Philips Research Laboratory Brussels
Av. E. Van Becelaere 2, Box 8
B-1170 Brussels, Belgium

A General Model for Causal Markov Fields

Christian Ronse

August 1985

Provisional working document

Distribution: X. Aubert, H. Bourlard, M. Dekesel, P. Devijver, Y. Kamp, V. Lacroix, C. Ronse, D. Snyers, A. Thayse, C. Wellekens.

Abstract: *We give a general framework for the study of causal Markov fields on arbitrary hierarchical structures. It includes as particular cases 1-D Markov chains and 2-D Markov fields (as defined by Abend, harley and Kanal). Several equivalent statements are given for the Markov field hypothesis and for the definition of a hidden Markov field. The applicability of Baum's forward-backward method is considered, and we show in particular that it can be implemented in the case where the hierarchical structure is a tree whose causal ordering is from the leaves to the root.*

I. Introduction

One of the simplest probabilistic models for one-dimensional signals is the nth order (causal) Markov chain: the state λ_t at time t is uniquely determined by the transition probabilities $p(\lambda_u/\lambda_{u-i},\ 1 \leq i \leq n)$. It has found applications in several domains, especially speech recognition (see in particular [10]). This model has been refined by supposing that the states λ_t are unknown (or rather hidden), and that one has for each t an observation ξ_t which depends upon the sequence of all λ_u by the probability $p(\xi_t/\lambda_t)$ only. One speaks then of a *hidden* Markov chain. The choice of the most probable λ_t given the entire observation can be computed in linear time thanks to Baum's *forward-backward algorithm*.

Several researchers have attempted to generalize this methodology, which is central in speech recognition, to image processing, in other words to extend Markov chains in one dimension to Markov fields in two or more dimensions [1,3,5,6,7,8]. As such a field is causal (like Markov chains), it relies on an ordering of the pixels of the image.

In this working document we study Markov fields on arbitrary hierarchical structures, which can be a one-dimensional chain, an ordered two-dimensional grid, a quadtree, etc.; the transition probabilities are expressed in terms of predecessors in the hierarchical structure. We formulate in a general framework certain results obtained in particular cases (mainly chains or two-dimensional images) and extend some of them; in particular we give several equivalent definitions of Markov fields on a hierarchical structure. We show also that the forward-backward algorithm can be applied to the case where a Markov field is defined on a tree whose causal ordering is from the leaves to the root.

Moreover we have adopted a new notation which, according to us, is more compact and less ambiguous than the one used in most works dealing with computations on hidden Markov fields.

In Chapter II we introduce the basic mathematical ground for the general definition of a causal Markovian model on an arbitrary hierarchical structure and analyze the possible definitions of a Markov field and the relations between them.

In Chapter III we apply this framework to several particular cases, and show for example that the forward-backward algorithm can be applied to trees, or to certain fields whose states are vectors of states in the original field (for example in a 2-D image, one can define a Markov chain on the rows, the columns or the diagonals of the image, see [3]).

II. Markov fields on hierarchical structures

This chapter is concerned with various general definitions of a (causal) Markov field or hidden Markov field on arbitrary hierarchical structures, and with the equivalence between these definitions.

II.1. Hierarchical structures

We take a *finite* set V whose elements will be called points. In practice V can be a set of consecutive integers (for a 1-D Markov chain), a rectangular grid consisting of pixels, the set of nodes of a quadtree, or any discrete image structure. As we will define on V a Markov field, we will require an ordering of V together with a neighborhood relation in terms of which we will describe the dependence between points of V. This will be done by endowing V with a precedence relation $\prec$.

This relation consists in a set of ordered pairs (p, q), with $p, q \in V$. Given any such pair (p, q), we will write $p \prec q$ or $q \succ p$ and say that p is a *predecessor* of q or that q is a *successor* of p. As it should be expected from a precedence relation, $\prec$ has no loops nor cycles. In other words:

— for any $p \in V$, $p \not\prec p$;

— for any $p_1, \ldots, p_n \in V$ (with $n \geq 2$), if $p_1 \prec \cdots \prec p_n$, then $p_n \not\prec p_1$.

Here $\not\prec$ means the negation of $\prec$; the definitions of $\not\succ$, $\preceq$, $\succeq$, $\not\preceq$ and $\not\succeq$ are straightforward.

The precedence relation $\prec$ induces on V an ordering relation $<$. For any $p, q \in V$, we write $p < q$ or $q > p$ and say that p is *before* q or that q is *after* p if there exist $p_1, \ldots, p_n \in V$ (with $n \geq 2$) such that $p = p_1 \prec \cdots \prec p_n = q$. The relation $<$ is a *strict order*. In other words, for any $p, q, r \in V$,

— $p \not< p$;

— if $p < q$, then $q \not< p$;

— if $p < q$ and $q < r$, then $p < r$.

Again, the meaning of $\not<$, $\not>$, $\leq$, $\geq$, $\not\leq$ and $\not\geq$ should be clear. Note that we can have $p \not> q$, $p \neq q$ and $p \not< q$ at the same time; thus $\not<$ is not equivalent to $\geq$.

Examples. 1) V is a rectangular grid with pixels labelled (i, j) ($1 \leq i \leq m$, $1 \leq j \leq n$); this labelling is consistent with matrix notation (i.e., the i axis points downwards and the j axis points rightwards).

We have $(i', j') \leq (i, j)$ iff $i' \leq i$ and $j' \leq j$. We have two possible choices for $\prec$, corresponding to the two adjacency relations on V:

a) For the 4-adjacency,
$(i', j') \prec (i, j) \quad \text{iff} \quad (i', j') = (i-1, j) \text{ or } (i, j-1).$

b) For the 8-adjacency,

$(i',j') \prec (i,j)$ iff $(i',j') = (i-1,j)$, $(i,j-1)$ or $(i-1,j-1)$.

2) If V is the set of nodes of a quadtree, then we have two possibilities:

(*a*) $p \prec q$ if p is a child node of q;

(*b*) $p \prec q$ if q is a child node of p.

Here (*a*) corresponds to the case where the quadtree is processed from the bottom to the top (analysis), while (*b*) corresponds to the case where the quadtree is processed from the top to the bottom (synthesis).

3) In speech processing, the sequence of phonemes is considered as a Markov chain. One could envisage the use of a syntactic Markovian model; for example, one can choose V as a set of utterances (phonemes, words or sentences), and write $p \prec q$ if q is a word and p is a phoneme of q, or if q is a sentence and p is a word of q. Here $\prec$ induces on V a tree structure.

Given an element p of V, we can define the following subsets of V:

$$\begin{aligned}
P(p) &\doteq \{q \in V \mid q \prec p\};\\
S(p) &\doteq \{q \in V \mid q \succ p\};\\
PE(p) &\doteq \{q \in V \mid q \preceq p\} = P(p) \cup \{p\};\\
SE(p) &\doteq \{q \in V \mid q \succeq p\} = S(p) \cup \{p\};\\
B(p) &\doteq \{q \in V \mid q < p\};\\
A(p) &\doteq \{q \in V \mid q > p\};\\
BE(p) &\doteq \{q \in V \mid q \leq p\} = B(p) \cup \{p\};\\
AE(p) &\doteq \{q \in V \mid q \geq p\} = A(p) \cup \{p\};\\
NB(p) &\doteq \{q \in V \mid q \not< p\} = V - B(p);\\
NA(p) &\doteq \{q \in V \mid q \not> p\} = V - A(p);\\
NBE(p) &\doteq \{q \in V \mid q \not\leq p\} = NB(p) - \{p\} = V - BE(p);\\
NAE(p) &\doteq \{q \in V \mid q \not\geq p\} = NA(p) - \{p\} = V - AE(p).
\end{aligned}$$

Here the letters P, S, B, A, E and N are mnemotechnical abbreviations for predecessor, successor, before, after, equal and not.

For the example 1) above (the rectangular grid), given a pixel $p = (i,j)$, $BE(p)$ is the set of pixels (i',j') with $i' \leq i$ and $j' \leq j$, $AE(p)$ is the set of pixels (i',j') with $i' \geq i$ and $j' \geq j$, $NBE(p)$ is the set of pixels (i',j') with $i' > i$ or $j' > j$, and $NAE(p)$ is the set of pixels (i',j') with $i' < i$ or $j' < j$. This is illustrated in Figure 1.

Given a subset V' of V, we will say that V' is *before-closed* if for any $x \in V'$, $B(x) \subseteq V'$ (or equivalently if $P(x) \subseteq V'$). The intersection or union of before-closed subsets of V is

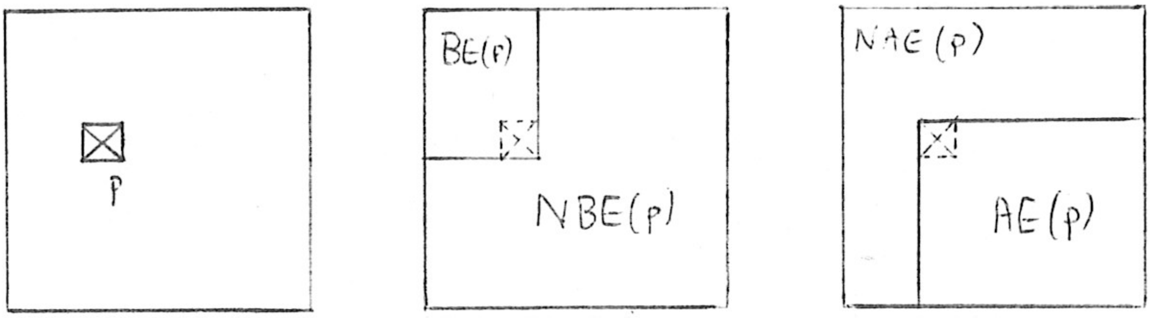

Figure 1. *$BE(p)$, $NBE(p)$, $AE(p)$ and $NAE(p)$ in the rectangular grid.*

before-closed. For example, for any $x \in V$, $BE(x)$ is the smallest before-closed set containing x (in other words, the intersection of all before-closed subsets of V containing x), while $NAE(x)$ is the largest before-closed set not containing x (in other words, the union of all before-closed subsets of V not containing x).

The following result will be useful in the next two sections, because it will allow us to prove by induction several results concerning before-closed subsets of V:

Lemma 1. *Let $k \geq 1$ and let $U_1 \subset \ldots \subset U_k = V$ be non-void before-closed subsets of V. Set $v = |V|$. Then there is a one to one numbering map $n : V \to \{1, \ldots, v\}$ such that for every $p, q \in V$, $n(p) < n(q)$ if $p < q$ or if $p \in U_i$ and $q \notin U_i$ for some i. In particular, if we set $u_i = |U_i|$ $(i = 1, \ldots, k)$, then $n(U_i) = \{1, \ldots, u_i\}$.*

Proof. We use induction on v. If $v = 1$, then the result is trivial. Suppose that $v > 1$ and that the result is true for $v - 1$. As V is finite, $U_k - U_{k-1}$ contains an element x which is maximal for $<$, in other words such that $x \not< y$ for any $y \in U_k - U_{k-1}$; as U_{k-1} is before-closed and $x \notin U_{k-1}$, $x \not< y$ for any $y \in U_{k-1}$. Thus $x \not< y$ for any $y \in V$, in other words $V - \{x\}$ is before-closed. We can now apply induction hypothesis on $V - \{x\}$: there is a one to one numbering map $n : V - \{x\} \to \{1, \ldots, v-1\}$ such that for every $p, q \in V - \{x\}$, $n(p) < n(q)$ if $p < q$ or if $p \in U_i$ and $q \notin U_i$ for some i. We extend then n to V by setting $n(x) = v$, and clearly n is still one to one and for $y \in V - \{x\}$, $n(y) < n(x)$; moreover we cannot have $x < y$ or $x \in U_i$ and $y \notin U_i$ for any i. Thus for every $p, q \in V$, $n(p) < n(q)$ if $p < q$ or if $p \in U_i$ and $q \in U_j$ for $i < j$. Thus the result holds for v too.

For each U_i, as $n(q) > n(p)$ for $p \in U_i$ and $q \in V - U_i$, $n(U_i) = \{1, \ldots, u_i\}$. ∎

II.2. Standard definitions of the Markov field

We assume that we have a set Γ of possible states which can be associated to elements of V. For example, if V is a set of pixels, Γ can be the set of possible grey levels.

Let S be a subset of V, and suppose that to every $x \in S$ we associate the state $\gamma(x) \in \Gamma$. This induces a map $\gamma : S \to \Gamma : x \mapsto \gamma(x)$. For the purpose of our discussion, such a map $\gamma : S \to \Gamma$, associating a state to each element of S, will be called a *configuration* for S. To this map corresponds a probability, namely that each $x \in S$ has state $\gamma(x)$, and we will write it $\mathbf{p}(\gamma)$. The set of all configurations for S will be written $\mathcal{C}[S]$.

Consider n pairwise disjoint subsets $S_1, \ldots, S_n$ of V $(n \geq 2)$, and suppose that to each S_i one associates a configuration γ_i. Then we obtain a joint configuration for $S_1 \cup \cdots \cup S_n$, which will be written $\gamma_1, \ldots, \gamma_n$; its probability will be written $\mathbf{p}(\gamma_1, \ldots, \gamma_n)$.

If T is a subset of S and γ is a configuration for S, then the restriction of γ to T is a configuration for T. We will write it γ_T; when $T = \{x\}$, we will write γ_x for $\gamma_{\{x\}}$.

Note that in [9], the authors give similar definitions and notations for a configuration and the restriction of a configuration to a subset.

Given two events X and Y, the conditional probability of X given Y is written $\mathbf{p}(X/Y)$.

In order to avoid quotients of the form $\frac{0}{0}$ in the conditional probabilities $\mathbf{p}(\lambda_S/\lambda_T)$ $(S, T \subseteq V,\ S \cap T = \emptyset)$, we will assume that $\mathbf{p}(\lambda) \neq 0$ for every $\lambda \in \mathcal{C}[V]$.

In the sequel, the following two results will be widely used:

Lemma 2. *Let A and B be two disjoint subsets of V, and let B' be a subset of B. Consider a configuration α for A. If for every configuration β for B, $\mathbf{p}(\alpha/\beta)$ is a function of α and $\beta_{B'}$ only, then $\mathbf{p}(\alpha/\beta) = \mathbf{p}(\alpha/\beta_{B'})$.*

Proof. Let β be a configuration for B. Given $\gamma \in \mathcal{C}[B - B']$, then the joint configuration $\beta_{B'}, \gamma \in \mathcal{C}[B]$ and so there is a function f such that

$$\mathbf{p}(\alpha/\beta_{B'}, \gamma) = \mathbf{p}(\alpha/\beta) = f(\alpha, \beta_{B'}).$$

We get then:

$$\begin{aligned}\mathbf{p}(\alpha/\beta_{B'}) &= \sum_{\gamma \in \mathcal{C}[B-B']} \mathbf{p}(\alpha, \gamma/\beta_{B'}) = \sum_{\gamma \in \mathcal{C}[B-B']} \mathbf{p}(\alpha/\gamma, \beta_{B'}) \cdot \mathbf{p}(\gamma/\beta_{B'}) \\ &= \sum_{\gamma \in \mathcal{C}[B-B']} f(\alpha, \beta_{B'}) \cdot \mathbf{p}(\gamma/\beta_{B'}) = f(\alpha, \beta_{B'}) \cdot \sum_{\gamma \in \mathcal{C}[B-B']} \mathbf{p}(\gamma/\beta_{B'}) = f(\alpha, \beta_{B'}).\end{aligned}$$

Thus $\mathbf{p}(\alpha/\beta_{B'}) = f(\alpha, \beta_{B'}) = \mathbf{p}(\alpha/\beta)$. ∎

Corollary 3. *Let A and B be two disjoint subsets of V, and let $B'' \subseteq B' \subseteq B$. Consider a configuration α for A. If for any configuration β for B, $\mathbf{p}(\alpha/\beta) = \mathbf{p}(\alpha/\beta_{B''})$, then $\mathbf{p}(\alpha/\beta) = \mathbf{p}(\alpha/\beta_{B'})$.*

This follows immediately from the fact that $\mathbf{p}(\alpha/\beta_{B''})$ is a function of α and $\beta_{B'}$ only.

We will now define a Markov field on V. We will generalize the definitions made by Abend, Harley and Kanal in the case of the rectangular grid [1]. Consider the following three hypotheses:

μ : For every $x \in V$ and every configuration λ for V,

$$\mathbf{p}(\lambda_x/\lambda_{NAE(x)}) = \mathbf{p}(\lambda_x/\lambda_{P(x)}).$$

μ' : For every before-closed subset U of V and every configuration λ for V,

$$\mathbf{p}(\lambda_U) = \prod_{z\in U} \mathbf{p}(\lambda_z/\lambda_{P(z)}).$$

μ^* : For every $x \in V$ and every configuration λ for V,

$$\mathbf{p}(\lambda_{BE(x)}) = \prod_{z\in BE(x)} \mathbf{p}(\lambda_z/\lambda_{P(z)}).$$

Here μ and μ^* correspond to equations (14) and (15) in [1]. As every $BE(x)$ ($x \in V$) is before-closed, μ^* is a particular case of μ'. It was shown in Theorem 1 of [1] that μ implies μ^* in the case of the rectangular grid; however we can prove the following more general result:

Theorem 4. *μ is equivalent to μ'.*

Proof. 1) $\underline{\mu \text{ implies } \mu'}$:
Let $u = |U|$ and let n be the numbering map defined in Lemma 1, with $U_1 = U$ and $n(U) = \{1,\dots,u\}$. The elements of V can be labelled $z_1,\dots,z_v$, where $n(z_i) = i$ for $i = 1,\dots,v$. Then $U = \{z_1,\dots,z_u\}$. For each $i = 1,\dots,u$ we set

$$S_i = \{z_j \mid j < i\}.$$

We recall that the map n gives $n(p) < n(q)$ for $p < q$. Let $1 \le i \le u$. Then

$$P(z_i) \subseteq S_i \subseteq NAE(z_i). \tag{1}$$

Indeed, for $z_h \in P(z_i)$, $z_h < z_i$ and so $h < i$; thus $P(z_i) \subseteq \{z_h \mid h < i\} = S_i$. On the other hand, for $j < i$, $z_j \not\geq z_i$, otherwise we would have $n(z_j) \geq n(z_i)$; thus $S_i = \{z_j \mid j < i\} \subseteq NAE(z_i)$. This justifies (1). Now μ implies that

$$\mathbf{p}(\lambda_{z_i}/\lambda_{NAE(z_i)}) = \mathbf{p}(\lambda_{z_i}/\lambda_{P(z_i)}).$$

By (1) and Corollary 3, this implies that

$$\mathbf{p}(\lambda_{z_i}/\lambda_{S_i}) = \mathbf{p}(\lambda_{z_i}/\lambda_{P(z_i)}). \tag{2}$$

Using (2), we obtain then the following by adding successively $z_1,\dots,z_u$ to $\emptyset$:

$$\mathbf{p}(\lambda_U) = \prod_{i=1}^{u} \mathbf{p}(\lambda_{z_i}/\lambda_{S_i}) = \prod_{i=1}^{u} \mathbf{p}(\lambda_{z_i}/\lambda_{P(z_i)}).$$

2) μ' implies μ:

Take $x \in V$. Given $y, z \in V$ with $y < z$, by the transitivity of $\geq$ and $>$, $z \not> x$ implies that $y \not> x$, and $z \not\geq x$ implies that $y \not\geq x$. Thus $NA(x)$ and $NAE(x)$ are before-closed. By μ', for any configuration λ for V we have:

$$\mathbf{p}(\lambda_{NA(x)}) = \prod_{z \in NA(x)} \mathbf{p}(\lambda_z/\lambda_{P(z)});$$
$$\mathbf{p}(\lambda_{NAE(x)}) = \prod_{z \in NAE(x)} \mathbf{p}(\lambda_z/\lambda_{P(z)}).$$

As $NA(x) = NAE(x) \cup \{x\}$, we get then:

$$\mathbf{p}(\lambda_x/\lambda_{NAE(x)}) = \frac{\mathbf{p}(\lambda_{NA(x)})}{\mathbf{p}(\lambda_{NAE(x)})} = \frac{\prod_{z \in NA(x)} \mathbf{p}(\lambda_z/\lambda_{P(z)})}{\prod_{z \in NAE(x)} \mathbf{p}(\lambda_z/\lambda_{P(z)})} = \mathbf{p}(\lambda_x/\lambda_{P(x)}).$$

Thus μ and μ' are equivalent. ∎

Let us now build a particular window around elements of V. For $x \in V$, we define

$$\begin{aligned} W(x) &\doteq \{z \in V \mid \exists y \in V, z \preceq y \succeq x\}. \\ &= \bigcup_{y \in SE(x)} PE(y). \end{aligned}$$

In other words, $W(x)$ is the union of all sets $PE(y)$ $(y \in V)$ which contain x. For example, in the case of the rectangular grid, with the two choices of $\prec$ corresponding to the 4- and 8-adjacencies respectively (see Section I), $W(x)$ takes the two forms shown in Figure 2 (see also Figure 3 (a,b) of [1]).

$$\begin{array}{ccc} & * & * \\ * & x & * \\ * & * & \end{array} \qquad\qquad\qquad \begin{array}{ccc} * & * & * \\ * & x & * \\ * & * & * \end{array}$$

4-adjacency 8-adjacency

Figure 2. *The window $W(x)$ in the rectangular grid.*

Other windows $W(x)$ based on different choices of $P(x)$ for the rectangular grid are shown in Figure 3 (c,d,e) of [1]. It was shown in this particular case that μ^* implies the following (see equation (16) of [1]):

μ° : For every $x \in V$ and every configuration λ for V,

$$\mathbf{p}(\lambda_x/\lambda_{V-\{x\}}) = \mathbf{p}(\lambda_x/\lambda_{W(x)-\{x\}}).$$

This result can be generalized as follows:

Theorem 5. *Suppose that either μ' holds, or μ^* holds and V contains some z such that $y < z$ for every $y \in V - \{z\}$ (in other words $V = BE(z)$). Then μ° holds.*

Proof. For every configuration λ' for V, we have

$$\mathrm{p}(\lambda') = \prod_{y \in V} \mathrm{p}(\lambda'_y / \lambda'_{P(y)}).$$

This follows from μ', and from μ^* if we have $V = BE(z)$. Let $x \in V$. Let λ be a configuration for V and let Λ be the set of all configurations λ' for V such that $\lambda'_{V-\{x\}} = \lambda_{V-\{x\}}$. Then

$$\mathrm{p}(\lambda_{V-\{x\}}) = \sum_{\lambda' \in \Lambda} \mathrm{p}(\lambda')$$

and so

$$\mathrm{p}(\lambda_x / \lambda_{V-\{x\}}) = \frac{\mathrm{p}(\lambda)}{\sum_{\lambda' \in \Lambda} \mathrm{p}(\lambda')} = \frac{\prod_{y \in V} \mathrm{p}(\lambda_y / \lambda_{P(y)})}{\sum_{\lambda' \in \Lambda} \prod_{y \in V} \mathrm{p}(\lambda'_y / \lambda'_{P(y)})}.$$

Now when $y \notin SE(x)$, $x \notin PE(y)$, in other words $\lambda'_y = \lambda_y$ and $\lambda'_{P(y)} = \lambda_{P(y)}$; but then $\mathrm{p}(\lambda'_y / \lambda'_{P(y)}) = \mathrm{p}(\lambda_y / \lambda_{P(y)})$. Thus these terms cancel out in the above expression, and we get

$$\mathrm{p}(\lambda_x / \lambda_{V-\{x\}}) = \frac{\prod_{y \in SE(x)} \mathrm{p}(\lambda_y / \lambda_{P(y)})}{\sum_{\lambda' \in \Lambda} \prod_{y \in SE(x)} \mathrm{p}(\lambda'_y / \lambda'_{P(y)})}.$$

The right-hand expression depends only upon the restriction of λ to $\bigcup_{y \in SE(x)} PE(y) = W(x)$. By Lemma 2, this implies that $\mathrm{p}(\lambda_x / \lambda_{V-\{x\}}) = \mathrm{p}(\lambda_x / \lambda_{W(x)-\{x\}})$. ∎

Theorems 4 and 5 are valid for arbitrary hierarchical structures. It is thus possible to apply them to some substructures of V when V satisfies μ or μ'. We have the following:

Proposition 6. *Let V' be a before-closed subset of V. Consider the restriction of $\prec$ to V'. If V satisfies μ, μ' or μ^*, then V' satisfies it.*

Proof. Let $x \in V'$. We have

$$P(x) \subseteq NAE(x) \cap V' \subseteq NAE(x). \tag{3}$$

Suppose that V satisfies μ. Then for every configuration λ,

$$P(\lambda_x / \lambda_{NAE(x)}) = P(\lambda_x / \lambda_{P(x)}).$$

By Corollary 3 and (3), this implies that

$$P(\lambda_x/\lambda_{NAE(x)\cap V'}) = P(\lambda_x/\lambda_{P(x)}),$$

and so V' satisfies μ.

It is obvious that if U is a before-closed subset of V', then U is before-closed in V. Thus V' satisfies μ' if V satisfies it.

As $BE(x) \subseteq V'$, it is obvious that if V satisfies μ^*, then V' satisfies it too. ∎

Before-closed subsets of V will intervene several times the next chapter, for example in the analysis of the case where $\prec$ induces on V a tree structure.

II.3. Separators and new definitions of the Markov field

The concept of a *separating set*, which will be introduced below, is an important one. From a theoretical point of view, it leads to equivalent definitions of the Markov field. For practical purposes, it will be used in the definition of Markov chains on some subsets of V and to the application of the forward-backward method to trees (see Chapter III).

Given two disjoint subsets S and T of V and a subset R of S, we say that R separates S from T if for every $z \in T$, $P(z) \cap S \subseteq R$. We call R a *separator* of S if R separates S from $V - S$.

For example, for every $x \in V$, $P(x)$ separates $NAE(x)$ from x. Now μ states that for every $x \in V$ and for every configuration λ for V, $\mathrm{p}(\lambda_x/\lambda_{NAE(x)}) = \mathrm{p}(\lambda(x)/\lambda_{P(x)})$. In fact, $NAE(x)$ and $NA(x) = NAE(x) \cup \{x\}$ are before-closed, and an interesting fact is that this equality can be generalized to separating subsets in before-closed sets:

μ_S : Let S and T be two disjoint subsets of V such that S and $S \cup T$ are before-closed, and let R be a subset of S separating S from T. Then for every configuration λ for V,

$$\mathrm{p}(\lambda_T/\lambda_S) = \mathrm{p}(\lambda_T/\lambda_R).$$

We have indeed the following:

Theorem 7. *μ is equivalent to μ_S.*

Proof. As μ is a particular case of μ_S (with $S = NAE(x)$, $T = \{x\}$, and $R = P(x)$), we have only to show that μ implies μ_S. Let $s = |S|$ and $t = |T|$. Consider the numbering map n defined in Lemma 1, with $U_1 = S$ and $U_2 = S \cup T$. We can label the elements of V as $x_1, \ldots, x_v$, where $n(x_i) = i$ for $i = 1, \ldots, v$. Then the elements of S are $x_1, \ldots, x_s$ and those of T are $x_{s+1}, \ldots, x_{s+t}$. For each $i = s+1, \ldots, s+t$, we define

$$S_i = \{x_j \mid 1 \le j < i\};$$
$$R_i = S_i \cap (T \cup R) = R \cup \{x_j \mid s < j < i\}.$$

We recall that the map n gives $n(p) < n(q)$ for $p < q$. Let $i = s+1, \ldots, s+t$. For $x_j \in P(x_i)$ we must have $j < i$; thus $P(x_i) \subseteq S_i$. Now as $x_i \in T$, $P(x_i) \cap S \subseteq R$, in other words $P(x_i) \subseteq T \cup R$; thus $P(x_i) \subseteq S_i \cap (T \cup R) = R_i$. Finally, it is clear that for $j < i$, $x_j \not\geq x_i$; thus $S_i \subseteq NAE(x_i)$. Hence:

$$P(x_i) \subseteq R_i \subseteq S_i \subseteq NAE(x_i).$$

By μ and Corollary 3, this implies that

$$\mathbf{p}(\lambda_{x_i}/\lambda_{R_i}) = \mathbf{p}(\lambda_{x_i}/\lambda_{S_i}). \tag{4}$$

By adding successively $x_{s+1}, \ldots, x_{s+t}$ to S, we obtain

$$\mathbf{p}(\lambda_T/\lambda_S) = \prod_{i=s+1}^{s+t} \mathbf{p}(\lambda_{x_i}/\lambda_{S_i}). \tag{5}$$

But by adding successively $x_{s+1}, \ldots, x_{s+t}$ to R, we obtain

$$\mathbf{p}(\lambda_T/\lambda_R) = \prod_{i=s+1}^{s+t} \mathbf{p}(\lambda_{x_i}/\lambda_{R_i}). \tag{6}$$

Combining $(4, 5, 6)$, we get $\mathbf{p}(\lambda_T/\lambda_S) = \mathbf{p}(\lambda_T/\lambda_R)$. ∎

A major consequence of this result is the following:

Corollary 8. *Let S and T be two disjoint before-closed subsets of V. Then μ implies that*

$$\mathbf{p}(\lambda_{S \cup T}) = \mathbf{p}(\lambda_S) \cdot \mathbf{p}(\lambda_T)$$

for every configuration λ for V.

Proof. Clearly $S \cup T$ is before-closed and the empty set $\emptyset$ separates S from T. Thus Theorem 7 implies that

$$\frac{\mathbf{p}(\lambda_T, \lambda_S)}{\mathbf{p}(\lambda_S)} = \mathbf{p}(\lambda_T/\lambda_S) = \mathbf{p}(\lambda_T/\lambda_\emptyset) = \mathbf{p}(\lambda_T),$$

and the result follows then by multiplying the first and last member of that equality by $\mathbf{p}(\lambda_S)$. ∎

We will now define another Markov field hypothesis which, like μ_S, is equivalent to μ, but is nevertheless an extension of μ to some subsets of V. This will require the extension of some definitions made in Section II.1.

Given $X \subseteq V$, we define:

$$PE(X) \doteq \{y \in V \mid \exists x \in X, y \preceq x\} = \bigcup_{x \in X} PE(x);$$
$$BE(X) \doteq \{y \in V \mid \exists x \in X, y \leq x\} = \bigcup_{x \in X} BE(x);$$
$$NBE(X) \doteq \{y \in V \mid \forall x \in X, y \nleq x\} = \bigcap_{x \in X} NBE(x) = V - BE(X);$$
$$AE(X) \doteq \{y \in V \mid \exists x \in X, y \geq x\} = \bigcup_{x \in X} AE(x);$$
$$NAE(X) \doteq \{y \in V \mid \forall x \in X, y \ngeq x\} = \bigcap_{x \in X} NAE(x) = V - AE(X).$$

It is easy to see that $BE(X)$ is the smallest before-closed subset of V containing X, while $NAE(X)$ is the largest before-closed subset of V disjoint from X.

We have now the following:

Lemma 9. *Let X be a non-void subset of V. Then the following conditions are equivalent:*

(i) *There does not exist $x, x' \in X$ and $y \in V - X$ such that $x < y < x'$.*

(ii) $BE(X) - X \subseteq NAE(X)$

(iii) $BE(X) \subseteq NAE(X) \cup X$

(iv) *$BE(X) - X$ is before-closed.*

(v) *$NAE(X) \cup X$ is before-closed.*

(vi) *There exist two before-closed subsets U and U' of V such that $U \subset U'$, $X \subseteq U' - U$ and $BE(X) - X \subseteq U$.*

Proof. (i) implies (ii):
Given $y \in BE(X) - X$, there is some $x' \in X$ such that $y < x'$. For any $x \in X$, $x \neq y$ (since $y \notin X$), and $x \nless y$, otherwise $x < y < x'$ and (i) is contradicted. Thus $x \nleq y$, in other words $y \in NAE(X)$.

(ii) implies (iii):
This follows by adding X to each member of (ii).

(iii) implies (v):
Let $y \in NAE(X) \cup X$ and $y' < y$. If $y \in X$, then $y' \in BE(X)$, and as $BE(X) \subseteq NAE(X) \cup X$, $y' \in NAE(X) \cup X$. If $y \in NAE(X)$, then $y' \in NAE(X)$, since $NAE(X)$ is before-closed. Thus $y' \in NAE(X) \cup X$ in any case, in other words $NAE(X) \cup X$ is before-closed.

(*v*) implies (*vi*):

We take $U = NAE(X)$ and $U' = NAE(X) \cup X$. As $X \subseteq U'$ and U' is before-closed, $BE(X) \subseteq U'$ and so $BE(X) - X \subseteq U' - X = U$.

(*vi*) implies (*iv*):

Let $y \in BE(X) - X$ and $y' < y$. Then clearly $y' \in BE(X)$, since $BE(X)$ is before-closed. Similarly, as $BE(X) - X \subseteq U$ and U is before-closed, $y' \in U$. Thus $y' \in BE(X) \cap U \subseteq BE(X) - (U' - U)$, and as $X \subseteq U - U'$, $y' \in BE(X) - X$. Thus $BE(X) - X$ is before-closed.

(*iv*) implies (*i*):

If $y < x'$ for $x' \in X$ and $y \notin X$, then $y \in BE(X) - X$. Then for $x < y$, $x \in BE(X) - X$, since $BE(X) - X$ is before-closed; in other words $x \notin X$. Thus (*i*) holds. ∎

A set X satisfying the conditions of Lemma 9 will be called *regular*. For example, in a two-dimensional grid, rows and columns are regular. For a regular set X we define:

$$P(X) \doteq PE(X) - X;$$
$$B(X) \doteq BE(X) - X;$$
$$NA(X) \doteq NAE(X) \cup X.$$

By (*iv*) and (*v*), $B(X)$ and $NA(X)$ are before-closed (while $BE(Y)$ and $NAE(Y)$ are before-closed for any $Y \subseteq V$). Now $P(X)$ separates $NAE(X)$ from X. Therefore μ implies the following by Theorem 7:

μ_R : For every regular subset X of V for every configuration λ for V,

$$\mathbf{p}(\lambda_X / \lambda_{NAE(X)}) = \mathbf{p}(\lambda_X / \lambda_{P(X)}).$$

But for every $x \in V$, x is regular, and so μ_R yields μ. We have thus the following:

Proposition 10. *μ is equivalent to μ_R.*

II.4. The hidden Markov field under memoryless noise

Let us consider the following situation: the states in Γ corresponding to the elements of V are not known, but we have a set Δ of observable states, and for every $x \in V$ one measures a corresponding state $\xi(x) \in \Delta$. One has thus a map $\xi : S \to \Delta : x \mapsto \xi(x)$, which for the purpose of the present discussion will be called an *observation*. Our goal is to find a highly probable configuration λ corresponding to the observation ξ. (In practice, we will attempt to maximize the probability $\mathbf{p}(\lambda_x / \xi)$ for each $x \in V$.) For this purpose, we assume that λ satisfies the Markov field hypothesis μ (or μ'), and that the transition probabilities $\mathbf{p}(\lambda_x / \lambda_{P(x)})$ and the dependence $\mathbf{p}(\xi / \lambda)$ are known for any configuration λ.

The configurations λ form thus what one calls a *hidden* Markov field. A particular case is when $\Delta = \Gamma$, and so ξ can be seen as the result of the corruption of λ by noise.

However this not the only possibility. For example, Δ can be a set of possible grey levels and ξ can be the measurement of grey levels of pixels on a rectangular grid; then Γ can be a set of textures which can be assigned to pixels by λ, or we can have $\Gamma = \{0,1\}$ and λ consists then in a binary image extracted from ξ by some form of probabilistic segmentation.

We will make the simplifying assumption that the dependence $\mathbf{p}(\xi/\lambda)$ corresponds to a memoryless transformation. This can be expressed by the following hypothesis:

ν : For every $D \subset C \subseteq V$ and $x \in C - D$ and for every configuration λ for V,

$$\mathbf{p}(\xi_x/\xi_D, \lambda_C) = \mathbf{p}(\xi_x/\lambda_x).$$

We will then consider the following apparently stronger hypothesis:

ν' : For every $S \subseteq T \subseteq V$ and $R \subseteq V$ such that $R \cap S = \emptyset$ and for every configuration λ for V,

$$\mathbf{p}(\xi_S/\xi_R, \lambda_T) = \prod_{z \in S} \mathbf{p}(\xi_z/\lambda_z) = \mathbf{p}(\xi_S/\lambda_S).$$

Proposition 11. *ν is equivalent to ν'.*

Proof. 1) $\underline{\nu \text{ implies } \nu'}$:
We use induction on the size of S. If $S = \emptyset$, then ν' is trivial. If $S = \{x\}$ for some $x \in T$, let $R' = R \setminus T = R - R \cap T$. Then we have

$$\mathbf{p}(\xi_x/\xi_R, \lambda_T) = \sum_{\phi_{R'} \in \mathcal{C}[R']} \mathbf{p}(\xi_x/\xi_R, \phi_{R'}, \lambda_T) \cdot \mathbf{p}(\phi_{R'}/\xi_R, \lambda_T). \tag{7}$$

Now by applying ν with $C = R' \cup T = R \cup T$ and $D = R$, we have $\mathbf{p}(\xi_x/\xi_R, \phi_{R'}, \lambda_T) = \mathbf{p}(\xi_x/\lambda_x)$, and so (7) becomes:

$$\begin{aligned}\mathbf{p}(\xi_x/\xi_R, \lambda_T) &= \sum_{\phi_{R'} \in \mathcal{C}[R']} \mathbf{p}(\xi_x/\lambda_x) \cdot \mathbf{p}(\phi_{R'}/\xi_R, \lambda_T) \\ &= \mathbf{p}(\xi_x/\lambda_x) \cdot \sum_{\phi_{R'} \in \mathcal{C}[R']} \mathbf{p}(\phi_{R'}/\xi_R, \lambda_T) = \mathbf{p}(\xi_x/\lambda_x).\end{aligned}$$

Thus ν' holds for $S = \{x\}$. Suppose finally that $|S| > 1$ and that the result is true for any proper subset S^* of S and any $R^* \subseteq V$ such that $R^* \cap S^* = \emptyset$. Let $x \in S$ and let $S' = S - \{x\}$. We have:

$$\begin{aligned}\mathbf{p}(\xi_S/\xi_R, \lambda_T) &= \mathbf{p}(\xi_{S'}, \xi_x/\xi_R, \lambda_T) = \mathbf{p}(\xi_{S'}/\xi_x, \xi_R, \lambda_T) \cdot \mathbf{p}(\xi_x/\xi_R, \lambda_T); \\ &= \mathbf{p}(\xi_{S'}/\xi_x, \xi_R, \lambda_T) \cdot \mathbf{p}(\xi_x/\lambda_x) \qquad \text{(by } \nu' \text{ with } S^* = \{x\} \text{ and } R^* = R); \\ &= \Big(\prod_{z \in S'} \mathbf{p}(\xi_z/\lambda_z)\Big) \cdot \mathbf{p}(\xi_x/\lambda_x) \qquad \text{(by } \nu' \text{ with } S^* = S' \text{ and } R^* = R \cup \{x\}); \\ &= \prod_{z \in S} \mathbf{p}(\xi_z/\lambda_z).\end{aligned}$$

We can apply the same decomposition to $\mathrm{p}(\xi_S/\lambda_S)$, and we obtain then the same result. Thus $\mathrm{p}(\xi_S/\lambda_S) = \mathrm{p}(\xi_S/\xi_R, \lambda_S)$ and ν' holds for S.

2) <u>ν' implies ν</u>:
This results from taking $S = \{x\}$, $R = D$ and $T = C$. ∎

Now ν' implies an interesting result, which will be applied several times in the next chapter:

Proposition 12. *Let S and T be two disjoint subsets of V, and let $R \subseteq S$. Suppose that the observation ξ satisfies ν' and that for every configuration λ for V, $\mathrm{p}(\lambda_T/\lambda_S) = \mathrm{p}(\lambda_T/\lambda_R)$. Then, given $L, X \subseteq S$ and $M, Y \subseteq T$ such that $R \subseteq L$, for every configuration λ for V we have $\mathrm{p}(\lambda_M, \xi_Y/\lambda_L, \xi_X) = \mathrm{p}(\lambda_M, \xi_Y/\lambda_R)$.*

Proof. We set $A = S - L$ and $B = T - M$. Then for every $\theta_B \in \mathcal{C}[B]$ and $\theta_A \in \mathcal{C}[A]$ we have by hypothesis

$$\mathrm{p}(\lambda_M, \theta_B/\lambda_L, \theta_A) = \mathrm{p}(\lambda_M, \theta_B/\lambda_R), \tag{8}$$

and ν' implies that

$$\mathrm{p}(\xi_Y, \xi_X/\lambda_M, \theta_B, \lambda_L, \theta_A) = \mathrm{p}(\xi_Y/\lambda_M, \theta_B, \lambda_R) \cdot \mathrm{p}(\xi_X/\lambda_L, \theta_A). \tag{9}$$

Applying (8) and (9) we get:

$$\begin{aligned}
\mathrm{p}(\lambda_M, \xi_Y, \lambda_L, \xi_X) &= \sum_{\theta_B \in \mathcal{C}[B]} \sum_{\theta_A \in \mathcal{C}[A]} \mathrm{p}(\lambda_M, \theta_B, \xi_Y, \lambda_L, \theta_A, \xi_X); \\
&= \sum_{\theta_B \in \mathcal{C}[B]} \sum_{\theta_A \in \mathcal{C}[A]} \mathrm{p}(\lambda_M, \theta_B, \lambda_L, \theta_A) \cdot \mathrm{p}(\xi_Y, \xi_X/\lambda_M, \theta_B, \lambda_L, \theta_A); \\
&= \sum_{\theta_B \in \mathcal{C}[B]} \sum_{\theta_A \in \mathcal{C}[A]} \mathrm{p}(\lambda_M, \theta_B, \lambda_L, \theta_A) \cdot \mathrm{p}(\xi_Y/\lambda_M, \theta_B, \lambda_R) \cdot \mathrm{p}(\xi_X/\lambda_L, \theta_A); \\
&= \sum_{\theta_B \in \mathcal{C}[B]} \sum_{\theta_A \in \mathcal{C}[A]} \mathrm{p}(\lambda_M, \theta_B/\lambda_L, \theta_A) \cdot \mathrm{p}(\lambda_L, \theta_A) \cdot \mathrm{p}(\xi_Y/\lambda_M, \theta_B, \lambda_R) \cdot \mathrm{p}(\xi_X/\lambda_L, \theta_A); \\
&= \sum_{\theta_B \in \mathcal{C}[B]} \sum_{\theta_A \in \mathcal{C}[A]} \mathrm{p}(\lambda_M, \theta_B/\lambda_R) \cdot \mathrm{p}(\lambda_L, \theta_A) \cdot \mathrm{p}(\xi_Y/\lambda_M, \theta_B, \lambda_R) \cdot \mathrm{p}(\xi_X/\lambda_L, \theta_A); \\
&= \sum_{\theta_B \in \mathcal{C}[B]} \mathrm{p}(\lambda_M, \theta_B/\lambda_R) \cdot \mathrm{p}(\xi_Y/\lambda_M, \theta_B, \lambda_R) \cdot \sum_{\theta_A \in \mathcal{C}[A]} \mathrm{p}(\lambda_L, \theta_A) \cdot \mathrm{p}(\xi_X/\lambda_L, \theta_A); \\
&= \sum_{\theta_B \in \mathcal{C}[B]} \mathrm{p}(\lambda_M, \theta_B, \xi_Y/\lambda_R) \cdot \sum_{\theta_A \in \mathcal{C}[A]} \mathrm{p}(\lambda_L, \theta_A, \xi_X) = \mathrm{p}(\lambda_M, \xi_Y/\lambda_R) \cdot \mathrm{p}(\lambda_L, \xi_X).
\end{aligned}$$

Dividing the first and last member of that equation by $\mathrm{p}(\lambda_L, \xi_X)$, we get the desired result. ∎

Note that if we assume μ and ν, then we can combine Proposition 12 with Theorem 7 and Corollary 8:

Corollary 13. *Assume that V satisfies μ and ν. Let S and T be two disjoint subsets of V such that S and $S \cup T$ are before-closed, and let R be a subset of S separating S from T. Then, given $L, X \subseteq S$ and $M, Y \subseteq T$ such that $R \subseteq L$, we have*

$$\mathbf{p}(\lambda_M, \xi_Y / \lambda_L, \xi_X) = \mathbf{p}(\lambda_M, \xi_Y / \lambda_R)$$

for every configuration λ for V.

Corollary 14. *Assume that V satisfies μ and ν. Let S and T be two disjoint before-closed subsets of V. Then, given $L, X \subseteq S$ and $M, Y \subseteq T$, we have*

$$\mathbf{p}(\lambda_{L \cup M}, \xi_{X \cup Y}) = \mathbf{p}(\lambda_M, \xi_Y) \cdot \mathbf{p}(\lambda_L, \xi_X)$$

for every configuration λ for V.

III. Applications

We will now assume that V satisfies the hypotheses μ (or equivalently μ', μ_S, μ_R) and ν (or equivalently ν'), and we will examine some methods for finding a probable configuration λ corresponding to the observation ξ. In fact, as we said at the beginning of Section II.4, we will attempt to maximize for each point x the probability of λ_x given ξ. This is opposed to "Viterbi type" methods, where one chooses the most probable λ given ξ. As explained in [5], there are several possible methods for choosing λ_x for each $x \in V$ when ξ is known:

(1°) *Closest value method* [8]
We choose λ_x such that $\mathbf{p}(\xi_x, \lambda_x)$ is maximum.

(2°) *Block constraint method* [8]
We choose λ_x such that $\mathbf{p}(\xi_{PE(x)}, \lambda_x)$ is maximum.

(3°) *Sequential compound method* [4,5]
We choose λ_x such that $\mathbf{p}(\xi_{BE(x)}, \lambda_x)$ is maximum.

(4°) *One step look-ahead method* [4,5]
Given $U(x) = BE(W(x))$ (the smallest before-closed set containing $W(x)$), we choose λ_x such that $\mathbf{p}(\xi_{U(x)}, \lambda_x)$ is maximum.

(5°) *Global method* [4,5]
We choose λ_x such that $\mathbf{p}(\xi_V, \lambda_x)$ is maximum.

Let us give a few comments on this. Method (1°) does not use the context, and according to [8] it should represent an upper error bound for the choice of λ. Method (2°) takes into account a small part of the context and can be implemented in linear time for any hierarchical structure. It can be generalized in order to incorporate a wider context; for example one can take λ_x in such a way that one maximizes $\mathbf{p}(\xi_{PE^n(x)}, \lambda_x)$, where we define $PE^1(x) = PE(x)$ and for $n > 1$,

$$PE^n(x) = PE(PE^{n-1}(x)) = \{y \in V \mid \exists z_1, \ldots, z_{n-1} \in V, y \preceq z_1 \preceq \cdots \preceq z_{n-1} \preceq x\}.$$

In the case where V is a rectangular grid, this is called in [8] the "(n, n) block constraint".

Method (3°) can be implemented with an exponential time complexity for an arbitrary hierarchical structure, but this complexity reduces to a linear one in the case of a Markov field in one dimension (*i.e.*, a Markov chain) [2], and even for a 2-D rectangular grid (provided that one makes certain additional assumptions [5]). Method (4°) takes into account the fact that $\mathbf{p}(\lambda_x/\lambda_{V-\{x\}}) = \mathbf{p}(\lambda_x/\lambda_{W(x)-\{x\}})$ for any configuration λ (see $\mu°$). Its complexity is essentially the same as that of (3°).

Method (5°) has a linear time complexity for a Markov chain, thanks to Baum's "forward-backward" algorithm [2,10]. In the case of a 2-D rectangular grid, to our knowledge no method exists for reducing the time complexity from exponential to linear [5].

Baum's forward-backward method was originally defined in the case of (1-D) Markov chains, but we will show that it can be applied to Markov fields on tree structures. Moreover, given any Markov field, Markov chains can be built on certain partitions of the hierarchical structure, for example on the set of rows, columns or secondary diagonals (that is, diagonals along the SW-NE direction) in a two-dimensional image [3]; here the states in a configuration are vectors of states in the original configuration. Thus the forward-backward method can also applied to these chains.

These partial results will be integrated in a general framework.

III.1. Markov chains built on sets of points

We will show how the hierarchical structure and two hypotheses μ and ν defined on V can be extended to certain types of partitions $V^* = \{V_1, \ldots, V_n\}$ of subsets of V (for example in a two-dimensional rectangular grid, the set of all rows, of all columns or all secondary diagonals, as in [3]), in such a way that V^* will be a first order hidden Markov chain. In this way it will be possible to apply to V^* the forward-backward method.

Let us describe our choice for V^*:

Proposition 15. *Let V be partitioned into k (non-void and pairwise disjoint) sets $V_1, \ldots, V_k$ such that for each $i = 1, \ldots, k$, $PE(V_i) \subseteq V_i \cup V_{i-1}$, with $V_0 = \emptyset$. For each $i = 1, \ldots, k$, set*

$$U_i \doteq \bigcup_{j=1}^{i} V_j$$

and set $U_0 = \emptyset$. Then for $i = 1, \ldots, k$,

- *(i) V_i is regular;*
- *(ii) V_i is a separator of U_i;*
- *(iii) U_i is before-closed.*

Setting $V^* = \{V_1, \ldots, V_k\}$, we have the following:

- *(iv) V^* is endowed with a hierarchical structure determined by the precedence relation $V_{i-1} \prec V_i$ for $i = 2, \ldots, k$. Every configuration λ for V induces a configuration λ^* for V^*, and an observation ξ for V induces an observation ξ^* for V^*.*
- *(v) If V satisfies μ, then V^* is a (one-dimensional) first order Markov chain.*
- *(vi) If V and ξ satisfy ν, then V^* and ξ^* satisfy it.*

Proof. (*i*): Given $x, x' \in V_i$ and $y' \in V - V_i$, $y < x'$ implies that $y \in PE(V_i) - V_i$, in other words that $i > 1$ and $y \in V_{i-1}$; but then $P(y) \subseteq PE(V_{i-1}) \subseteq V_{i-1} \cup V_{i-2}$, which is disjoint from V_i; this means then that $x \not< y'$. Thus V_i is regular by Lemma 9(*i*).

(*ii*): If $x \notin U_i$, then $x \in V_j$ for some $j > i$, and so $P(x) \subseteq PE(V_j) \subseteq V_j \cup V_{j-1}$. Thus $P(x) \cap V_r = \emptyset$ for $r < i$, in other words $P(x) \cap U_i \subseteq V_i$. Hence V_i is a separator of U_i.

(iii): If $x \in U_i$, then $x \in V_j$ for some $j \leq i$, and $P(x) \subseteq V_j \cup V_{j-1} \subseteq U_i$. Thus U_i is before-closed.

(iv) is obvious. Given $S^* \subseteq V^*$, it is induced by some $S \subseteq V$, where S is the union of all $V_i \in S^*$; then for a configuration λ for V, we define naturally $\mathbf{p}(\lambda^*_{S^*}) \doteq \mathbf{p}(\lambda_S)$, and similarly for the observation ξ we define $\mathbf{p}(\xi^*_{S^*}) \doteq \mathbf{p}(\xi_S)$.

(v): Assume that V satisfies μ. It is clear that for $i = 1, \ldots, k$,

$$P(V_i) \subseteq V_{i-1} \subseteq U_{i-1} \subseteq NAE(V_i).$$

As V_i is regular, Corollary 3 and μ_R imply that

$$\mathbf{p}(\lambda_{V_i}/\lambda_{P(V_i)}) = \mathbf{p}(\lambda_{V_i}/\lambda_{V_{i-1}}) = \mathbf{p}(\lambda_{V_i}/\lambda_{U_{i-1}}) = \mathbf{p}(\lambda_{V_i}/\lambda_{NAE(V_i)}).$$

We get thus

$$\mathbf{p}(\lambda^*_{V_i}/\lambda^*_{V_{i-1}}) = \mathbf{p}(\lambda^*_{V_i}/\lambda^*_{U^*_{i-1}}).$$

As $U^*_{i-1} = \{V_j \mid j < i\}$, V^* is a first-order Markov chain.

(vi): Assume that ξ satisfies ν w.r.t. λ. Take $D^* \subset C^* \subseteq V^*$ and $V_i \in C^* - D^*$. To C^* and D^* correspond $D \subset C \subseteq V$, with $V_i \subseteq C - D$. We can apply ν' with $S = V_i$, $T = C$ and $R = D$, and so we get:

$$\mathbf{p}(\xi_{V_i}/\xi_D, \lambda_C) = \mathbf{p}(\xi_{V_i}/\lambda_{V_i}).$$

Therefore

$$\mathbf{p}(\xi^*_{V_i}/\xi^*_{D^*}, \lambda*_{C^*}) = \mathbf{p}(\xi^*_{V_i}/\lambda^*_{V_i}),$$

and so V^* satisfies ν. ∎

Thus one can apply to V^* the treatment applied to first order hidden Markov chains. This is done in [3] with V being a two-dimensional square grid and V^* the set of rows (or of columns) of V. As suggested by Devijver, a more promising approach is to chose for V^* the set of all secondary columns (i.e., those along the SW-NE direction), because it is then possible to make assumptions of relative independence between the λ_x for $x \in V_i$, something impossible when V_i is a row or a column.

III.2. The forward-backward method

Given a hierarchical structure V satisfying μ and ν, we will see to which extent Baum's forward-backward method can be applied directly to V, and not only to the Markov chain V^* described above.

Let us first give a very brief description of that method in the case of Markov chains. Here $V = \{1, \ldots, n\}$, and we define for $i = 1, \ldots, n$ and for a configuration λ the two numbers $\mathcal{F}(i; \lambda)$ and $\mathcal{B}(i; \lambda)$ (Baum's α and β respectively [2]) as follows:

$$\mathcal{F}(i; \lambda) = \mathbf{p}(\lambda_i, \xi_1, \ldots, \xi_i) \qquad (i = 1, \ldots, v),$$
$$\mathcal{B}(i; \lambda) = \mathbf{p}(\xi_{i+1}, \ldots, \xi_v/\lambda_i) \qquad (i = 1, \ldots, v-1),$$

with $\mathcal{B}(v;\lambda) = 1$. Here $\mathcal{F}$ and $\mathcal{B}$ are abbreviations of "forward" and "backward". Indeed, it can be shown that:

(i) For $i = 1, \ldots, v$, $\mathbf{p}(\xi_V, \lambda_i) = \mathcal{F}(i;\lambda) \cdot \mathcal{B}(i;\lambda)$;

(ii) $\mathcal{F}(i;\lambda)$ can be computed by a forward iteration from $i = 1$ to $i = v$;

(iii) $\mathcal{B}(i;\lambda)$ can be computed by a forward iteration from $i = v$ to $i = 1$;

(iv) Thanks to (i), (ii), and (iii), the set of all $\mathbf{p}(\xi_V, \lambda_i)$ ($i = 1, \ldots, v$, $\lambda \in \mathcal{C}[V]$) can be computed in linear time.

We can generalize these quantities to an arbitrary Markov field. Given $S \subseteq V$ and a classification ϕ for S, we define the following two numbers:

$$\mathcal{F}(S;\phi) = \mathbf{p}(\xi_{BE(S)}, \phi);$$
$$\mathcal{B}(S;\phi) = \mathbf{p}(\xi_{NBE(S)}/\phi).$$

When $S = \{x\}$, we write $\mathcal{F}(x;\phi)$ and $\mathcal{B}(x;\phi)$ for $\mathcal{F}(\{x\};\phi)$ and $\mathcal{B}(\{x\};\phi)$ respectively (as we did above for $V = \{1, \ldots, n\}$).

We did not restrict the definitions of $\mathcal{F}$ and $\mathcal{B}$ to points, but did also consider sets, in light of what was done with the set V^* at the end of the previous section. Let us note as an example that the function $\mathcal{G}_{a,b}(r,s)$ introduced in [5] for the rectangular grid is in fact $\mathcal{F}(P,\lambda)$, where P is the pair $\{(a-1,b),(a,b-1)\}$.

We can now give the basic forward and backward decomposition rules for $\mathcal{F}$ and $\mathcal{B}$; they can be applied only when S satisfies certain requirements. Let us beforehand make the following obsevation:

If S is regular, then $B(S) = BE(P(S))$. Indeed, if $x \in B(S)$, then $x < y$ for some $y \in S$, and $x \notin S$; there is thus a chain $x = z_1 \prec \ldots \prec z_r = y$ ($r > 1$); take the smallest $u \in \{1, \ldots, r\}$ such that $z_u \in S$; then $u > 1$, $z_{u-1} \in P(S)$ and $x \in BE(z_u)$. Hence $B(S) \subseteq BE(P(S))$. Now $P(S) \subseteq B(S)$, and as $B(S)$ is before-closed (see Lemma 9(iv)), $BE(P(S)) \subseteq B(S)$.

We can now give three results corresponding to the properties (i) to (iii) stated above for $\mathcal{F}$ and $\mathcal{B}$ in the case of Markov chains.

Proposition 16. *For any regular subset S of V and any configuration λ_S for S,*

$$\mathcal{F}(S;\lambda_S) = \mathbf{p}(\xi_S/\lambda_S) \cdot \sum_{\phi \in \mathcal{C}[P(S)]} \mathbf{p}(\lambda_S/\phi) \cdot \mathcal{F}(P(S);\phi).$$

Proof. As S is regular, $B(S) = BE(S) - S$ is before-closed (see Lemma 9(iv)), and $P(S) = PE(S) - S$ separates $B(S)$ from S. Thus Corollary 13 implies that for every $\phi \in \mathcal{C}[P(S)]$,

$$\mathbf{p}(\lambda_S/\xi_{B(S)}, \phi) = \mathbf{p}(\lambda_S/\phi). \qquad (10)$$

As observed above, $B(S) = BE(P(S))$. Hence $BE(S) = B(S) \cup S = BE(P(S)) \cup S$ and we obtain thus the following decomposition:

$$\begin{aligned}
\mathcal{F}(S;\lambda_S) &= \mathbf{p}(\xi_{BE(S)}, \lambda_S) = \mathbf{p}(\xi_S, \xi_{BE(P(S))}, \lambda_S); \\
&= \mathbf{p}(\xi_S/\lambda_S) \cdot \mathbf{p}(\xi_{BE(P(S))}, \lambda_S) \qquad (\text{by } \nu'); \\
&= \mathbf{p}(\xi_S/\lambda_S) \cdot \sum_{\phi \in \mathcal{C}[P(S)]} \mathbf{p}(\xi_{BE(P(S))}, \phi, \lambda_S); \\
&= \mathbf{p}(\xi_S/\lambda_S) \cdot \sum_{\phi \in \mathcal{C}[P(S)]} \mathbf{p}(\lambda_S/\xi_{BE(P(S))}, \phi) \cdot \mathbf{p}(\xi_{BE(P(S))}, \phi); \\
&= \mathbf{p}(\xi_S/\lambda_S) \cdot \sum_{\phi \in \mathcal{C}[P(S)]} \mathbf{p}(\lambda_S/\phi) \cdot \mathbf{p}(\xi_{BE(P(S))}, \phi) \qquad (\text{by } (10)); \\
&= \mathbf{p}(\xi_S/\lambda_S) \cdot \sum_{\phi \in \mathcal{C}[P(S)]} \mathbf{p}(\lambda_S/\phi) \cdot \mathcal{F}(P(S);\phi). \blacksquare
\end{aligned}$$

Let us make one further definition. A set R will be called *before-separating* if R is a separator of $BE(R)$, in other words if for every $y \in NBE(R)$, $P(y) \cap BE(R) \subseteq R$.

Proposition 17. *For any regular and before-separating subset S of V and any configuration $\lambda_{P(S)}$ for $P(S)$,*

$$\mathcal{B}(P(S);\lambda_{P(S)}) = \sum_{\theta \in \mathcal{C}[S]} \mathbf{p}(\xi_S/\theta) \cdot \mathbf{p}(\theta/\lambda_{P(S)}) \cdot \mathcal{B}(S;\theta).$$

Proof. As S is before-separating, Corollary 13 implies that for every $\theta \in \mathcal{C}[S]$,

$$\mathbf{p}(\xi_{NBE(S)}/\theta, \lambda_{P(S)}) = \mathbf{p}(\xi_{NBE(S)}/\theta). \tag{11}$$

As S is regular, $BE(P(S)) = B(S)$. We have thus $NBE(P(S)) = V - BE(P(S)) = V - B(S) = V - (BE(S) - S) = NBE(S) \cup \{S\}$. Hence we get the following decomposition:

$$\begin{aligned}
\mathcal{B}(P(S);\lambda_{P(S)}) &= \mathbf{p}(\xi_{NBE(P(S))}/\lambda_{P(S)}) = \mathbf{p}(\xi_{NBE(S)}, \xi_S/\lambda_{P(S)}); \\
&= \sum_{\theta \in \mathcal{C}[S]} \mathbf{p}(\xi_{NBE(S)}, \xi_S, \theta/\lambda_{P(S)}); \\
&= \sum_{\theta \in \mathcal{C}[S]} \mathbf{p}(\xi_S/\xi_{NBE(S)}, \theta, \lambda_{P(S)}) \cdot \mathbf{p}(\xi_{NBE(S)}/\theta, \lambda_{P(S)}) \cdot \mathbf{p}(\theta/\lambda_{P(S)}); \\
&= \sum_{\theta \in \mathcal{C}[S]} \mathbf{p}(\xi_S/\theta) \cdot \mathbf{p}(\xi_{NBE(S)}/\theta, \lambda_{P(S)}) \cdot \mathbf{p}(\theta/\lambda_{P(S)}) \qquad (\text{by } \nu'); \\
&= \sum_{\theta \in \mathcal{C}[S]} \mathbf{p}(\xi_S/\theta) \cdot \mathbf{p}(\xi_{NBE(S)}/\theta) \cdot \mathbf{p}(\theta/\lambda_{P(S)}) \qquad (\text{by } (11)); \\
&= \sum_{\theta \in \mathcal{C}[S]} \mathbf{p}(\xi_S/\theta) \cdot \mathbf{p}(\theta/\lambda_{P(S)}) \cdot \mathcal{B}(S;\theta). \blacksquare
\end{aligned}$$

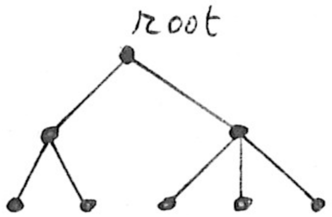

Figure 3. *A tree with a root*

Proposition 18. *For any before-separating subset S of V and any configuration λ_S for S,*

$$\mathrm{p}(\xi_V, \lambda_S) = \mathcal{F}(S; \lambda_S) \cdot \mathcal{B}(S; \lambda_S).$$

Proof. As S is before-separating, Corollary 13 implies that for any configuration λ for V,

$$\mathrm{p}(\xi_{NBE(S)}/\xi_{BE(S)}, \lambda_S) = \mathrm{p}(\xi_{NBE(S)}/\lambda_S). \tag{12}$$

We obtain thus:

$$\begin{aligned}\mathrm{p}(\xi_V, \lambda_S) &= \mathrm{p}(\xi_{NBE(S)}, \xi_{BE(S)}, \lambda_S) = \mathrm{p}(\xi_{NBE(S)}/\xi_{BE(S)}, \lambda_S) \cdot \mathrm{p}(\xi_{BE(S)}, \lambda_S);\\ &= \mathrm{p}(\xi_{NBE(S)}/\lambda_S) \cdot \mathrm{p}(\xi_{BE(S)}, \lambda_S) \qquad \text{(by (12))};\\ &= \mathcal{B}(S; \lambda_S) \cdot \mathcal{F}(S; \lambda_S). \ \blacksquare\end{aligned}$$

In the case where $V = \{1, \ldots, n\}$ mentioned at the beginning of this Section, every point i is before-separating and regular. Thus the assertions (ii), (iii) and (i) made there follow from Propositions 16, 17 and 18 respectively. These three propositions indicate also how the forward-backward method can be applied to a one-dimensional Markov chain, for example to the set $V^* = \{V_1, \ldots, V_k\}$ mentioned in Proposition 15.

In the case of a 2-D rectangular grid, if one chooses the rows, columns or secondary diagonals as elements of V^*, it is easy to see that the subsets V_i are regular and before-separating; thus Proposition 15 follows from Propositions 16, 17 and 18 in this case.

We will also be able to apply the forward-backward method to Markov fields on trees, thanks to new formulas giving a decomposition of $\mathcal{F}(S; \lambda_S)$ and $\mathcal{B}(S; \lambda_S)$ (where S is regular and before-separating) in terms of $\mathcal{F}(x; \lambda_x)$ and $\mathcal{B}(x; \lambda_x)$ for $x \in S$. This will be the object of the next section.

III.3. Bottom-up trees and forests

A non-oriented graph G is called a *forest* if G contains neither loops, multiple edges, nor cycles; it is called a *tree* if it is a connected forest.

In many branches of computer science, one generally associates to a tree a particular vertex called the root, and views that tree as if it was oriented from the leaves to the root in a bottom-up fashion (see Figure 3). In fact, every vertex of a tree can be chosen as a root.

Given an oriented graph G, we will call it a *bottom-up tree* if the corresponding non-oriented graph is a tree (with a root) and the arrows of G are oriented from the leaves to the root. We illustrate this in Figure 4. We will call it a *bottom-up forest* if it is the disjoint union of bottom-up trees; or equivalently if from every vertex of G points at most one arrow.

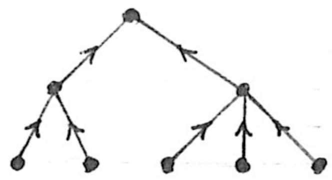

Figure 4. *A bottom-up tree*

For the set V endowed by $\prec$ with a hierarchical structure, we make the arrows point from every point to its successors. Thus V is a *bottom-up forest* if for every $z \in V$, z has at most one successor. An equivalent condition is that for every $x, y \in V$, either $x \leq y$, $x \geq y$, or $BE(x) \cap BE(y) = \emptyset$. Moreover V is a *bottom-up tree* if it is a bottom-up forest and if there exists some $z \in V$ such that $V = BE(z)$.

We will consider the case where $\prec$ induces on V a bottom-up tree or forest structure, and we will show that it is possible to implement the forward-backward method in that situation. This will be possible thanks to a few decomposition formulas which will be proven below. Let us make beforehand one more definition:

For $S, T \subseteq V$, we say that S and T are *before-disjoint* if $BE(S) \cap BE(T) = \emptyset$.

We have then the following two results:

Lemma 19. *Let $R_1, \ldots, R_n$ be pairwise before-disjoint subsets of V and let $R = R_1 \cup \cdots \cup R_n$. Then*

$$\mathcal{F}(R; \lambda_R) = \prod_{i=1}^{n} \mathcal{F}(R_i; \lambda_{R_i}).$$

Proof. We use induction. The result is trivial for $n = 1$. Suppose that $n > 1$ and that the result is true for $n - 1$. Then $R' = R_1 \cup \cdots \cup R_{n-1}$ and R_n are before-disjoint. By Corollary 14, $\mathrm{p}(\xi_{BE(R')}, \xi_{BE(R_n)}, \lambda'_R, \lambda_{R_n}) = \mathrm{p}(\xi_{BE(R')}, \lambda'_R) \cdot \mathrm{p}(\xi_{BE(R_n)}, \lambda_{R_n})$, in other words

$$\mathcal{F}(R; \lambda_R) = \mathcal{F}(R' \cup R_n; \lambda_{R' \cup R_n}) = \mathcal{F}(R'; \lambda_{R'}) \cdot \mathcal{F}(R_n; \lambda_{R_n}).$$

Now by induction hypothesis

$$\mathcal{F}(R'; \lambda'_R) = \prod_{i=1}^{n-1} \mathcal{F}(R_i; \lambda_{R_i}).$$

The result follows then by combining both equations. ∎

Lemma 20. *Let $R_1, \ldots, R_n$ be before-separating and pairwise before-disjoint subsets of V and let $R = R_1 \cup \cdots \cup R_n$. Then for $i = 1, \ldots, n$,*

$$\mathcal{B}(R_i; \lambda_{R_i}) = \sum_{\eta \in \mathcal{C}[R-R_i]} \mathcal{B}(R; \lambda_{R_i}, \eta) \cdot \prod_{j \neq i} \mathcal{F}(R_j; \eta_{R_j}).$$

Proof. We set $S = BE(R_i)$, $T = BE(R - R_i)$, and $N = V - (S \cup T)$. We have then $BE(R) = S \cup T$, $NBE(R) = N$ and $NBE(R_i) = N \cup T$. Let η be any configuration for $R - R_i$. As $R - R_i$ and R_i are before-disjoint, Corollary 13 (or 14) implies that

$$\mathbf{p}(\xi_T, \eta / \lambda_{R_i}) = \mathbf{p}(\xi_T, \eta). \tag{13}$$

As each R_j is before-separating, $R = R_1 \cup \cdots \cup R_n$ separates $BE(R) = S \cup T$ from N and Corollary 13 implies that

$$\mathbf{p}(\xi_N / \xi_T, \eta, \lambda_{R_i}) = \mathbf{p}(\xi_N / \eta, \lambda_{R_i}). \tag{14}$$

We get thus:

$$\begin{aligned}
\mathcal{B}(R_i; \lambda_{R_i}) = \mathbf{p}(\xi_{N \cup T} / \lambda_{R_i}) &= \sum_{\eta \in \mathcal{C}[R-R_i]} \mathbf{p}(\xi_N, \xi_T, \eta / \lambda_{R_i}); \\
&= \sum_{\eta \in \mathcal{C}[R-R_i]} \mathbf{p}(\xi_N / \xi_T, \eta, \lambda_{R_i}) \cdot \mathbf{p}(\xi_T, \eta / \lambda_{R_i}); \\
&= \sum_{\eta \in \mathcal{C}[R-R_i]} \mathbf{p}(\xi_N / \eta, \lambda_{R_i}) \cdot \mathbf{p}(\xi_T, \eta) \qquad \text{by (13) and (14)}; \\
&= \sum_{\eta \in \mathcal{C}[R-R_i]} \mathcal{B}(R; \lambda_{R_i}, \eta) \cdot \mathcal{F}(R - R_i; \eta).
\end{aligned}$$

The result follows then by applying Lemma 19 to $\mathcal{F}(R - R_i; \eta)$. ∎

We can now use these two results, together with Propositions 16, 17 and 18, in order to implement the forward-backward method in the case where $\prec$ induces on V a bottom-up tree. Here $R_1, \ldots, R_n$ will be the elements of $P(x)$ for $x \in V$:

Theorem 21. *Suppose that V is a bottom-up tree, i.e., there is some $z \in V$ such that $V = BE(z)$ and for every $y \in V - \{z\}$, $|S(y)| = 1$. Then for every $x \in V$,*

$$\mathbf{p}(\xi_V, \lambda_x) = \mathcal{F}(x; \lambda_x) \cdot \mathcal{B}(x; \lambda_x),$$

where

$$\mathcal{F}(x; \lambda_x) = \mathbf{p}(\xi_{BE(x)}, \lambda_x),$$
$$\mathcal{B}(x; \lambda_x) = \mathbf{p}(\xi_{NBE(x)} / \lambda_x),$$

and the quantities $\mathcal{F}(x;\lambda_x)$ and $\mathcal{B}(x;\lambda_x)$ can be computed iteratively as follows:

$$\begin{aligned}
\mathcal{F}(x;\lambda_x) &= \mathrm{p}(\lambda_x)\cdot\mathrm{p}(\xi_x/\lambda_x) && \textit{if } P(x)=\emptyset;\\
&= \mathrm{p}(\xi_x/\lambda_x)\cdot\sum_{\phi\in\mathcal{C}[P(x)]}\mathrm{p}(\lambda_x/\phi)\cdot\prod_{y\in P(x)}\mathcal{F}(y;\phi_y) && \textit{otherwise.}\\
\mathcal{B}(z;\lambda_z) &= 1;\\
\mathcal{B}(x;\lambda_x) &= \sum_{\eta\in\mathcal{C}[P(w)-\{x\}]}\ \sum_{\theta\in\mathcal{C}[w]}\mathrm{p}(\xi_w/\theta)\cdot\mathrm{p}(\theta/\lambda_x,\eta)\cdot\\
&\qquad \mathcal{B}(w;\theta)\cdot\prod_{y\in P(w)-\{x\}}\mathcal{F}(y;\eta_y) && \textit{if } S(x)=\{w\}.
\end{aligned}$$

Proof. The first equation follows from Proposition 18. We have only to justify the last equations giving the decomposition formulas for $\mathcal{F}(x,\lambda_x)$ and $\mathcal{B}(x,\lambda_x)$.

The decomposition of $\mathcal{F}(x,\lambda_x)$ is trivial for $P(x)=\emptyset$. If $P(x)\neq\emptyset$, we have by Proposition 16 (with $S=x$):

$$\mathcal{F}(x;\lambda_x)=\mathrm{p}(\xi_x/\lambda_x)\cdot\sum_{\phi\in\mathcal{C}[P(x)]}\mathrm{p}(\lambda_x/\phi)\cdot\mathcal{F}(P(x);\phi).$$

By Lemma 19 (with $R_1,\ldots,R_n$ being the elements of $P(x)$), we get:

$$\mathcal{F}(P(x);\phi)=\prod_{y\in P(x)}\mathcal{F}(y;\phi_y).$$

The decomposition follows then by combining these two equalities.

The decomposition of $\mathcal{B}(x,\lambda_x)$ is trivial for $S(x)=\emptyset$ (in other words for $x=z$). If $S(x)\neq\emptyset$, then $S(x)=w$ for a unique $w\in V$. By Lemma 20 (with $R_1,\ldots,R_n$ being the elements of $P(w)$), we get:

$$\mathcal{B}(x;\lambda_x)=\sum_{\eta\in\mathcal{C}[P(w)-\{x\}]}\mathcal{B}(P(w);\lambda_x,\eta)\cdot\prod_{y\in P(w)-\{x\}}\mathcal{F}(y;\eta_y).$$

We have by Proposition 17 (with $S=w$):

$$\mathcal{B}(P(w);\lambda_x,\eta)=\sum_{\theta\in\mathcal{C}[w]}\mathrm{p}(\xi_w/\theta)\cdot\mathrm{p}(\theta/\lambda_x,\eta)\cdot\mathcal{B}(w;\theta).$$

The decomposition formula follows then by combining these two equalities. ∎

Let us now consider the case where V is a bottom-up forest. Here V can be partitioned into the bottom-up trees $V_1=BE(z_1),\ldots,V_n=BE(z_n)$. Then Theorem 21 is still valid if we replace in it "$V-\{z\}$" by "$V-\{z_1,\ldots,z_n\}$" and "$\mathcal{B}(z,\lambda_z)$" "$\mathcal{B}(z_i,\lambda_{z_i})$" $(i=1,\ldots,n)$. However, there is a more economical solution, because we can restrict the computation of $\mathcal{F}$ and $\mathcal{B}$ to the sets V_i, as follows:

Given $S \subseteq V_i$ and a configuration λ, we write $\mathcal{F}_i(S;\lambda_S)$ and $\mathcal{B}_i(S;\lambda_S)$ for the values of $\mathcal{F}$ and $\mathcal{B}$ computed in V_i; in other words

$$\mathcal{F}_i(S;\lambda_S) \doteq \mathbf{p}(\xi_{BE(S)\cap V_i}, \lambda_S).$$
$$\mathcal{B}_i(S;\lambda_S) \doteq \mathbf{p}(\xi_{NBE(S)\cap V_i}/\lambda_S).$$

Then it is clear that $\mathcal{F}_i(S;\lambda_S) = \mathcal{F}(S;\lambda_S)$. On the other hand, we will have

$$\mathcal{B}(S;\lambda_S) = \mathcal{B}_i(S;\lambda_S) \cdot \prod_{j\neq i} \mathbf{p}(\xi_{V_j}). \tag{15}$$

This follows from Corollary 14, because $NBE(S) = (NBE(S)\cap V_i) \cup \bigcup_{j\neq i} V_j$ and the sets $V_1, \ldots, V_n$ are before-closed and pairwise disjoint. We obtain then the following result:

Theorem 22. *Suppose that V is a bottom-up forest which can be partitioned into the bottom-up trees $V_1 = BE(z_1), \ldots, V_n = BE(z_n)$. For $x \in V_i$, write $\mathcal{F}_i(x;\lambda_x)$ and $\mathcal{B}_i(x;\lambda_x)$ for the $\mathcal{F}(x;\lambda_x)$ and $\mathcal{B}(x;\lambda_x)$ of Theorem 21 computed in the tree V_i. Then*

$$\mathbf{p}(\xi_V, \lambda_x) = \mathcal{F}_i(x;\lambda_x) \cdot \mathcal{B}_i(x;\lambda_x) \cdot \prod_{j\neq i} \mathbf{p}(\xi_{V_j}),$$

with

$$\mathbf{p}(\xi_{V_j}) = \sum_{\theta_{z_j} \in \mathcal{C}[z_j]} \mathcal{F}_j(z_j;\theta_{z_j}).$$

Let us now give an interpretation of Theorem 21. The numbers $\mathcal{F}(x;\lambda_x)$ $(x \in V)$ can be computed iteratively from the bottom to the top of the tree; afterwards the numbers $\mathcal{B}(x;\lambda_x)$ $(x \in V)$ can be computed iteratively from the top to the bottom of the tree with the help of the $\mathcal{F}(y,\lambda_y)$. Thus the forward-backward method can be applied to the tree. Moreover, if the feature space Γ has size c and if the number of children nodes of a node in the tree is at most u, then the computational complexity of the decomposition of $\mathcal{F}(x;\lambda_x)$ and $\mathcal{B}(x;\lambda_x)$ is in at most c^u, and so the computation of all $\mathbf{p}(\xi_V, \lambda_x)$ $(x \in V,\ \lambda_x \in \mathcal{C}[x])$ can be achieved with a complexity in $O(v \cdot c^u)$.

Let us now make a few practical remarks. In a bottom-up tree, the children nodes of a node x are the elements of $P(x)$, while the parent node of x is the successor of x. One might argue that the transition probabilities $\mathbf{p}(\lambda_x/\lambda_{P(x)})$ are too restricted, that the feature state on a node should be determined by the feature states not only of its children nodes, but also of the nodes having the same parent. In other words, the transition probabilities could take the form $\mathbf{p}(\lambda_x/\lambda_{P(x)}, \lambda_{P(S(x))-\{x\}})$ One can remedy to this limitation by modifying the signification of the state λ_x on a node x: we can assume that this state describes features of the children nodes of x; in other words, if the children nodes of x are $y_1, \ldots, y_t$, then $\lambda_x = (\hat{\lambda}_{y_1}, \ldots, \hat{\lambda}_{y_t})$, where each $\hat{\lambda}_{y_i}$ is a feature state for y_i. In this way, the features on y_i depend on those on y_j $(j \neq i)$ thanks to the dependence of $\lambda_x = (\hat{\lambda}_{y_1}, \ldots, \hat{\lambda}_{y_t})$ upon

$(\lambda_{y_1}, \ldots, \lambda_{y_t})$. Note that for leaf nodes (those having no children), the state λ_x is then undefined.

Second, an interesting question to investigate is the possibility of applications of Markov fields to bottom-up trees. Consider for example the quadtree of an image. Here the nodes represent image portions having various sizes, and the states λ_x corresponding to them belong to the same set Γ, irrespectively of their size. One should thus choose features whose interpretation is independent of the size of these image portions.

References

[1] K. Abend, T.J. Harley, L.N. Kanal: Classification of Binary Random Patterns. *IEEE Trans. Information Theory*, Vol. IT-11, no. 4, pp. 538–544, Oct. 1965.

[2] L.E. Baum: An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes. *Inequalities*, Vol. 3, pp. 1–8, 1972.

[3] H. Derin, H. Elliott, R. Cristi, D. Geman: Bayes Smoothing Algorithm for Segmentation of Binary Images Modeled by Markov Random Fields. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-6, no. 6, pp. 707–720, Nov. 1984.

[4] P.A. Devijver: A Comparative Study of Decision making Algorithms in Hidden Markov Chains. *PRLB Research Report R481*, Oct. 1984.

[5] P.A. Devijver: Optimal Probabilistic Labelings in Markov Fields. *PRLB Research Report R483*, Dec. 1984.

[6] P.A. Devijver: An Improved F-G-H Algorithm. *PRLB Manuscript M98*, Jan. 1985.

[7] L.N. Kanal: Markov Mesh Models. *Computer Graphics & Image Processing*, Vol. 12, no. 4, pp.371–375, Apr. 1980.

[8] M. Kanefsky, M.G. Strintzis: A Decision Theory Approach to Picture Smoothing. *IEEE Trans. Computers*, Vol. C-27, no. 1, pp. 32–36, Jan. 1978.

[9] R. Kindermann, J.L. Snell: *Markov Random Fields and their Applications.* Contemporary Mathematics, Vol. 1, AMS (Providence, RI), 1980.

[10] S.E. Levinson, L.R. Rabiner, M.M. Sondhi: An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition. *Bell System Technical J.*, Vol. 62, no. 4, pp. 1035–1074, Apr. 1983.