \sum_{i_K,\lambda_K} r_{i,i_2,\ldots,i_K}(\lambda, \lambda_2, \ldots, \lambda_K) \\ \times p_{i_2}(\lambda_2) \cdots p_{i_K}(\lambda_K) \end{aligned} \qquad (14)$$

and the average local consistency is still a homogeneous polynomial (of degree K). The symmetry condition complicates considerably here, and becomes

$$\begin{aligned} \sum_{\sigma \in C_K} r_{i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(K)}}\left(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \ldots, \lambda_{\sigma(K)}\right) \\ = K r_{i_1, i_2, \ldots, i_K}(\lambda_1, \lambda_2, \ldots, \lambda_K) \end{aligned} \qquad (15)$$

for all $i_1, i_2, \ldots, i_K = 1 \ldots n$ and $\lambda_1, \lambda_2, \ldots, \lambda_K \in \mathbf{\Lambda}$, where C_K is the set of cyclic permutations on K objects. In the symmetric case we have $\nabla A(\bar{p}) = K\bar{q}$. All the preceding arguments apply immediately here for the symmetric and the asymmetric case, and we omit the details.

8. Summary and Discussion

Nonlinear relaxation labeling processes represent a fundamental tool in the computer vision toolkit. They were derived heuristically in the mid-1970s to solve certain constraint satisfaction problems arising in the interpretation of ambiguous line drawings, and since then have been employing successfully in a variety of practical tasks. At the same time, however, their theory has never been fully understood, and their heuristic nature has always been indicated as one of the most serious drawbacks of the model.

In this paper, we have sought to provide a definite answer to many long-standing questions about the original nonlinear relaxation scheme. We have offered a unified treatment of the algorithm and have proven a number of properties which show that, in the face of its completely heuristic derivation, it turns out to be intimately related to a well-known mathematical theory of consistency developed by Hummel and Zucker. In particular, based on a powerful result due to Baum and Eagon, we have shown that in the case of symmetric compatibilities the algorithm possesses a Liapunov function which is precisely the measure of (in)consistency proposed by Hummel and Zucker in [6]. This amounts to stating that each relaxation labeling iteration actually increases the labeling's consistency, and the algorithm eventually approaches the nearest consistent solution. We have then proceeded in our analysis by relaxing the symmetry requirement and have found that most of the essential dynamical properties of the algorithm hold true even in the general case of unrestricted compatibilities. Finally, a generalization of these results to higher-order relaxation schemes has been discussed.

Apart from their theoretical value, the results presented in this paper have a number of applications and implications that now we wish to discuss briefly. Perhaps, the most immediate application that comes to mind is in the field of optimization. In a classical paper [46], Hopfield and Tank demonstrated how certain networks of simple locally interacting processing units are capable of solving difficult optimization problems. To do so, they simply exploited the property of such networks to have an associated energy function, and

proposed a method for mapping optimization problems directly onto specific networks. The fact that the nonlinear relaxation algorithm possesses essentially the same energy function as the Hopfield neural network (Theorem 5.1) implies that the model can be utilized to solve optimization problems in exactly the same way as Hopfield and Tank did in their original work. We note, however, that the two models operate in different state spaces. In fact, while the Hopfield model works on a "fuzzy" domain where every variable is simply constrained to be in the range [0, 1], relaxation labeling operates on a "probabilistic" state space where the variables are subject to an additional normalization constraint. This is a remarkably positive fact that leads to formulate simpler energy functions containing less easily-determined operational parameters [47]. As a matter of fact, experimental evidence has recently demonstrated the power of Rosenfeld et al.'s relaxation algorithm in solving hard optimization problems such as the traveling salesman [48] or the maximum clique problems [49], and the results show that the algorithm clearly outperforms the Hopfield model, whose effectiveness has recently been questioned [50]. Now, a wide range of tasks in computer vision and pattern recognition can be formulated in terms of minimizing some cost function [51], and in fact many authors are currently employing Hopfield-style networks to solve them (see, e.g., [52] and references therein). From the very encouraging results obtained so far, it is expected that nonlinear relaxation labeling has a good chance of being capable of solving such vision problems within this energy-minimization framework. This offers a new and systematic approach for designing relaxation labeling applications which, at present, are instead typically guided by simple heuristic reasoning.

A further application of the results proven in this paper arises in the context of associative memories. The existence of stable attractors in state space is in fact a highly desirable property that can be practically exploited to construct an associative memory device [53]: memory patterns can be stored as attractive fixed points of the system, so that when started in their vicinity the memory will eventually "recall" the nearest one. This is a property that is conveniently exploited in such practical tasks as, for example, face processing [54], and object recognition [55]. From Theorems 5.2 and 6.4 we know that strictly consistent labelings are asymptotically stable equilibrium points, and hence local attractors, for the nonlinear relaxation dynamical system, and this naturally suggests the idea of using the algorithm as an associative memory model. In this kind of memory, patterns are first mapped onto unambiguous labeling assignments by means of a 1-of-m encoding and then made strictly consistent by solving a system of linear equations (the storing phase). A more detailed description of this kind of memory along with experimental results which confirm its validity can be found in [19, 56]. A remarkable feature of the approach is that it naturally allows us to develop asymmetric multi-valued associative memories, which are believed to be more closely related to biology and turn out to be more useful in practical applications [57].

Our concluding considerations are rather speculative, and relate to the plausibility of relaxation labeling as a model of biological visual computation. Because of their parallel and cooperative nature, in fact, it has been clear since the beginning that intriguing similarities exist between relaxation labeling processes and certain mechanisms in the early stages of biological visual systems [1, 6]. This observation is indeed supported by much physiological and anatomical evidence. To begin, extensive collateral connections between pyramidal cells have been found in the cerebral cortex [58] and, as pointed out by Anderson [59], these strongly resemble the kind of connections that are implemented in a relaxation labeling network. Interestingly, such collaterals rarely extend over long distances, and this led Crick and Asanuma to state that "in mathematical terms, for one cortical area the connections seem to be 'near-diagonal,' assuming that we have a two-dimensional arrangement of cells and a four-dimensional connection matrix" [58, p. 366]. We observe that this is perfectly in accordance with the vast body of computational experience with relaxation labeling, where the compatibility matrix turns out to have actually such a "near-diagonal" form. This is so because, due to computational requirements, in real-world applications it is customary to neglect the influence of distant labels. More recently, the biological significance of relaxation labeling has also been strongly advocated by Zucker et al. [35] who explicitly hypothesized that the first 2–3 iterations of the algorithm could be implemented by the pyramidal neurons connecting the striate and extrastriate cortices. Finally, until recently one of the biggest obstacles to consider relaxation labeling as a realistic brain model, came from its (apparent) lack of learning abilities, which is of course one of the most fundamental aspects of the human brain. Anderson and Hinton, classified in fact relaxation labeling algorithms as "systems of simple units

with fixed interconnections" [60]. Indeed, this misconception stemmed more from the utter lack of any procedure that made relaxation learn, than from a real, intrinsic inability of the algorithm to do this. Recently, this obstacle has been removed and relaxation labeling processes have instead exhibited interesting learning capabilities [17–19].

The results presented in this paper may serve to strengthen the biological plausibility hypothesis discussed above in at least two important ways. First, though it is still an unanswered question as to what style of computation is actually implemented in the brain, increasing evidence supports an appealing alternative to the conventional von Neumann's paradigm which is clearly unsatisfactory for biological computation. The key idea behind this style of computation consists of viewing information processing as the motion of a dynamical system which evolves in a state space dominated by few locally stable attractors [53, 61]. Hogg and Huberman [62] demonstrated that this computational paradigm exhibits a number of interesting properties that biological neural systems are known to possess, such as adaptation to the environment, self-repair, and conditional learning. They concluded that "computing with attractors may be applicable to biological computation" [62, p. 6875]. Now, the fact that, under symmetric connectivity restriction, nonlinear relaxation labeling possesses a Liapunov function (Theorem 5.1) that makes the algorithm evolve towards stable attractors (i.e., strictly consistent labelings, see Theorem 5.2), is a remarkable property which demonstrates how the algorithm, in view of the above discussion, exhibits computational properties that are identical to those the human brain is supposed to have. On the other hand, this makes the algorithm intimately related to other well-known neural network models for which similar collective properties have been discovered [53, 63, 64]. As to this last point, we mention that some effort has recently been made in an attempt to establish a formal correspondence between the fields of relaxation labeling and artificial neural networks [65, 66], and the results discussed in this paper can also be intended as a contribution towards this direction.

The second, and perhaps most remarkable feature of nonlinear relaxation discussed in this paper, which makes it even more attractive as a computational brain model, is that all its fundamental dynamical properties are retained when the symmetry restriction is relaxed, and the algorithm continues to perform useful computations in this case (see Section 6 and, particularly, Theorem 6.4). This contrasts sharply with existing neural network models, whose convergence properties are instead known to hold only in the case of symmetric connections [53, 63, 64] and, in the asymmetric case, cyclic or chaotic behavior can occur. We emphasize that the symmetry restriction is widely recognized to be biologically implausible, since empirical evidence clearly suggests otherwise (e.g., [61]).

In conclusion, whatever the value of such speculations, the theoretical analysis offered in this paper definitively clarifies the dynamical behavior of the standard Rosenfeld et al.'s relaxation labeling algorithm, thereby making it as well-founded and mathematically justified as other notoriously more formal relaxation schemes [3, 4, 6, 9]. It is hoped that our work may serve to explain the success of the algorithm, and also to guide practitioners in its future applications.

Notes

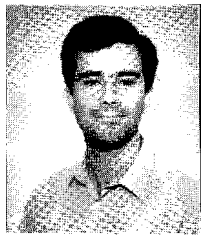
1. We remark that all the results presented in this paper hold true in the more general case when objects are associated with different label sets.
2. Henceforth, when it will be clear from context, the dependence on $\bar{p}$ will be understood.
3. See [23] for a different proof.

References

1. A. Rosenfeld, R.A. Hummel, and S.W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst. Man Cybern.*, Vol. 6, pp. 420–433, 1976.
2. J. Kittler and J. Illingworth, "Relaxation labeling algorithms—A review," *Image Vision Comput.*, Vol. 3, pp. 206–216, 1985.
3. S. Peleg, "A new probabilistic relaxation scheme," *IEEE Trans. Pattern Machine Intell.*, Vol. 2, pp. 362–369, 1980.
4. J. Kittler and E.R. Hancock, "Combining evidence in probabilistic relaxation," *Int. J. Pattern Recognition Artif. Intell.*, Vol. 3, pp. 29–51, 1989.
5. W.J. Christmas, J. Kittler, and M. Petrou, "Structural matching in computer vision using probabilistic relaxation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 17, pp. 749–764, 1995.
6. R.A. Hummel and S.W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 5, pp. 267–287, 1983.
7. A.J. Stoddart, M. Petrou, and J. Kittler, "On the foundations of probabilistic relaxation with product support," submitted for publication, 1996.
8. S. Ullman, "Relaxation and constrained optimization by local processes," *Comput. Graph. Image Processing*, Vol. 10, pp. 115–125, 1979.
9. O.D. Faugeras and M. Berthod, "Improving consistency and reducing ambiguity in stochastic labeling: An optimization

approach," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 3, pp. 412–424, 1981.
10. J. Illingworth and J. Kittler, "Optimization algorithms in probabilistic relaxation labeling," in *Pattern Recognition Theory and Applications*, P.A. Devijver and J. Kittler (Eds.), Springer-Verlag: Berlin, 1987, pp. 109–117.
11. E.R. Hancock and J. Kittler, "Discrete relaxation," *Pattern Recognition*, Vol. 23, pp. 711–733, 1990.
12. S.A. Lloyd, "An optimization approach to relaxation labeling algorithms," *Image Vision Comput.*, Vol. 1, pp. 85–91, 1983.
13. T. Elfving and J.-O. Eklundh, "Some properties of stochastic labeling procedures," *Comput. Graph. Image Processing*, Vol. 20, pp. 158–170, 1982.
14. M. Levy, "A new theoretical approach to relaxation—Application to edge detection," in *Proc. 9th Int. Conf. Pattern Recognition*, Rome, Italy, 1988, pp. 208–212.
15. S. Peleg and A. Rosenfeld, "Determining compatibility coefficients for curve enhancement relaxation processes," *IEEE Trans. Syst. Man Cybern.*, Vol. 8, pp. 548–555, 1978.
16. R.C. Wilson and E.R. Hancock, "A Bayesian compatibility model for graph matching," *Pattern Recognition Lett.*, Vol. 17, pp. 263–276, 1996.
17. M. Pelillo and M. Refice, "Learning compatibility coefficients for relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 16, pp. 933–945, 1994.
18. M. Pelillo, F. Abbattista, and A. Maffione, "An evolutionary approach to training relaxation labeling processes," *Pattern Recognition Lett.*, Vol. 16, pp. 1069–1078, 1995.
19. M. Pelillo and A.M. Fanelli, "Autoassociative learning in relaxation labeling networks," in *Proc. 13th Int. Conf. Pattern Recognition*, Vienna, Austria, 1996.
20. R.L. Kirby, "A product rule relaxation method," *Comput. Graph. Image Processing*, Vol. 13, pp. 158–189, 1980.
21. E.R. Hancock and J. Kittler, "Edge-labeling using dictionary-based relaxation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 12, pp. 165–181, 1990.
22. S.W. Zucker, E.V. Krishnamurthy, and R.L. Haar, "Relaxation processes for scene labeling: Convergence, speed, and stability," *IEEE Trans. Syst. Man Cybern.*, Vol. 8, pp. 41–48, 1978.
23. X. Zhuang, R.M. Haralick, and H. Joo, "A simplex-like algorithm for the relaxation labeling process," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, pp. 1316–1321, 1989.
24. L.E. Baum and J.A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bull. Amer. Math. Soc.*, Vol. 73, pp. 360–363, 1967.
25. G.R. Blakley, "Homogeneous nonnegative symmetric quadratic transformations," *Bull. Amer. Math. Soc.*, Vol. 70, pp. 712–715, 1964.
26. L.E. Baum and G.R. Sell, "Growth transformations for functions on manifolds," *Pacific J. Math.*, Vol. 27, pp. 211–227, 1968.
27. J.L. Mohammed, R.A. Hummel, and S.W. Zucker, "A gradient projection algorithm for relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 5, pp. 330–332, 1983.
28. L.E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, Vol. 41, pp. 164–171, 1970.
29. S.E. Levinson, L.R. Rabiner, and M.M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition," *Bell. Syst. Tech. J.*, Vol. 62, pp. 1035–1074, 1983.
30. P. Stebe, "Invariant functions of an iterative process for maximization of a polynomial," *Pacific J. Math.*, Vol. 43, pp. 765–783, 1972.
31. D.S. Passman, "The Jacobian of a growth transformation," *Pacific J. Math.*, Vol. 44, pp. 281–290, 1973.
32. P.S. Gopalakrishnan, D. Kanevsky, A. Nádas, and D. Nahamoo, "An inequality for rational functions with applications to some statistical estimation problems," *IEEE Trans. Inform. Theory*, Vol. 37, pp. 107–113, 1991.
33. D.G. Luenberger, *Introduction to Dynamic Systems*, Wiley: New York, 1979.
34. J.P. LaSalle, *The Stability and Control of Discrete Processes*, Springer-Verlag: New York, 1986.
35. S.W. Zucker, A. Dobbins, and L. Iverson, "Two stages of curve detection suggest two styles of visual computation," *Neural Computat.*, Vol. 1, pp. 68–81, 1989.
36. G. Zoutendijk, *Mathematical Programming Methods*, North-Holland: Amsterdam, 1976.
37. P. Parent and S.W. Zucker, "Radial projection: An efficient update rule for relaxation labeling," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, pp. 886–889, 1989.
38. A.M. Ostrowski, *Solution of Equations and Systems of Equations*, Academic Press: New York, 1966.
39. J.-O. Eklundh and A. Rosenfeld, "Some relaxation experiments using triples of pixels," *IEEE Trans. Syst. Man Cybern.*, Vol. 10, pp. 150–153, 1980.
40. S. Peleg, "Ambiguity reduction in handwriting with ambiguous segmentation and uncertain interpretation," *Comput. Graph. Image Processing*, Vol. 10, pp. 235–245, 1979.
41. J. Kittler and E.R. Hancock, "Contextual decision rule for region analysis," *Image Vision Comput.*, Vol. 5, pp. 145–154, 1987.
42. A.M. Finch, R.C. Wilson, and E.R. Hancock, "Matching Delaunay triangulations by relaxation labeling," in *Computer Analysis of Images and Patterns*, V. Hlavác and R. Sára (Eds.), Springer: Berlin, 1995, pp. 350–358.
43. C.L. Giles and T. Maxwell, "Learning, invariance, and generalization in high-order neural networks," *Appl. Optics*, Vol. 26, pp. 4972–4978, 1987.
44. D. Psaltis and C.H. Park, "Nonlinear discriminant functions and associative memories," in *Neural Networks for Computing*, J.S. Denker (Ed.), American Institute of Physics: New York, 1986, pp. 370–375.
45. R. Durbin and D.E. Rumelhart, "Product units: A computationally powerful and biologically plausible extension to backpropagation networks," *Neural Computat.*, Vol. 1, pp. 133–142, 1989.
46. J.J. Hopfield and D.W. Tank, "Neural computation of decisions in optimization problems," *Biol. Cybern.*, Vol. 52, pp. 141–152, 1985.
47. D.E. Van den Bout and T.K. Miller, "A traveling salesman objective function that works," in *Proc. IEEE Int. Conf. Neural Networks*, San Diego, CA, 1988, pp. 299–303.
48. M. Pelillo, "Relaxation labeling processes for the traveling salesman problem," in *Proc. Int. J. Conf. Neural Networks*, Nagoya, Japan, 1993, pp. 2429–2432.
49. M. Pelillo, "Relaxation labeling networks for the maximum clique problem," *J. Artif. Neural Networks*, Special issue on

"Neural Networks for Optimization," Vol. 2, pp. 313–328, 1995.
50. G.V. Wilson and G.S. Pawley, "On the stability of the traveling salesman problem algorithm of Hopfield and Tank," *Biol. Cybern.*, Vol. 58, pp. 63–70, 1988.
51. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, Vol. 317, pp. 314–319, 1985.
52. N.M. Nasrabadi and C.Y. Choo, "Hopfield network for stereo vision correspondence," *IEEE Trans. Neural Networks*, Vol. 3, pp. 5–13, 1992.
53. J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci. USA*, Vol. 79, pp. 2554–2558, 1982.
54. D. Valentin, H. Abdi, A.J. O'Toole, and G.W. Cottrell, "Connectionist models of face processing: A survey," *Pattern Recognition*, Vol. 27, pp. 1209–1230, 1994.
55. H. Wechsler, "Network representations and match filters for invariant object recognition," in *Pattern Recognition Theory and Applications*, P.A. Devijver and J. Kittler (Eds.), Springer-Verlag: Berlin, 1987, pp. 269–276.
56. M. Pelillo and A.M. Fanelli, "An asymmetric associative memory model based on relaxation labeling processes," in *Proc. ESANN'95—3rd Europ. Symp. Artif. Neural Networks*, Brussels, Belgium, 1995, pp. 223–228.
57. G.A. Kohring, "Finite-state neural networks: A step toward the simulation of very large systems," *J. Stat. Phys.*, Vol. 62, pp. 563–576, 1991.
58. F. Crick and C. Asanuma, "Certain aspects of the anatomy and physiology of the cerebral cortex," in *Parallel Distributed Processing. Vol 2: Psychological and Biological Models*, J.L. McClelland and D.E. Rumelhart (Eds.), MIT Press: Cambridge, MA, 1986, pp. 333–371.
59. J.A. Anderson, "Cognitive and psychological computation with neural models," *IEEE Trans. Syst. Man Cybern.*, Vol. 13, pp. 799–815, 1983.
60. J.A. Anderson and G.E. Hinton, "Models of information processing in the brain," in *Parallel Models of Associative Memory*, G.E. Hinton and J.A. Anderson (Eds.), Erlbaum: Hillsdale, NJ, 1981, pp. 9–48.
61. D.J. Amit, *Modeling Brain Function*, Cambridge University Press: Cambridge, UK, 1989.
62. T. Hogg and B.A. Huberman, "Understanding biological computation: Reliable learning and recognition," *Proc. Natl. Acad. Sci. USA*, Vol. 81, pp. 6871–6875, 1984.
63. M.A. Cohen and S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," *IEEE Trans. Syst. Man Cybern.*, Vol. 13, pp. 815–826, 1983.
64. G.E. Hinton and T.J. Sejnowski, "Learning and relearning in Boltzmann machines," in *Parallel Distributed Processing. Vol 1: Foundations*, D.E. Rumelhart and J.L. McClelland (Eds.), MIT Press: Cambridge, MA, 1986, pp. 282–317.
65. C. Torras, "Relaxation and neural learning: Points of convergence and divergence," *J. Parallel Distrib. Computing*, Vol. 6, pp. 217–244, 1989.
66. S.S. Yu and W.H. Tsai, "Relaxation by the Hopfield neural network," *Pattern Recognition*, Vol. 25, pp. 197–209, 1992.



Marcello Pelillo was born in Taranto, Italy, on June 1, 1966. He received the "Laurea" degree with honors in Computer Science from the University of Bari, Italy, in 1989.

From 1988 to 1989 he spent about a year at the IBM Scientific Center in Rome, where he was involved in studies on natural language and speech processing. In 1991 he joined the Department of Computer Science at the University of Bari, Italy, as Assistant Professor. Since 1995 he is with the Department of Applied Mathematics and Computer Science at the University of Venice "Ca' Foscari". In 1995 and 1996 he was a Visiting Professor at the Department of Computer Science of the University of York, UK, and in 1995 he visited the Center for Intelligent Machines of the Department of Electrical Engineering at the McGill University, Montréal, Canada.

His research interests are in the areas of computer vision, neural networks and pattern recognition, where he has published over 40 papers in refereed journals and conference proceedings. He has been the program co-chair of the *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (EMMCVPR '97) held in Venice, Italy, in May 1997. He is member of the IEEE Computer Society and the Pattern Recognition Society.