

Champs de Markov cachés et estimation conditionnelle itérative.

Hidden Markov Fields and Iterative Conditional Estimation

par Wojciech PIECZYNSKI

Département Signal et Image
Institut National des Télécommunications
9, rue Charles Fourier - 91011 Evry - France

Résumé

Notre étude a pour objet le problème posé par l'estimation des paramètres dans les champs de Markov cachés. L'objectif principal en est la comparaison des algorithmes obtenus à partir d'une méthode générale récente d'estimation dans le cas des données cachées, dite Estimation Conditionnelle Itérative (ECI), avec certains algorithmes existants. En premier lieu nous rappelons l'importance du problème de l'estimation dans les champs markoviens cachés en segmentation non supervisée d'images. Nous comparons ensuite les méthodes de type ECI avec certains algorithmes existants. Nous discutons les différences entre les principes régissant les différentes démarches, la généralité des modèles pouvant être traités et les difficultés de mise en œuvre. Nous présentons une étude comparative détaillée avec les algorithmes Gradient Stochastique de L. Younes et EM Gibbsien de B. Chalmond. Le principe de ECI, qui diffère de ceux de toutes les méthodes existantes, permet la conception des algorithmes applicables dans un cadre relativement général. Dans la mesure où les formules de l'algorithme EM peuvent être obtenues par une procédure ECI particulière, il apparaît comme en étant un cas particulier.

Mots clés : champs de Markov cachés, estimation des paramètres, EM, gradient stochastique, ECI, segmentation non supervisée.

Abstract

This work deals with the parameter estimation problem in hidden Markov fields. The principal goal is the comparison of methods deriving from a recent general procedure of estimation in the case of hidden data, Iterative Conditional Estimation (ICE), with some existing algorithms. The paper starts with the recall of the importance of the hidden Markov fields estimation problem in unsupervised image segmentation. Then we compare ICE methods with some existing algorithms at different levels : principle, generality of models, difficulty of its implementation. The Stochastic Gradient of L. Younes and the Gibbsian EM of B. Chalmond are compared with ICE methods in some detail. The principle of ICE, different from principles of all existing methods, allows the conception of algorithms applicable in a quite general framework. Furthermore, the EM formulae can be obtained by ICE and, in this sense, it can be seen as particular ICE case.

Key words : hidden Markov fields, parameter estimation, EM, stochastic gradient, ICE, unsupervised segmentation.

1. Introduction

Le problème de la segmentation figure parmi les plus importants se posant en imagerie. Dans certaines situations la segmentation constitue un stade préliminaire ouvrant la voie à d'autres traitements, dans d'autres elle fournit directement les résultats recherchés. Parmi les nombreux algorithmes existants, la famille des méthodes statistiques s'avère, dans certains cas, d'une remarquable efficacité. Les segmentations statistiques s'appuient sur la modélisation par champs aléatoires. S étant l'ensemble de pixels, on considère deux collections de variables aléatoires

$X = (X_s)_{s \in S}, Y = (Y_s)_{s \in S}$ dites « champs aléatoires ». Chaque X_s prend ses valeurs dans un ensemble fini Ω de classes et nous supposons dans toute la suite, afin de simplifier, que chaque Y_s prend ses valeurs dans \mathcal{R} . Notons que Ω désigne usuellement l'ensemble des épreuves et nous l'identifions ici à l'ensemble des classes, ce qui est possible en vertu du théorème de transfert. Dans ce cadre, le problème de la segmentation est celui de l'estimation de la réalisation invisible de $X = x$ à partir de la réalisation observée $Y = y$, y étant l'image numérique à segmenter. La procédure d'estimation généralement adoptée est la classification bayésienne. N étant le cardinal de S , on se donne une fonction L de $\Omega^N \times \Omega^N$ dans \mathcal{R} , dite « fonction de perte ». Pour $x_i, x_j \in \Omega^N$, $L(x_i, x_j)$ modélise le degré de gravité que

l'on accorde à l'erreur « choix erroné de x_i au lieu de la vraie valeur x_j ». La classification bayésienne s_B est alors celle parmi les applications s de \mathcal{R}^N dans Ω^N pour laquelle l'espérance de $L(X, s(Y))$ est minimale.

La stratégie bayésienne s_B dépend ainsi de la loi de (X, Y) , qui sera notée $P_{(X,Y)}$ dans toute la suite, et de L . Le calcul explicite de s_B , ainsi que celui de son approximation, ne sont pas possibles dans le cas général. Les difficultés viennent principalement de l'absence d'un ordre naturel dans S et de son cardinal élevé. On est ainsi amené, pour pouvoir envisager la mise en œuvre des stratégies bayésiennes, à renoncer au traitement du problème dans sa généralité et à introduire des modélisations particulières de $P_{(X,Y)}$. La simplification la plus brutale est celle qui suppose l'indépendance des variables $(Z_s)_{s \in S}$, où $Z_s \triangleq (X_s, Y_s)$. La stratégie bayésienne correspondante, dite « aveugle », est alors explicitement calculable et autorise l'utilisation d'un grand éventail de fonctions de perte L . L'inconvénient de cette démarche est la perte de l'information spatiale : la recherche de chaque x_s est effectuée à partir du seul y_s . Une autre simplification aboutit aux méthodes contextuelles : la recherche de chaque x_s est effectuée à partir de y_{V_s} , restriction de y à V_s , un voisinage de s . Sous la réserve V_s de petite taille les stratégies bayésiennes sont encore explicitement calculables et les stratégies contextuelles ont été largement utilisées avant l'introduction des modélisations markoviennes ([25], [34], [40], [41]). Ces stratégies gardent encore, dans certains cas, un intérêt non négligeable ([7], [29], [31], [35], [45]).

Notre travail concerne un troisième type de modélisation de $P_{(X,Y)}$, fréquemment utilisé par les auteurs depuis une dizaine d'années et donnant, dans certains cas, des résultats spectaculaires. L'importance de la modélisation que nous rappelons ci-dessous est due aux résultats de S. Geman et al. ([20]) qui ont permis de résoudre d'importants problèmes algorithmiques. X est un champ markovien et les variables aléatoires Y_s sont indépendantes conditionnellement à X . Dans ces conditions, la distribution de X conditionnelle à Y , dite « a posteriori », est également celle d'un champ markovien, ce qui rend possible la simulation de ses réalisations par des procédures itératives de type « échantillonneur de Gibbs ». Cette possibilité est à la base de deux algorithmes itératifs permettant l'approximation de deux stratégies bayésiennes correspondant à deux fonctions de perte L différentes : algorithme de recuit simulé de S. Geman et al. ([20]) et l'algorithme de Marroquin et al. ([30]). Le premier permet de construire une suite d'images convergeant vers une configuration $X = x_0$ dont la probabilité conditionnelle à $Y = y$, dite « a posteriori », est maximale. La méthode de segmentation consistant à associer à $Y = y$ un tel x_0 est dite méthode du maximum a posteriori (MAP). Le deuxième permet la recherche de x^0 tel que pour tout $s \in S$, la probabilité de x_s^0 conditionnelle à $Y = y$, dite probabilité a posteriori marginale, est maximale. Une telle méthode est dite « maximum des marginales a posteriori » (MPM). Signalons également l'algorithme « Iterated Conditional Modes » (ICM) de J. Besag ([5]). Cette méthode déterministe permet de définir une suite de réali-

sations de X convergeant vers des maxima locaux, dépendant de l'initialisation, de sa loi a posteriori. Bien que la convergence vers une solution du MAP ne soit pas garantie, ICM est fréquemment utilisée du fait de sa rapidité. L'avantage de cette modélisation par rapport aux précédentes est de permettre la recherche de chaque x_s à partir de tout le champ y . Cette possibilité de tenir compte de toute l'information spatiale simultanément est à la base de l'efficacité parfois surprenante des méthodes MAP, MPM, ICM ([3], [6], [7], [8], [12], [13], [16], [19], [20], [27], [28] [29], [30], [39], [42], [43], [45], [47]).

La mise en œuvre des trois algorithmes ci-dessus requiert la connaissance de certains paramètres de la loi $P_{(X,Y)}$. Lorsqu'on souhaite intégrer l'algorithme de segmentation choisi dans une chaîne complexe de traitement d'images, utilisable de façon automatique dans des problèmes faisant intervenir des images de types différents, le problème d'estimation de ces paramètres devient crucial. Leur estimation est entièrement automatique si elle est faite uniquement à partir de $Y = y$, image numérique à segmenter. Plusieurs propositions de telles méthodes ont été avancées par les auteurs. Parmi celles-ci la récente procédure générale d'estimation dans le cas de données cachées, dite « estimation conditionnelle itérative » (ECI), peut générer, dans le contexte de modélisations markoviennes, plusieurs méthodes particulières ([37]). Trois algorithmes de segmentation non supervisée, où la phase « estimation » est traitée par une variante de ECI, ont été programmés et les résultats numériques sont encourageants ([1], [7], [8]).

Dans ce travail, nous nous intéressons essentiellement au problème de l'estimation des paramètres. Notre objectif est de préciser davantage les différentes possibilités d'utilisation de ECI dans le contexte de champs de Markov cachés et de situer sa démarche par rapport à certaines méthodes d'estimation des paramètres à partir de $Y = y$ existantes. Nous nous intéressons tout particulièrement à l'algorithme du Gradient Stochastique de L. Younes ([47]) qui sera noté (GSY) dans la suite, et à l'algorithme EM Gibbsien de B. Chalmond ([13]) que nous noterons EMG. Concernant ces deux algorithmes nous discutons les points suivants :

- (i) Fondement des algorithmes et degré de généralité de leur applicabilité. Nous nous intéressons plus particulièrement au degré de généralité de modèles dans lesquels les principes régissant les divers algorithmes restent applicables.
- (ii) Difficultés de mise en œuvre. Ce point concerne les simplifications qu'on est amené à faire afin de contourner les difficultés rencontrées dans l'application exacte des principes définissant les méthodes en présence.
- (iii) Généralité des algorithmes. ECI apparaît comme plus générale par rapport à l'algorithme EMG dans la mesure où les formules de re-estimation auxquelles aboutit ce dernier peuvent être obtenues par une démarche ECI particulière.

Nous mentionnons également deux autres algorithmes et situons, de manière moins détaillée, la démarche de ECI par rapport à celle de ces méthodes. La première est une méthode générale utilisant

les segmentations effectuées sur la base de la valeur courante des paramètres ([5], [16], [22]). La deuxième est l'algorithme de S. Lakshmanan et H. Derin ([28]).

L'organisation de l'article est la suivante :

Dans la section suivante nous rappelons rapidement le modèle markovien généralement adopté par les auteurs.

La troisième section est consacrée à la procédure ECI; nous rappelons son principe général et montrons son originalité par rapport aux algorithmes Expectation-Maximisation (EM, [15]) et Stochastique EM (SEM, [11]) dans le cas simple de mélange gaussien. Nous décrivons ensuite les possibilités de son application dans le contexte des champs markoviens cachés.

La quatrième section est consacrée à l'étude comparative de l'ECI et de l'algorithme du Gradient Stochastique de L. Younes.

Dans la cinquième section, nous comparons ECI à l'EM Gibbsien de B. Chalmond.

Deux autres algorithmes sont rappelés dans la sixième section et nous présentons une rapide discussion les situant par rapport à l'ECI.

La septième et dernière section contient la conclusion.

2. Modèle markovien

Nous rappelons brièvement dans ce paragraphe, le modèle markovien caché.

Chaque variable X_s du champ X prend ses valeurs dans $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ et toutes les réalisations de X se produisent avec une probabilité non nulle. On peut alors écrire :

$$P[X = x] = ce^{-U(x)} \quad (1)$$

où U est dite « énergie » et c la constante de normalisation. X sera supposé markovien : la loi de chaque X_s conditionnelle à tous les autres X_t est égale à la loi de X_s conditionnelle à $(X_t)_{t \in V_s}$, où V_s est un voisinage de s . Nous supposons que la forme géométrique de V_s est indépendante de s . Le théorème d'Hammersley-Clifford ([5], [24]) permet alors d'affirmer que U est une somme de fonctions définies sur les cliques, une clique étant soit un singleton, soit un ensemble de pixels mutuellement voisins au sens du système des voisinages V_s . Nous supposons également que la loi de X_s conditionnelle à $(X_t)_{t \in V_s}$, est stationnaire. Nous n'explicitons pas davantage cette modélisation largement décrite dans la littérature ([5], [7], [8], [12], [13], [16], [18], [19], [20], [27], [28], [30], [37], [42], [43], [44], [45], [46], [47]).

L'intérêt de son utilisation en imagerie est dû à deux facteurs suivants :

(a) Lorsque la taille de V_s est raisonnable, celle des cliques est petite. P_X est ainsi définie par un nombre limité de paramètres dont la variation assure une grande variété d'images.

(b) La loi de X_s conditionnelle à $(X_t)_{t \in V_s}$ est explicitement calculable. La possibilité de simulation des réalisations de chaque X_s selon cette loi implique la possibilité de simulation, par des procédures itératives de type d'échantillonneur de Gibbs, des réalisations de X selon la loi donnée par (1).

Dans la suite, nous noterons α l'ensemble des paramètres définissant U , donc P_X .

Les variables aléatoires $(Y_s)_{s \in S}$ seront supposées indépendantes conditionnellement à X , de plus la loi de chaque Y_s conditionnelles à X sera supposée égale à sa loi conditionnelle à X_s . Les m lois de Y_s conditionnelles à $X_s = \omega_1, \dots, \omega_m$ respectivement déterminent ainsi la loi de Y conditionnelle à X . Nous noterons f_1, \dots, f_m les densités de ces lois par rapport à la mesure de Lebesgue et β_1, \dots, β_m les paramètres définissant ces densités. Le paramètre $\beta \triangleq (\beta_1, \dots, \beta_m)$ détermine ainsi toutes les lois de Y conditionnelles à X .

Notons $h(x_s)$ l'indice de la classe x_s . La loi du couple (X, Y) est donnée par :

$$f(x, y) = ce^{-U(x)} \prod_{s \in S} f_{h(x_s)}(y_s) \quad (2)$$

En notant $V_x(y) \triangleq - \sum_{s \in S} \text{Log}(f_{h(x_s)}(y_s))$, nous avons :

$$f(x, y) = ce^{-[U(x) + V_x(y)]} \quad (3)$$

Cela implique que la loi « a posteriori » de X , i.e. conditionnelle à Y , est également une loi d'un champ markovien et les propriétés (a), (b) ci-dessus restent valables. La propriété (b) est importante pour la suite et sera systématiquement utilisée.

Notons $\theta = (\alpha, \beta)$ l'ensemble des paramètres définissant la loi de $P_{(X, Y)}$. Notre étude concerne le problème de l'estimation de θ à partir de Y .

3. Procédure ECI

Ce paragraphe est consacré à la récente procédure générale d'estimation dans le cas des données cachées dite « Estimation Conditionnelle Itérative », ECI ([37]). Nous exposons son principe dans un contexte général et précisons les possibilités de ses applications dans le cas des champs de Markov cachés.

3.1. CONTEXTE GÉNÉRAL

Considérons deux collections de variables aléatoires $X = (X_i)_{i \in I}$, $Y = (Y_i)_{i \in I}$ dont la loi conjointe $P_{(X, Y)}$ dépend d'un paramètre θ . Nous cherchons une méthode d'estimation de θ à partir de Y . Il existe généralement des estimateurs $\hat{\theta} = \hat{\theta}(X, Y)$ de θ à partir de (X, Y) et nous nous placerons dans cette hypothèse. Si tel n'est pas le cas, on peut en effet raisonnablement considérer le problème de l'estimation de θ à partir du seul Y

comme inextricable. L'idée de l'ECI est de considérer un tel estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ et de rechercher son approximation dans l'ensemble de variables aléatoires fonctions de Y , le seul observable. La meilleure approximation, au sens de l'erreur quadratique moyenne, est l'espérance conditionnelle. L'espérance conditionnelle de $\hat{\theta} = \hat{\theta}(X, Y)$ par rapport à Y , qui sera notée $E[\hat{\theta}/Y]$, dépend du paramètre θ . En effet, c'est la projection orthogonale par rapport au produit scalaire de L^2 , qui dépend de θ , sur l'ensemble de variables aléatoires fonctions de Y . Lorsqu'on cherche à calculer, ou à approcher $E[\hat{\theta}/Y]$, on est ainsi amené à considérer la valeur « courante » θ_k du paramètre et la valeur suivante est donnée par $\theta_{k+1} = E_k[\hat{\theta}/Y]$, où E_k est l'espérance conditionnelle correspondant à θ_k . Nous retrouvons ainsi le caractère itératif, propre à toutes les méthodes d'estimation dans le cas des données cachées, de ECI. Finalement la démarche est la suivante :

- on considère une valeur initiale θ_0
- on pose

$$\theta_{k+1} = E_k[\hat{\theta}/Y = y] \quad (4)$$

En raison du degré de difficulté de l'étude théorique de la suite aléatoire (θ_k) ainsi obtenue, nous ne pouvons prétendre à l'obtention d'un résultat précis concernant son comportement asymptotique. Notons cependant que tel est généralement le cas pour les algorithmes existants. En particulier, à l'exception de l'algorithme de L. Younes pour lequel il existe un résultat « local » (voir section 4.6), aucun des quatre algorithmes faisant l'objet de notre étude comparative dans le cas des champs de Markov cachés n'admet de justification mathématique complète.

La justification de ECI reste ainsi intuitive et s'appuie sur la propriété fondamentale de l'espérance conditionnelle énoncée plus haut. Lorsque l'estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ possède des bonnes propriétés, en particulier lorsqu'il converge dans L^2 , $E_k[\hat{\theta}/Y]$ assure sa meilleure approximation, lorsque $\theta_k \rightarrow \theta$, en fonction des données observables. On peut ainsi espérer que la « perte » des données X implique, dans le cas de ECI, la « perte » minimale des bonnes propriétés de $\hat{\theta}$. Ces considérations sont confirmées par les premières études des segmentations non supervisées d'images, où la phase « estimation » est traitée par une variante de ECI : les algorithmes ainsi obtenus s'avèrent compétitifs vis-à-vis des méthodes existantes ([1], [7], [8]).

Le principe de l'ECI implique la possibilité de son application dans une grande variété de situations. Contrairement au principe de l'algorithme EM, la notion de vraisemblance n'intervient pas, on peut ainsi considérer les lois de probabilité contenant simultanément une partie continue et des masses de Dirac. De telles situations se présentent dans la théorie des files d'attente ou en classification bayésienne floue ([9], [10]).

Lorsque $E_k = [\hat{\theta}/Y = y]$ n'est pas explicitement calculable mais la simulation des réalisations de X selon la loi conditionnelle à Y possible, on peut utiliser une approximation stochastique. En effet, l'espérance conditionnelle est aussi l'espérance selon la loi conditionnelle et peut être approchée, en vertu de la loi des grands nombres, par la moyenne empirique. On peut ainsi simuler N

réalisations x_1, \dots, x_N de X selon la loi conditionnelle à $Y = y$ et poser :

$$\theta_{k+1} \triangleq \frac{1}{N} [\hat{\theta}(x_1, y) + \dots + \hat{\theta}(x_N, y)] \quad (5)$$

Nous verrons dans la section suivante que le calcul de $E_k = [\hat{\theta}/Y = y]$ n'est pas possible dans le cadre du modèle exposé dans la section précédente, par contre la propriété (b) permettra la mise en œuvre de l'approximation stochastique ci-dessus.

En conclusion, l'utilisation de ECI est possible dès que :

- I. On dispose d'un estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ de θ à partir des données complètes.
- II. On dispose d'une méthode de simulation, valable pour tout θ , des réalisations de X selon la loi conditionnelle à $Y = y$.

3.2. MÉLANGE GAUSSIEN

L'objectif de cette sous-section est de mettre en évidence la différence entre ECI et les algorithmes EM et SEM dans le cas simple de mélange gaussien.

Considérons une suite de n variables aléatoires $Z_i = (X_i, Y_i)$ indépendantes et équidistribuées, chaque X_i prenant ses valeurs dans $\Omega = \{\omega_1, \omega_2\}$ et Y_i dans \mathbf{R} . Les lois des Y_i conditionnelles à $X_i = \omega_1, \omega_2$ respectivement sont gaussiennes et seront notées $N(\mu_1, 1)$, $N(\mu_2, 1)$. En notant $\alpha = P[X_i = \omega_1]$ les paramètres $\theta = (\alpha, \mu_1, \mu_2)$ définissent entièrement la loi de $(X, Y) \triangleq ((X_1, \dots, X_n), (Y_1, \dots, Y_n))$. Nous considérons le problème d'estimation de $\theta = (\alpha, \mu_1, \mu_2)$ à partir de $Y = (Y_1, \dots, Y_n) = (y_1, \dots, y_n)$.

3.2.1. Méthode ECI

Supposons, conformément au principe général de l'ECI, que $X = (X_1, \dots, X_n)$ est observable. Notons

$$U \triangleq 1_{[X_1=\omega_1]} + \dots + 1_{[X_n=\omega_1]}$$

et

$$V \triangleq Y_1 1_{[X_1=\omega_1]} + \dots + Y_n 1_{[X_n=\omega_1]}$$

avec, pour A ensemble, 1_A fonction indicatrice de A . α et μ_1 peuvent être estimés par :

$$\hat{\alpha} = \frac{U}{n} \quad (6)$$

$$\hat{\mu}_1 = \frac{V}{U} \quad (7)$$

qui sont fréquence et moyenne empiriques.

Notons $\pi_{1,k}, \dots, \pi_{n,k}$, les probabilités

$$P[X_1 = \omega_1 / Y_1 = y_1], \dots, P[X_n = \omega_1 / Y_n = y_n]$$

calculées sur la base du paramètre courant θ_k . Selon le principe de l'ECI les valeurs suivantes des paramètres sont données par

$E_k[\hat{\alpha}/Y = y]$ et $E_k[\hat{\mu}_1/Y = y]$. Le calcul explicite de la première expression est possible et on trouve :

$$\alpha_{k+1} = E_k[\hat{\alpha}/Y = y] = \frac{1}{n}(\pi_{1,k} + \dots + \pi_{n,k}) \quad (8)$$

Le calcul de $E_k[\hat{\mu}_1/Y = y]$ n'est pas possible explicitement et on doit procéder à son approximation en utilisant (5). On simule N réalisations x^1, \dots, x^N du vecteur $X = (X_1, \dots, X_n)$ selon sa loi conditionnelle à $Y = (Y_1, \dots, Y_n) = (y_1, \dots, y_n)$. Ainsi une composante donnée x_i^j du vecteur x^j est la réalisation du tirage dans $\Omega = \{\omega_1, \omega_2\}$ selon la probabilité $\pi_{i,k}, 1 - \pi_{i,k}$. En appliquant (7) aux réalisations x^1, \dots, x^N , on obtient $\mu_{1,k}^1, \dots, \mu_{1,k}^N$. La valeur suivante du paramètre μ_1 est alors donnée par :

$$\mu_{1,k+1} = \frac{1}{N}(\mu_{1,k}^1 + \dots + \mu_{1,k}^N) \quad (9)$$

On procède de manière analogue pour la re-estimation de μ_2 , ce qui donne la valeur suivante θ_{k+1} de l'ensemble des paramètres.

3.2.2. Méthode EM

Le principe même de l'algorithme EM est différent du principe de ECI. Nous discutons les différences entre les deux principes dans la section 5.1. et nous nous contentons ici de donner directement les formules de re-estimation. L'application du principe de EM à notre exemple aboutit aux formules suivantes :

$$\alpha_{k+1} = \frac{1}{n}(\pi_{1,k} + \dots + \pi_{n,k}) \quad (10)$$

$$\mu_{1,k+1} = \frac{y_1 \pi_{1,k} + \dots + y_2 \pi_{n,k}}{\pi_{1,k} + \dots + \pi_{n,k}} \quad (11)$$

On a ainsi la même formule de re-estimation pour le paramètre α et une formule différente, ne faisant pas appel aux simulations, pour les paramètres μ_1, μ_2 . Notons que les paramètres μ_1, μ_2 interviennent dans le calcul des probabilités $\pi_{i,k}$, les deux suites (α_k) obtenues par ECI et EM sont donc différentes.

3.2.3. Méthode SEM

L'algorithme SEM est présenté par les auteurs comme une variante stochastique de EM ([11]). Les re-estimations de tous les paramètres sont faites à partir de simulations, comme pour la re-estimation de μ_1, μ_2 dans le cas de ECI, en faisant un seul tirage ($N = 1$). Ainsi, les moyennes μ_1, μ_2 sont re-estimées par la formule (9) avec $N = 1$. Par contre α est re-estimé à partir du résultat $x^1 = (x_1^1, \dots, x_n^1)$ de tirage par la fréquence empirique :

$$\alpha_{k+1} = \frac{1}{n}(1_{[x_1^1=\omega_1]} + \dots + 1_{[x_n^1=\omega_1]}) \quad (12)$$

La méthode ECI aboutit ainsi, dans le cas de mélange gaussien, à une procédure d'estimation de paramètres différente de celles obtenues par les méthodes EM et SEM.

3.3. CHAMPS DE MARKOV CACHÉS

Nous considérons le modèle défini dans la section 2, avec $\theta = (\alpha, \beta)$, où α est l'ensemble des paramètres définissant la loi de X et β l'ensemble des paramètres définissant les lois de Y conditionnelles à X . Suivant le principe général de l'ECI, nous recherchons dans un premier temps des estimateurs $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ de $\theta = (\alpha, \beta)$ à partir de (X, Y) . Concernant l'estimation de α , quatre estimateurs au moins $\hat{\alpha} = \hat{\alpha}(X)$ ont été proposés par les auteurs : [5], [14], [16], [46]. Tous ces estimateurs présentent de bonnes propriétés asymptotiques. Une démonstration récapitulative en est présentée dans [24]. En particulier l'algorithme de L. Younes ([46]), que nous utiliserons dans la suite, permet d'approcher le maximum de vraisemblance. Insistons cependant sur le fait que toutes ces méthodes permettent d'estimer α à partir de X et les justifications théoriques de leurs convergences ne sont généralisables au cas bruité (estimation de α à partir de Y) pour aucune d'entre elles.

Nous proposons pour $\hat{\beta} = \hat{\beta}(X, Y)$ l'estimation classique à partir des moments empiriques. Posons $\Omega = \{\omega_1, \dots, \omega_m\}$. $X = x$ étant observé, considérons la partition S_1, \dots, S_m de S , avec $s \in S_q$ si et seulement si $x_s = \omega_q$. Sur chaque S_q la réalisation de Y_s est une réalisation de Y_s selon la loi P_q donnée par le paramètre β_q . Dans les cas classiques où P_q est donnée par la densité de forme connue, il existe généralement un lien direct entre β_q et les moments, ces derniers pouvant être estimés, à partir de la restriction de Y à S_q , par les moments empiriques. Notons que le cas non paramétrique peut également être envisagé. Considérons, à titre d'exemple, le cas de trois classes où P_1 est une loi exponentielle, P_2 une loi gaussienne et P_3 une loi admettant une densité de forme inconnue. β_1 est donné par la seule moyenne, β_2 par la moyenne et la variance et β_3 est composé des paramètres correspondant à la méthode d'estimation non paramétrique choisie. β_1 et β_2 peuvent donc être estimés, à partir des restrictions de Y à S_1, S_2 respectivement, par les moments empiriques et β_3 , à partir de la restriction de Y à S_3 , par la méthode non paramétrique donnée.

Considérons $\hat{\alpha} = \hat{\alpha}(X)$, l'un des quatre estimateurs cités et $\hat{\beta} = \hat{\beta}(X, Y)$ l'estimateur décrit ci-dessus. Nous disposons donc d'un estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ et pouvons envisager l'application de l'ECI. Le calcul explicite de l'espérance conditionnelle donnée par (4) est impossible dans ce contexte, par contre on peut utiliser son approximation (5) à partir des simulations des réalisations de X selon la loi a posteriori. En effet, nous avons vu dans la section 2 que la structure du modèle a été conçue de façon à ce que de telles simulations soient possibles.

En conclusion, la démarche est la suivante :

1. Initialisation : on se donne θ_0
2. Calcul de θ_{k+1} à partir de θ_k et $Y = y$:
 - simulation de N réalisations x_1, \dots, x_N de X selon la loi conditionnelle à $Y = y$ sur la base de θ_k

- calcul de θ_{k+1} par :

$$\theta_{k+1} = \frac{1}{N} \left[\hat{\theta}(x_1, y) + \dots + \hat{\theta}(x_N, y) \right] \quad (13)$$

L'initialisation θ_0 peut être obtenue en appliquant $\hat{\theta}$ à la première segmentation x_0 faite par la stratégie bayésienne aveugle utilisant les paramètres estimés par l'algorithme SEM, dont le bon comportement dans ce contexte a été montré dans [33].

Nous obtenons ainsi quatre procédures d'estimation de θ à partir de Y , chacune correspondant à un estimateur $\hat{\alpha} = \hat{\alpha}(X)$ donné. Notons que cette richesse peut donner lieu à la conception d'une procédure d'estimation « combinée ». Le paramètre θ contient des paramètres réels de nature différente et il est vraisemblable qu'aucune des quatre procédures décrites ci-dessus ne soit meilleure que toutes les autres pour chacun de ces paramètres. On pourrait ainsi les programmer « en parallèle » et garder, dans chacune des quatre valeurs de θ obtenues, les composantes estimées avec le plus de précision.

Notons les possibilités d'application des procédures ci-dessus à la segmentation non supervisée d'images. En combinant les trois algorithmes de segmentation MAP, MPM, ICM mentionnés dans l'introduction avec les quatre estimateurs ECI ci-dessus nous obtenons douze algorithmes de segmentation non supervisée. On peut également envisager l'utilisation de MAP, MPM ou ICM avec la procédure « combinée » ci-dessus. Le comportement d'un algorithme de segmentation non supervisée est difficile à étudier, il dépend en effet des qualités des deux phases « estimation » et « segmentation » mais aussi de la robustesse de la méthode de segmentation choisie par rapport aux paramètres composant θ . Certains résultats relatifs à la robustesse de MAP, MPM et ICM par rapport au paramètre α sont exposés dans [18], [44].

4. ECI et gradient stochastique

Cette section comporte six sous-sections. Dans la première nous énonçons le principe de GSY, décrivons rapidement les calculs et donnons les formules de re-estimation auxquelles ils aboutissent. Les trois sous-sections suivantes sont consacrées à la discussion dont l'objectif est de situer la démarche ECI par rapport à celle de GSY. Ces sous-sections reprennent les points (i), (ii), (iii) mentionnés dans l'introduction. La cinquième sous-section contient des résultats numériques et dans la sixième nous rappelons un résultat concernant la convergence de GSY.

4.1. PRINCIPE ET FORMULES DE RE-ESTIMATION DE GSY

Notons f_θ la densité, par rapport au produit de la mesure de dénombrement par celle de Lebesgue, de la loi de (X, Y) . f_θ

est donnée par (voir (3) section 2) :

$$f_\theta(x, y) = ce^{-[U_\alpha(x) + V_{\beta,x}(y)]} \quad (14)$$

avec $\theta = (\alpha, \beta)$. Notons ϕ_θ , la densité de P_Y

$$\phi_\theta(y) = \sum_x f_\theta(x, y) \quad (15)$$

Le principe de GSY est celui du maximum de vraisemblance : pour y donné, on recherche θ pour lequel $\phi_\theta(y)$ est maximal. En notant $\Delta(\theta, x, y) \triangleq U_\alpha(x) + V_{\beta,x}(y)$, l'énergie de la distribution (14) et $\Delta'(\theta, x, y)$ son gradient par rapport à θ , l'auteur montre ([47], page 631) :

$$\frac{\partial}{\partial \theta} \text{Log}[\phi_\theta(y)] = E_\theta[\Delta'] - E_\theta[\Delta'/Y = y] \quad (16)$$

θ recherché annule donc (16). Notons que l'on procède par condition nécessaire : θ vérifiant (16) peut correspondre à un maximum local et ne pas maximiser (15).

L'idée de GSY est de présenter l'équation

$$E_\theta[\Delta'] - E_\theta[\Delta'/Y = y] = 0$$

sous la forme

$$E_\theta[Z_\theta] = 0 \quad (17)$$

où Z_θ est une certaine variable aléatoire dépendant de θ . Les équations de cette forme se prêtent bien à l'application des algorithmes du type « gradient stochastique » ([4]). La présentation de $E_\theta[\Delta'] - E_\theta[\Delta'/Y = y] = 0$ sous la forme $E_\theta[Z_\theta] = 0$ est possible en considérant trois champs aléatoires A, B, C où la distribution de (A, B) est celle de (X, Y) et la distribution de C celle de X conditionnelle à $Y = y$. En considérant la variable aléatoire $Z_\theta \triangleq \Delta'(\theta, A, B) - \Delta'(\theta, C, y) \triangleq h(\theta, A, B, C)$ l'équation $E_\theta[\Delta'] - E_\theta[\Delta'/Y = y] = 0$ équivaut à (17). L'application du gradient stochastique à (17) mène aux formules de re-estimation suivantes :

$$\theta_{k+1} = \theta_k + \frac{1}{d(k+1)} h(\theta_k, A_k, B_k, C_k) \quad (18)$$

où d est une constante et A_k, B_k, C_k prennent les valeurs, qui peuvent être simulées, des champs A, B, C dont la distribution correspond à la valeur courante θ_k du paramètre.

On obtient ainsi une méthode itérative différente de ECI mais dont la démarche générale est la même : la valeur suivante du paramètre est obtenue à partir de y , de sa valeur courante et des réalisations des simulations effectuées selon les lois rattachées à $P_{(X,Y)}$. Nous retrouvons ici l'importance de la possibilité d'effectuer de telles simulations.

4.2. APPLICABILITÉ DE GSY ET ECI

La généralité des modèles pouvant être traités par GSY et ECI est comparable. La démarche GSY n'utilise à aucun moment des propriétés liées à une forme particulière des densités f_1, \dots, f_m , elle est donc valable dans un cadre général, en particulier, ces densités peuvent avoir des formes différentes. Notons toutefois que si l'une de ces densités est non bornée, l'application du principe du maximum de vraisemblance à ϕ_θ pourrait s'avérer impossible. Cela constitue un avantage pour la ECI qui, en utilisant l'estimateur $\hat{\beta}$ décrit dans le paragraphe 3.2, reste applicable dans de telles situations. Cet avantage ne paraît pas déterminant dans le contexte où nous nous sommes placés : dans les problèmes de segmentation d'images, ces densités sont généralement supposées bornées.

4.3. DIFFICULTÉS DE MISE EN ŒUVRE

Les deux méthodes ne posent pas de difficultés de mise en œuvre particulières.

4.4. GÉNÉRALITÉ DE GSY ET ECI

Nous montrons dans ce paragraphe que les formules de re-estimation de GSY peuvent être obtenues différemment, par une démarche utilisant l'espérance conditionnelle.

Supposons X observable et considérons l'estimateur du gradient stochastique $\hat{\theta}$, que nous noterons EY et dont le principe est exposé dans [46]. Comme précédemment $\Delta(\theta, x, y)$ désigne l'énergie de la distribution du couple (X, Y) et $\Delta'(\theta, x, y)$ son gradient par rapport aux paramètres θ . Le principe de EY mène à la recherche de θ pour lequel

$$E_\theta[\Delta'(\theta, X, Y)] - \Delta'(\theta, x, y) = 0 \quad (19)$$

En prenant l'espérance conditionnelle de (19), on obtient

$$E_\theta[\Delta'(\theta, X, Y)] - E_\theta[\Delta'(\theta, X, Y)/Y = y] = 0 \quad (20)$$

Ce qui est précisément le principe de GSY.

Notons cependant que considérer l'espérance conditionnelle de l'égalité (19) ne revient pas à considérer l'espérance conditionnelle de l'estimateur $\hat{\theta}$ qu'elle définit implicitement. Par conséquent GSY n'apparaît pas comme étant nécessairement du type ECI.

4.5. EXEMPLE DE RÉSULTATS NUMÉRIQUES

On considère l'algorithme GSY noté, uniquement à l'intérieur de cette sous-section, A1. On applique ECI à $\hat{\theta} = \hat{\theta}(X, Y) = (\hat{\alpha}(X), \hat{\beta}(X, Y))$, où $\hat{\alpha}$ est EY et $\hat{\beta}$ l'estimateur décrit dans la

section 3.3. On obtient ainsi l'un des quatre estimateurs mentionnés dans la section 3.3, que l'on notera A2. Les algorithmes A1 et A2 peuvent être considérés comme deux généralisations différentes de l'algorithme EY au cas bruité; en effet, lorsque les variances du bruit tendent vers 0, on retrouve, dans les deux cas, l'algorithme EY.

Différents résultats concernant le comportement de A1 et A2 sont exposés dans [1]. Ces deux estimateurs ont été testés dans diverses situations et leur comportement est comparable, sauf en ce qui concerne les variances du bruit, dont l'estimation par A2 est toujours plus précise. Deux méthodes S1, S2 de segmentation statistique non supervisée, obtenues par l'adjonction à A1, A2 respectivement de l'algorithme du recuit simulé, ont également été testées. Les résultats obtenus par S2 sont toujours supérieurs, ce qui montre accessoirement que le recuit simulé est peu robuste par rapport aux variances du bruit.

Nous donnons ci-dessous quelques exemples numériques extraits de [1].

Considérons un champ de Markov X dont l'énergie est définie par

$$U(x) = \sum_{s \in S} \alpha(x_s) + \sum_{s, t \text{ voisins}} \delta(x_s, x_t) \quad (21)$$

Pour m classes la fonction α est définie par m valeurs

$$\alpha_1 = \alpha(\omega_1), \dots, \alpha_m = \alpha(\omega_m)$$

et nous considérons la fonction δ de la forme

$$\delta(x_s, x_t) = \begin{cases} \delta & \text{si } x_s = x_t \\ -\delta & \text{si } x_s \neq x_t \end{cases}$$

Les paramètres du bruit sont donnés par m moyennes μ_1, \dots, μ_m et m écarts type $\sigma_1, \dots, \sigma_m$. On considère $m = 4$. Les vraies valeurs des paramètres (VV) et les valeurs estimées par A1 et A2 sont données dans le tableau 1 ci-dessous (on utilise 256 niveaux de gris).

On constate que la qualité des deux algorithmes est équivalente, sauf en ce qui concerne l'estimation de la variance où A2 se montre plus performant. Lorsqu'on procède à la segmentation de l'image bruitée par l'algorithme de recuit simulé sur la base des paramètres estimés, les taux de bonne classification sont respectivement de 0.92 et 0.98. Notons que la qualité de l'estimation de δ est peu satisfaisante dans les deux cas, cependant le recuit simulé semble peu sensible à ce paramètre.

4.6. CONVERGENCE DE GSY

Comme nous l'avons mentionné plus haut, l'algorithme GSY apparaît comme une généralisation au cas de données incomplètes de l'algorithme EY. L'étude théorique de ce dernier, montrant la convergence presque sûre sous certaines conditions, est exposée dans [46]. La démonstration n'est pas transposable aux données incomplètes, toutefois il est possible de montrer la convergence de

Tableau 1

VV	A1	A2
$\alpha_1 = 0$	$\alpha_1 = -0.001$	$\alpha_1 = -0.002$
$\alpha_2 = 0$	$\alpha_2 = 0.049$	$\alpha_2 = 0.050$
$\alpha_3 = 0$	$\alpha_3 = -0.026$	$\alpha_3 = -0.019$
$\alpha_4 = 0$	$\alpha_4 = -0.022$	$\alpha_4 = -0.029$
$\delta = -5$	$\delta = -0.62$	$\delta = -0.65$

VV	A1	A2
$\mu_1 = 32$	$\mu_1 = 34.34$	$\mu_1 = 33.19$
$\mu_2 = 96$	$\mu_2 = 96.02$	$\mu_2 = 96.37$
$\mu_3 = 160$	$\mu_3 = 157.68$	$\mu_3 = 158.28$
$\mu_4 = 224$	$\mu_4 = 219.72$	$\mu_4 = 220.81$

VV	A1	A2
$\sigma_1 = 30$	$\sigma_1 = 17.89$	$\sigma_1 = 24.66$
$\sigma_2 = 30$	$\sigma_2 = 18.38$	$\sigma_2 = 26.23$
$\sigma_3 = 30$	$\sigma_3 = 18.81$	$\sigma_3 = 26.91$
$\sigma_4 = 30$	$\sigma_4 = 17.83$	$\sigma_4 = 24.72$

GSY à condition que l'initialisation soit « suffisamment proche » de la vraie valeur des paramètres ([47]). Bien que ce type d'hypothèse soit malaisé à vérifier dans la pratique ce résultat présente, sur le plan théorique, un avantage certain de l'algorithme GSY vis-à-vis de toutes les méthodes discutées dans cette étude.

5. ECI et EM Gibbsien

Nous reprenons dans cette section le plan de la section précédente en remplaçant l'algorithme GSY par l'algorithme EM Gibbsien.

5.1. PRINCIPE ET FORMULES DE RE-ESTIMATION DE EM GIBBSIEN

Le principe de l'algorithme EM Gibbsien est le principe général de l'algorithme EM. On considère, comme dans la section précédente, (voir (14)), la densité

$$f_\theta(x, y) = ce^{-[U_\alpha(x) + V_{\beta, x}(y)]} \quad (22)$$

avec $\theta = (\alpha, \beta)$. Le maximum de vraisemblance consisterait, si X était observable, en recherche de θ maximisant $\text{Log}[f_\theta(x, y)]$. X n'étant pas observable, on remplace la variable aléatoire $\text{Log}[f_\theta(X, Y)]$ par

$$E_\theta [\text{Log}[f_\theta(X, Y)/Y = y]] \quad (23)$$

Lorsqu'on désire rechercher la valeur de θ qui maximise (23), on se heurte à la même difficulté que dans le cas de l'ECI : l'espérance conditionnelle (23) dépend du paramètre. On arrive ainsi, comme dans les cas ECI ou GSY, à une méthode itérative : la valeur suivante du paramètre θ_{k+1} est donnée à partir de sa valeur courante θ_k et $Y = y$ par

$$E_{\theta_k} [\text{Log}[f_{\theta_{k+1}}(X, Y)/Y = y]] = \max_{\theta} E_{\theta_k} [\text{Log}[f_{\theta}(X, Y)/Y = y]] \quad (24)$$

La différence entre les principes de ECI et EM peut être mise en lumière à l'aide de l'exemple suivant. Supposons que l'estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ choisi en vue de l'application de l'ECI soit celui du maximum de vraisemblance. On recherche donc, pour $(X, Y) = (x, y)$, θ maximisant $\text{Log}[f_\theta(x, y)]$ puis on considère l'espérance conditionnelle, calculée sur la base de la valeur courante θ_k du paramètre, de la variable aléatoire $(X, Y) \rightarrow \theta(X, Y)$ ainsi obtenue. En suivant le principe du EM, on considère d'abord l'espérance conditionnelle, calculée sur la base de la valeur courante θ_k du paramètre, de $\text{Log}[f_\theta(X, Y)]$ puis on recherche θ maximisant la fonction ainsi obtenue. Ainsi, les deux principes s'intéressent à $\text{Log}[f_\theta(x, y)]$ et utilisent les opérations « recherche du maximum de vraisemblance » et « calcul de l'espérance conditionnelle » (qui ne commutent pas dans le cas général) dans un ordre différent.

Le principe de EM ne peut être rigoureusement appliqué dans le cadre de la modélisation retenue et l'auteur est amené à remplacer la densité de la loi P_X , qui intervient dans $\text{Log}[f_\theta(x, y)]$, par une fonction différente, dite « pseudo-vraisemblance ». Ce point fait l'objet du paragraphe 5.3 ci-dessous.

Dans la modélisation retenue par l'auteur, le paramètre α n'apparaît pas dans l'énergie mais directement au niveau des probabilités conditionnelles. En considérant une image binaire ($\Omega = \{\omega_1, \omega_2\}$) et les voisinages composés des quatre plus proches voisins, on distingue cinq types de réalisations de la restriction de X , au voisinage donné : $j = 0, \dots, 4$. Un voisinage est de type j s'il contient j classes ω_1 . On pose alors

$$\alpha = (\alpha_{i,j})_{1 \leq i \leq 2, 0 \leq j \leq 4},$$

avec

$$\alpha_{i,j} = P[X_s = \omega_i / V_s \text{ est de type } j].$$

Le champ Y est supposé gaussien conditionnellement à X et les variables $(Y_s)_{s \in S}$ indépendantes conditionnellement à X . Moyennant la substitution de la pseudo-vraisemblance à la vraisemblance, il est possible de déduire du principe général les formules de re-estimation suivantes :

$$\alpha_{ij}^{k+1} = \frac{E_k \left[\sum_{s \in S} 1_{[X_s = \omega_i \text{ et } V_s \text{ de la forme } j]} \right]}{\sum_i E_k \left[\sum_{s \in S} 1_{[X_s = \omega_i \text{ et } V_s \text{ de la forme } j]} \right]} \quad (25)$$

où E_k est l'espérance conditionnelle à $Y = y$ basée sur θ_k .

$$\mu_j^{k+1} = \frac{\sum_{s \in S} P_{s,y}(\omega_i) y_s}{\sum_{s \in S} P_{s,y}(\omega_i)} \quad (26)$$

$$\sigma^{2,k+1} = \frac{\sum_i \sum_{s \in S} P_{s,y}(\omega_i) (y_s - \mu_i^{k+1})^2}{\sum_i \sum_{s \in S} P_{s,y}(\omega_i)} \quad (27)$$

où $1 \leq i \leq 2$ et $P_{s,y}$ désigne la loi de X_s conditionnelle à $Y = y$, calculée sur la base de la valeur courante du paramètre.

5.2. APPLICABILITÉ DE EMG ET ECI

Les conditions de l'applicabilité de l'EMG apparaissent, par rapport à celles de l'applicabilité de ECI exposées dans le paragraphe 4.2, comme moins générales. Les densités f_1, \dots, f_m sont gaussiennes et ne se différencient que par leurs moyennes, la variance étant supposée commune à toutes les classes. Notons cependant que cette dernière restriction peut être aisément levée par une modification de la formule (27) : elle peut être remplacée par deux formules permettant la re-estimation de deux variances, éventuellement différentes, correspondant aux deux classes. Les résultats numériques présentés dans la sous-section 5.5 ont été obtenus par une version de l'EMG ainsi modifiée. Le caractère gaussien du bruit étant toutefois difficilement contournable, l'EMG apparaît, vis-à-vis de ECI ou GSY, comme plus restrictif sur le plan de la généralité des modèles pouvant être traités.

5.3. DIFFICULTÉS DE MISE EN ŒUVRE

Comme dans le paragraphe 4.3, nous nous intéressons ici aux difficultés liées à la recherche des formules de re-estimation à partir du principe général. Dans le cas de l'EMG, on est amené à remplacer la vraisemblance de la loi a priori $f_\alpha(x) = ce^{-U_\alpha(x)}$ par la pseudo-vraisemblance, qui est le produit, pour $s \in S$, des lois des X_s conditionnelles à $(X_t)_{t \neq s}$. Nous avons vu dans la section 2 que ces probabilités sont explicitement calculables. Ainsi, contrairement à la vraisemblance f_α , la valeur de la pseudo-vraisemblance est calculable en tout $X = x$. Notons que la pseudo-vraisemblance est égale à la vraisemblance dans le cas des données indépendantes. On peut ainsi imaginer que la validité de son utilisation augmente lorsque la dépendance des données diminue. Par ailleurs, lorsque la dépendance des données n'est pas suffisamment importante, l'utilisation du modèle markovien peut ne plus être justifiée. L'utilisation simultanée du modèle markovien et de la pseudo-vraisemblance pose ainsi un problème de cohérence. Notons cependant qu'il est possible de montrer la convergence presque sûre de l'estimateur du pseudo-maximum de vraisemblance ainsi obtenu ([24]).

5.4. GÉNÉRALITÉ DE EMG ET ECI

Nous montrons dans ce paragraphe que les formules de re-estimation de l'EMG peuvent être obtenues par une procédure ECI particulière.

Conformément au principe général de l'ECI supposons X observable et recherchons $\hat{\theta} = \hat{\theta}(X, Y)$ un estimateur de θ . Considérons la fréquence empirique pour $\hat{\alpha}$ et l'estimateur $\hat{\beta}$ défini dans la section 3. Plus précisément $\hat{\alpha}$ et $\hat{\beta}$ sont définis par :

$$\hat{\alpha}_{i,j}(X) = \frac{\sum_{s \in S} 1_{[X_s = \omega_i \text{ et } V_s \text{ de type } j]}}{\sum_{s \in S} 1_{[V_s \text{ de type } j]}} \quad (28)$$

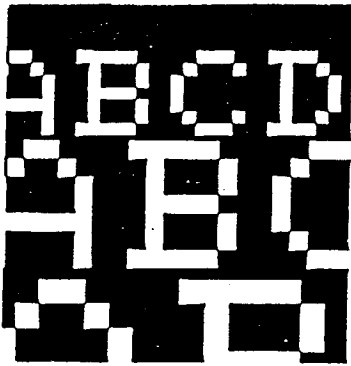
$$\hat{\mu}_i(X, Y) = \frac{\sum_{s \in S} Y_s 1_{[X_s = \omega_i]}}{\sum_{s \in S} 1_{[X_s = \omega_i]}} \quad (29)$$

$$\hat{\sigma}^2(X, Y) = \frac{\sum_i \sum_{s \in S} (Y_s 1_{[X_s = \omega_i]} - \mu_i)^2}{\sum_i \sum_{s \in S} 1_{[X_s = \omega_i]}} \quad (30)$$

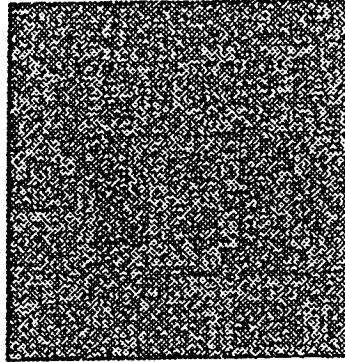
Notons que le dénominateur de (30) est égal au cardinal de S , nous l'avons présenté sous cette forme en vue de la suite de l'étude.

Chacun des estimateurs (28), (29), (30) est de la forme $\frac{U}{V}$. Nous devons, afin d'obtenir les formules de re-estimation de ECI, considérer leur espérance conditionnelle $E_k[\frac{U}{V}]$, calculée sur la base de la valeur courante du paramètre. L'application de E_k à $\frac{U}{V}$ ne mène pas aux formules (25), (26), (27) mais donne un algorithme d'estimation original. Le calcul explicite de $E_k[\frac{U}{V}]$ n'est pas possible et nous devons faire appel à l'approximation (5). Notons que tel est également le cas lorsqu'on utilise l'EMG : les probabilités marginales a posteriori $P_{s,y}$ figurant dans les formules (25), (26) ne sont pas calculables et nous devons les estimer à partir des réalisations de X simulées selon la loi a posteriori.

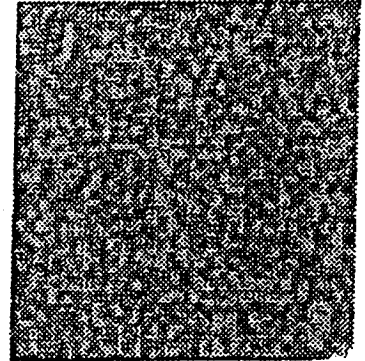
Les formules (25), (26), (27) peuvent être obtenues par une procédure ECI en considérant le problème du calcul de $E_k[\frac{U}{V}]$ sous un angle légèrement différent. Chacune des variables aléatoires U, V est un estimateur d'un certain paramètre. Supposons que l'on s'intéresse à chacun d'eux séparément : on considère, à chaque étape de l'ECI, les espérances conditionnelles $E_k[U/Y = y]$, $E_k[V/Y = y]$ et, après les avoir calculées ou approchées, on pose pour la valeur suivante du paramètre $\frac{E_k[U/Y=y]}{E_k[V/Y=y]}$. L'application de cette démarche à (28), (29), (30) respectivement donne directement (25), (26), (27).



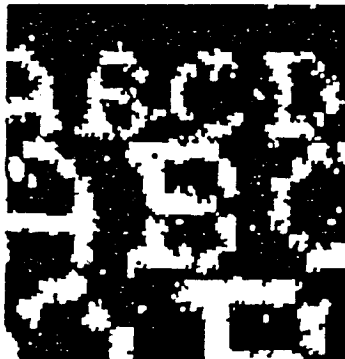
IM.1



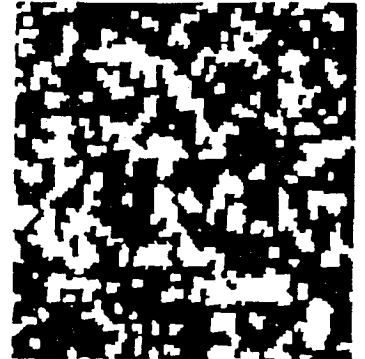
IM.2
IM.1+Bruit indépendant



IM.3
IM.1+Bruit corrélé



IM.4
Segmentée de IM.2 par
ECI+MPM



IM.5
Segmentée de IM.3
par ECI+MPM

5.5. RÉSULTATS NUMÉRIQUES

Nous donnons ci-dessous deux exemples d'estimation de paramètres par EMG et l'ECI décrite dans la section 5.4. L'image originale (IM.1) est bruitée par un bruit blanc (IM.2) et un bruit corrélé (IM.3) dont les paramètres sont donnés dans les tableaux Tab.2 et Tab. 3 respectivement. Les images IM.2 et IM.3 sont ensuite segmentées par l'algorithme Marroquin et al. (MPM) sur la base des paramètres estimés. Les images IM.4 et IM.5 sont les résultats de la segmentation de IM.2 et IM.3 respectivement par la méthode non supervisée ECI+MPM. Les résultats de leur segmentation par la méthode non supervisée EMG+MPM, que nous ne présentons pas, sont équivalents. VV désigne les vraies valeurs de paramètres. L'initialisation de deux méthodes est faite à partir de la segmentation classique « pixel par pixel » utilisant les paramètres estimés par le SEM aveugle. τ (MPM) désigne le taux de pixels mal classés par l'algorithme de Marroquin et al. fondé sur les valeurs estimées des paramètres.

Tableau 2 (bruit blanc)

	μ_1	μ_2	σ_1	σ_2	τ (MPM)
VV	110	110	15	20	
EMG	109,5	110,1	14,9	30,8	7,65
ECI	109,5	110	14,9	31,6	7,87

Tableau 3 (bruit corrélé)

	μ_1	μ_2	σ_1	σ_2	τ (MPM)
VV	110	110	15	30	
EMG	102,4	128,6	15,1	14,2	33,68
ECI	102,8	129,8	15,2	14	32,74

Nous pouvons constater l'équivalence des comportements des deux méthodes dans ce cas précis : le comportement de la segmentation non supervisée semble ainsi être principalement régi

par la phase « segmentation », l'application de l'ECI ou l'EMG dans la phase « estimation des paramètres » apparaissant comme secondaire.

Notons que le bruitage est relativement important : dans le cas de la segmentation aveugle (« pixel par pixel ») utilisant les vraies valeurs des paramètres le taux théorique de pixels mal classé est de 25%. Ce taux est par ailleurs identique dans les deux cas, en effet, les deux bruitages se différencient uniquement par les corrélations et sont identiques lorsqu'on considère les pixels séparément. Cet exemple confirme la grande efficacité des méthodes markoviennes lorsqu'on est en présence d'une bonne adéquation du modèle (bruit blanc : le taux passe de 25% à 8%) mais également leur sensibilité à la corrélation du bruit (le taux passe de 25% à 33%), ce qui corrobore l'utilité des méthodes contextuelles, moins sensibles à la corrélation du bruit, dans certaines situations.

6. Méthodes « segmentation courante » et l'algorithme de S. Lakshaman et H. Derin

Cette section est consacrée à la revue de deux méthodes d'estimation dans les champs markoviens cachés. Dans chacun des cas étudiés, nous proposons une description rapide de l'algorithme et situons sa démarche par rapport à celle de l'ECI.

6.1. MÉTHODES « SEGMENTATION COURANTE »

Nous appellerons une méthode « segmentation courante » (SC) toute méthode itérative dans laquelle la valeur suivante du paramètre θ_{k+1} est obtenue en appliquant une méthode d'estimation donnée à (x_k, y) , où $x_k = s_k(y)$ est une segmentation de y par une stratégie s_k utilisant la valeur courante θ_k du paramètre. Le principe est donc le suivant : on se donne un estimateur $\hat{\theta} = \hat{\theta}(X, Y)$, une stratégie de segmentation s_θ et on pose :

$$\theta_{k+1} = \hat{\theta}(s_{\theta_k}(y), y) \quad (31)$$

Un certain nombre de telles méthodes a été proposé par les auteurs ([5], [16], [22]).

Notons que cette démarche ressemble à la ECI. Les deux méthodes utilisent un estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ et on re-estime les paramètres en l'appliquant à des réalisations de X obtenues « artificiellement ». Lorsqu'on considère une méthode de type SC, ces réalisations sont obtenues par une segmentation et, dans le cas d'une méthode de type ECI, elles le sont par des tirages aléatoires. Nous avons vu qu'il existe, pour l'ECI, une justification partielle utilisant la minimisation de l'erreur quadratique moyenne. Dans le cas de la SC, il est difficile de mettre en avant une justification précise. La qualité d'une méthode de segmentation répond en effet à un critère de qualité précis, lié à la fonction de perte. Le choix de

la fonction de perte étant indépendant du paramétrage, la méthode SC contient, en associant deux problèmes de type différent, une certaine incohérence interne. Considérons, afin d'illustrer cette incohérence, un exemple simple de mélange de deux lois, avec la fonction de perte $L(\omega_1, \omega_2) = 1$, $L(\omega_2, \omega_1) = \delta$, $\delta > 1$. π_1, π_2 étant la loi a priori et f_1, f_2 les densités conditionnelles, la règle bayésienne s'écrit :

$$s_B(y) = \begin{cases} \omega_1 & \text{si } \delta\pi_1 f_1(y) > \pi_2 f_2(y) \\ \omega_2 & \text{si } \delta\pi_1 f_1(y) < \pi_2 f_2(y) \end{cases} \quad (32)$$

Lorsqu'on désire appliquer la procédure SC à un échantillon y_1, \dots, y_n de Y on doit, afin d'obtenir un échantillon « artificiel » de X , appliquer le règle ci-dessus à chaque y_i . On constate que l'hypothèse $\delta > 1$ implique la sur-représentation de ω_1 dans cet échantillon, ce qui entraîne une surestimation de π_1 . Ce dernier fait implique, au pas suivant de l'algorithme, une sur-représentation de ω_1 encore plus forte. Le phénomène se reproduisant à chaque itération, nous pouvons constater comment la dissymétrie de la fonction de perte, qui constitue dans certains cas son principal intérêt, peut impliquer une instabilité de la procédure SC.

6.2. ALGORITHME DE S. LAKSHMANAN ET H. DERIN

La démarche originale de cet algorithme ([28]), que nous appellerons ALD, le situe à mi-chemin entre les méthodes de type « estimation conditionnelle itérative » et celles de type « segmentation courante ». Les auteurs proposent la solution simultanée des problèmes d'estimation des paramètres et de segmentation, cette dernière étant traitée par le recuit simulé. L'algorithme du recuit simulé utilise un paramètre T , dit « température », tel que pour $T = 1$ la loi de X a posteriori est sa vraie loi a posteriori. Cet algorithme consiste à diminuer T par paliers, en faisant fonctionner, sur chaque palier, l'échantillonneur de Gibbs. On montre que la suite aléatoire des x ainsi obtenue converge vers la solution du MAP. L'idée de ALD est d'interrompre régulièrement le déroulement du recuit simulé et d'utiliser la valeur courante x du champ X afin de re-estimer, comme dans ECI ou SC, tous les paramètres. Ainsi ALD est du type ECI au début de son déroulement : lorsque T est proche de 1, la loi a posteriori utilisant T est proche de la vraie loi a posteriori. Lorsque T est proche de 0, ALD devient du type SC, en effet, les simulations par l'échantillonneur de Gibbs s'apparentent à la segmentation. Finalement nous pouvons dire qu'au cours de son déroulement ALD passe, d'une manière « continue », du type ECI au type SC.

L'algorithme ADL admet une justification théorique partielle : supposant les variances du bruit connues, les auteurs montrent la convergence de la suite de paramètres obtenue vers le maximum de vraisemblance et celle de la suite d'images de classes obtenue vers une solution de MAP.

7. Conclusion

Nous avons traité, dans cet article, du problème de l'estimation des paramètres dans les champs de Markov cachés. La modélisation par de tels champs permet la conception de nombreux algorithmes de segmentation statistique dont la grande efficacité, dans de nombreuses situations, a été montrée par certains auteurs ces dernières années. La solution du problème de l'estimation des paramètres mène à la possibilité de conception des algorithmes non supervisés. L'importance du caractère non supervisé des algorithmes devient primordiale lorsqu'on cherche à les intégrer dans un système plus complexe, et automatique, de traitement d'images.

L'objectif de notre étude était de montrer la richesse des possibilités de conception des algorithmes d'estimation des paramètres dans les champs de Markov cachés à partir d'une méthode récente d'estimation dans le cas de données cachées, dite « Estimation Conditionnelle Itérative » (ECI). Le principe général de ECI étant différent de ceux des méthodes « Expectation-Maximization » (EM) et « Gradient Stochastique » (GS), les algorithmes que l'on peut en déduire sont, en général, originaux par rapport à ceux déduits des méthodes EM ou GS. Par ailleurs, n'utilisant pas la notion de vraisemblance, le principe de l'ECI reste valable dans des modélisations relativement générales. Enfin, l'ECI peut apparaître comme une généralisation de l'EM dans la mesure où les formules de re-estimation de ce dernier peuvent être obtenues par une démarche de type ECI particulière. Ce type de situations reste toutefois exceptionnel et la famille de méthodes présentée apparaît plutôt comme complémentaire aux méthodes existantes que comme leur généralisation. En particulier, l'ECI apparaît comme complémentaire par rapport à la méthode utilisant les segmentations effectuées sur la base de la valeur courante des paramètres ([5], [16], [22]) et par rapport à l'algorithme de S. Lakshmanan et H. Derin ([28]).

Les résultats numériques obtenus avec des algorithmes de type ECI montrent leur bonne compétitivité vis-à-vis des méthodes de type EM ou GS, cependant, nous ne pouvons conclure à sa supériorité dans tout type de situations. Des études complémentaires, théoriques ou fondées sur des simulations, sont nécessaires pour mieux situer le comportement des diverses méthodes dans les différentes situations rencontrées. Il est vraisemblable qu'aucune méthode particulière ne soit meilleure que toutes les autres dans tout type de situations et de telles études pourraient mener à la conception d'une méthode « généralisée » contenant plusieurs méthodes particulières et, en amont, un algorithme de choix automatique de la méthode la mieux adaptée aux données en présence.

BIBLIOGRAPHIE

[1] O. ALLAGNAT, J.M. BOUCHER, D.C. HE, W. PIECZYNSKI (1992). Hidden Markov fields and unsupervised segmentation of images, *IAPR 11th*

International Conference on Pattern Recognition (ICPR 92), Delft, Pays-Bas, septembre 92.

[2] R. AZENCOTT (1988). Image analysis and Markov fields, *Proc. of the Int. Conf. on Ind. and Appl. Math.*, SIAM, Paris 1987. Siam Philadelphia 1988.

[3] R. AZENCOTT Ed. (1992). Simulated annealing : Parallelization techniques, Wiley.

[4] A. BENVENISTE, M. MÉTIVIER, B. PRIOURET (1987). Algorithmes adaptatifs et approximations stochastiques, Collection *Techniques Stochastiques*, Masson, Paris.

[5] J. BESAG (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society*, Series B, 6, pp. 192-236.

[6] J. BESAG (1986). On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society*, Series B, 48, pp. 259-302.

[7] B. BRAATHEN, N. MARHIC, P. MASSON, W. PIECZYNSKI (1991). Sur une nouvelle approche de segmentation Bayésienne non supervisée d'images, *Treizième colloque GRETSI*, Juan-les-Pins, France, septembre 1991.

[8] B. BRAATHEN, W. PIECZYNSKI, P. MASSON (1993). Global and local methods of unsupervised Bayesian segmentation of images, *Machine Graphics & Vision*, Vol. 2, N°1, pp. 39-52.

[9] H. CAILLOL, A. HILLON, W. PIECZYNSKI (1993). Fuzzy random fields and unsupervised image segmentation, *IEEE Transactions on GRS*, Vol. 31, N°4, pp. 801-810.

[10] H. CAILLOL, A. HILLON, W. PIECZYNSKI (1994). Segmentation contextuelle non supervisée utilisant une modélisation statistique floue, *RFIA '94*, 11-14 janvier 94, Paris, France.

[11] G. CELEUX, J. DIEBOLT (1986). L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités, *Revue de Statistique Appliquée*, Vol. 34, N°2.

[12] B. CHALMOND (1988). Image restauration using an estimated Markov model, *Signal Processing*, 15, pp. 115-129.

[13] B. CHALMOND (1989). An iterative Gibbsian technique for reconstruction of m-ary images, *Pattern Recognition*, Vol. 22, N°6, pp. 747-761.

[14] C.C. CHEN (1988). Markov random field models in image analysis, Ph. D. Thesis, East Lansing, Michigan State University.

[15] M.M. DEMPSTER, N.M. LAIRD, D.B. RUBIN (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, Series B, 39, pp. 1-38.

[16] H. DERIN, H. ELLIOT (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields, *IEEE Transactions on PAMI*, Vol. 9, pp. 39-55.

[17] P.A. DEVIJVER, M. DEKESEL (1988). Champs aléatoires de Pickard et modélisation d'images digitales, *Traitement du Signal*, Vol. 5, N°5.

[18] J.M. DINTEN, X. GUYON, J.F. YAO (1988). On the choice of the regularization parameter : the case of binary images in the Bayesian restoration framework, *Proc. of an AMS-IMS-SIAM Joint Summer Research Conference* (Ed. A. Possolo), Maine, pp. 55-77.

[19] R.C. DUBES, A.K. JAIN (1989). Random field models in images analysis, *Journal of Applied Statistics*, Vol. 16, N°2, pp. 131-164.

[20] S. GEMAN, D. GEMAN (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on PAMI*, Vol. 6, pp. 721-741.

[21] S. GEMAN, D. GEMAN, C. GRAFFIGNE (1987). Locating texture and object boundaries - *Pattern recognition theory and applications*, P.A. Devijver et J. Kittler ed., Heidelberg, Springer-Verlag.

[22] S. GEMAN, D.E. McCLURE (1985). Bayesian image analysis : An application to single photon emission tomography, *Proc. Amer. Stat. Assoc. : Statistical Computing Section*, pp. 12-18.

[23] X. GUYON (1982). Parameter estimation for a stationary process on a d-dimensional lattice, *Biometrika* 69, pp. 95-105.

- [24] X. GUYON (1993). Champs aléatoires sur un réseau, *Collection Techniques Stochastiques*, Masson, Paris.
- [25] R. HARALICK, J. HYONAM (1986). A context classifier, *IEEE Transactions on GRS GE-24*, pp. 997-1007.
- [26] A. HILLION (1992). Les approches statistiques pour la reconnaissance des images de télédétection, *Atti della XXXVI Riunione Scientifica*, SIS, Vol. 1, pp. 287-297.
- [27] P.A. KELLY, H. DERIN, K.D. HARTT (1988). Adaptive segmentation of speckled images using a hierarchical random field model, *IEEE Transactions on ASSP*, Vol. 36, N° 10 pp. 1628-1641.
- [28] S. LAKSHMANAN, H. DERIN (1989). Simultaneous parameter estimation and segmentation of Gibbs random fields, *IEEE Transactions on PAMI*, Vol. 11, pp. 799-813.
- [29] N. MARHIC, P. MASSON, W. PIECZYNSKI (1991). Mélange de lois et segmentation non supervisée des données SPOT, *Statistique et Analyse des Données*, Vol. 16, N° 2.
- [30] J. MARROQUIN, S. MITTER, T. POGGIO (1987). Probabilistic solution of ill-posed problems in computational vision, *Journal of the American Statistical Association*, 82, pp. 76-89.
- [31] A. MARZOUKI, Y. DELIGNON, H.C. QUELLE, W. PIECZYNSKI (1993). Segmentation non supervisée d'images satellite utilisant un modèle hiérarchique généralisé, *Quatorzième Colloque GRETSI 93*, Juan-les-Pins, France, septembre 1993.
- [32] P. MASSON, G. DUBOIS, G. LE LANN, A. HILLION, C. ROUX (1990). Classification d'images du satellite SPOT pour la cartographie bathymétrique du littoral, *Congrès « PIX-ILES 90 » Journées Internationales*, Nouméa-Tahiti, novembre 1990.
- [33] P. MASSON, W. PIECZYNSKI (1993). SEM algorithm and unsupervised segmentation of satellite images, *IEEE Transactions on GRS*, Vol. 31, N° 3, pp. 618-633.
- [34] E. MOHN, N. HJORT, G. STORVIC (1987). A simulation study of some contextual classification methods for remotely sensed data, *IEEE Transactions on GRS*, GE, 25, pp. 796-804.
- [35] L. MOREAU, P.M. ALLIOUX, C. BANGA, F. GHORBEL. « Segmentation bayésienne par échantillonnage Bootstrap : application à une séquence d'images de fond de l'œil », *Quatorzième Colloque GRETSI 93*, Juan-les-Pins, France, septembre 1993.
- [36] W. PIECZYNSKI (1989). Estimation of context in random fields, *Journal of Applied Statistics*, Vol. 16, N° 2, pp. 283-290.
- [37] W. PIECZYNSKI (1990). Mixture of distributions, Markov random fields and unsupervised Bayesian segmentation of images, Rapport Technique N° 122, L.S.T.A., Université de Paris VI.
- [38] B. PRUM (1986). Processus sur un réseau et mesures de Gibbs, Applications, Masson.
- [39] W. QIAN, D.M. TITTERINGTON (1989). On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions, *Journal of Applied Statistics*, Vol. 16, N° 2, pp. 267-282.
- [40] P. SWAIN, S. VARDEMAN, J. TILTON (1981). Contextual classification of multispectral image data, *Pattern Recognition*, 13, pp. 429-441.
- [41] J. TILTON, S. VARDEMAN, P. SWAIN (1982). Estimation of context for statistical classification of multispectral image data, *IEEE Transactions on GRS*, GE-20, pp. 445-452.
- [42] A. VEIJANEN (1991). A simulation-based estimator for hidden Markov random fields, *IEEE Transactions on PAMI*, Vol. 13, N° 8, pp. 825-830.
- [43] C.S. WON, H. DERIN (1992). Unsupervised segmentation of noisy and textured images using Markov random fields, *CVGIP : Graphical models and image processing*, Vol. 54, N° 4.
- [44] J.F. YAO (1988). Méthodes Bayésiennes en segmentation d'images : *Bayesian Statistics, Proc. 8-th Franco-Belgian Meeting of Statisticians*, CORE, Université Catholique de Louvain, Belgique.
- [45] J.F. YAO (1989). Segmentation bayésienne d'images : comparaisons de méthodes contextuelles et globales, *Cahier du Centre d'Études et de Recherches Op. 30 (4)*, pp. 269-290.
- [46] L. YOUNES (1988). Estimation and annealing for Gibbsian fields, *Annales de l'Institut Henri Poincaré*, 2.
- [47] L. YOUNES (1989). Parametric inference for imperfectly observed Gibbsian fields, *Probability Theory and Related Fields*, 82, pp. 625-645.

L'AUTEUR

Titulaire d'un Doctorat d'État en Mathématiques obtenu à l'Université Paris VI en 1986, Wojciech Pieczynski a enseigné au Centre Universitaire de Tiaret, Algérie, à l'Université de Brazzaville, Congo, et à l'École Nationale Supérieure des Télécommunications de Bretagne, Brest. Actuellement Professeur à l'Institut National des Télécommunications, Évry. Ses recherches portent sur les modélisations probabilistes et les traitements statistiques des images.