FORMULES DE RÉESTIMATION POUR UN MODÈLE DE CHAÎNE DE MARKOV CACHÉE STATIONNAIRE RÉVERSIBLE(1)

Jérôme IDIER† et Yves GOUSSARD‡

†Laboratoire des signaux et systèmes – Supélec, Plateau de Moulon 91192 Gif-sur-Yvette Cedex, France – idier@lss.supelec.fr

‡École polytechnique – Biomedical Engineering Institute – P.O. Box 6079, Station "centre-ville", Montréal (Québec) H3C 3A7, Canada – yves@grbb.polymtl.ca

RÉSUMÉ

Cette communication présente une méthode d'estimation EM pour la loi d'une chaîne de Markov cachée, dans le cas d'une loi paramétrée de type « télégraphique ». Cette classe de modèles assure la stationnarité et la réversibilité de la loi, propriétés utiles et naturelles en segmentation de signaux et d'images. Il s'agit d'autre part, à notre connaissance, d'un cas unique d'algorithme EM adapté à un modèle paramétré de chaîne de Markov cachée. A titre d'application, PERROT *et coll.* [8] envisagent le problème de la segmentation d'images médicales.

ABSTRACT

This communication presents an EM estimation method in a specific case of parametric hidden Markov chains, namely under the "telegraphic" assumption. Such models ensure stationarity and reversibilty of the probability law, both desirable and natural properties for the purpose of signal and image segmentation. On the other hand, we are aware of no other parametric family of hidden Markov chains which yields a devisable EM algorithm. Application to medical image segmentation is proposed in an accompanying paper by PERROT *et al. [8].*

1. MOTIVATION

Une chaîne de Markov homogène quelconque sur N niveaux se décrit à l'aide de sa probabilité initiale $\boldsymbol{p}$ et de sa matrice de transition $\boldsymbol{P}$, soit au total $N(N+1)$ paramètres et $N+1$ contraintes d'égalité, donc N^2-1 degrés de liberté. C'est à l'identification de ce modèle général que s'appliquent les formules de réestimation de Baum-Welch, algorithme EM adapté au cas où la chaîne est indirectement observée [1].

Dans cette communication, l'idée de départ est de reconnaître aux équations de Baum-Welch le défaut de leur généralité, c'est-à-dire qu'elles ne permettent ni d'incorporer des contraintes *a priori* naturelles telle que stationnarité, ou même réversibilité ; ni d'incorporer un *a priori* plus fort qui se traduirait par une paramétrisation parcimonieuse de la mesure de probabilité de la chaîne, *via* une reparamétrisation de $\boldsymbol{p}$ et de $\boldsymbol{P}$. Stationnarité et réversibilité fournissent respectivement $N(N-1)$ et $(N/2+1)(N-1)$ degrés de liberté (la seconde impliquant la première pour un modèle homogène). Dans les deux cas, le nombre de paramètres reste de l'ordre du carré du nombre N de niveaux occupés par la chaîne, donc un apprentissage réaliste ne peut être envisagé que pour des chaînes à petit nombre de niveaux.

L'objectif de ce travail est de réduire le nombre de paramètres par l'introduction d'un *a priori* pertinent, en respectant simultanément deux contraintes : la réversibilité d'une part, la possibilité de mettre en œuvre un algorithme EM d'autre part. La pertinence du modèle est relative aux applications qui nous intéressent en traitement du signal et des images. Il s'agit essentiellement de problèmes de segmentation, pour lesquels le prototype du signal à restaurer est une fonction constante par morceaux. Outre l'hypothèse markovienne classique, la clé de notre travail réside dans une hypothèse supplémentaire d'indépendance entre les niveaux constants occupés par le modèle. Autrement dit, le passage d'un état à un état différent est considéré comme une rupture engendrant la décorrélation. Ce type d'hypothèse est déjà courant en traitement d'image : elle est implicite pour la plupart des champs de Markov composites proposés dans la littérature, dont l'énergie est une somme de potentiels positifs annulé par activation de variables de ligne : [4], [2]. Le contraire est plus rare : par exemple [6], qui proposent des variables de ligne réduisant les potentiels sans les annuler.

[1] Le travail présenté a bénéficié d'un soutien partiel du Ministère des affaires étrangères français et du Ministère de l'enseignement supérieur et de la science du Québec, projet N°04-02-94.

2. MODÈLE DE CHAÎNE DE MARKOV TÉLÉGRAPHIQUE

Le modèle proposé est une généralisation simple de la chaîne « télégraphique » utilisée par [5] pour la segmentation de signaux sismiques. La matrice de transition du modèle généralisé s'écrit

$$\boldsymbol{P} = (P_{mn}) = \boldsymbol{\Lambda} + (\boldsymbol{1}-\boldsymbol{\lambda})\boldsymbol{\mu}^t, \tag{1}$$

avec $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda})$ et $\boldsymbol{1} = (1,\ldots,1)^t$. Dans ces conditions,

- partant d'un état m, on décide de changer d'état avec la probabilité $1-\lambda_m$; le choix du nouvel état est

indépendant de l'état courant m et la probabilité de choisir l'état n est μ_n ;
• comme on peut retomber sur l'ancien état m avec la probabilité μ_m, la probabilité de rester dans l'état m est $\lambda_m + \mu_m - \lambda_m\mu_m$;
• la probabilité stationnaire de la chaîne est

$$\boldsymbol{p} = (p_n) = (\boldsymbol{I} - \boldsymbol{\Lambda} + \boldsymbol{\mu}\boldsymbol{\lambda}^{\mathrm{t}})^{-1}\boldsymbol{\mu},$$

vecteur qui s'exprime plus simplement composante par composante :

$$p_n = \frac{\mu_n}{1-\lambda_n} / \sum_m \frac{\mu_m}{1-\lambda_m} \; ; \tag{2}$$

• dans son état stationnaire, la chaîne est réversible : $P_{mn}p_m = P_{nm}p_n$ pour tout m, n indexant l'espace d'état.

La paramétrisation de la chaîne se réduit aux deux vecteurs $\boldsymbol{\lambda}$ et $\boldsymbol{\mu}$, soit $2N$ paramètres. Pour le vecteur de probabilité $\boldsymbol{\mu}$, on a les contraintes classiques(1).

$$\mathbf{1}^{\mathrm{t}}\boldsymbol{\mu} = 1 \quad \text{et} \quad \boldsymbol{\mu} > \mathbf{0}. \tag{3}$$

Les contraintes affectant $\boldsymbol{\lambda}$, qui n'est pas un vecteur de probabilité, sont plus inattendues :

$$\boldsymbol{\lambda} < \mathbf{1} \quad \text{et} \quad \boldsymbol{\lambda} > -\boldsymbol{\mu}/(\mathbf{1} - \boldsymbol{\mu}) \; ; \tag{4}$$

le second jeu d'inégalités est moins intuitif que la condition naturelle $\boldsymbol{\lambda} > \mathbf{0}$, car il retire à $\boldsymbol{\lambda}$ la signification d'un vecteur de probabilité de changement d'état. Il s'avère néanmoins suffisant (et nécessaire) pour assurer la validité mathématique du modèle.

3. CHAÎNE DE MARKOV CACHÉE (CMC)

3.1. Définition

Une CMC est décrite par deux composantes : l'une est une chaîne de Markov $\{X_t\}$: pour tout T, $\boldsymbol{X} = (X_1, \ldots, X_T)$, et tout $\boldsymbol{x} = (x_1, \ldots, x_T)$,

$$P(\boldsymbol{X} = \boldsymbol{x}) = p_{x_1} \prod_{t=2}^{T} P_{x_{t-1}x_t} \; ; \tag{5}$$

l'autre un signal observé $\{Y_t\}$ dont les composantes résultent d'une transformation aléatoire instantanée de $\{X_t\}$:

$$f_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{t=1}^{T} f_{Y|X}(y_t \mid x_t).$$

Le signal $\{Y_t\}$ est ici supposé à valeurs réelles, *via* la densité de probabilité $f_{Y|X}$. La loi gaussienne est la plus utilisée. Sa description nécessite le vecteur $\boldsymbol{m}$ des moyennes conditionnelles $m_n = \mathrm{E}(Y|X = n)$ et celui des variances $\sigma_n^2 = \mathrm{var}(Y|X = n)$, soit au total les N couples de moyenne et de variance $\boldsymbol{\theta}_{Y|X} = (\boldsymbol{m}, \boldsymbol{\sigma}^2)$.

[1] Ici et dans la suite du texte, les inégalités vectorielles sont des écritures compactes à interpréter composante par composante.

Finalement, la paramétrisation complète d'une CMC gaussienne s'écrit $\boldsymbol{\theta} = (\boldsymbol{\theta}_X, \boldsymbol{\theta}_{Y|X})$, avec $\boldsymbol{\theta}_X = (\boldsymbol{p}, \boldsymbol{P})$ dans le cas standard, soit globalement $N^2 + 2N - 1$ degrés de liberté, et $\boldsymbol{\theta}_X = (\boldsymbol{\lambda}, \boldsymbol{\mu})$ dans le cas d'une chaîne télégraphique cachée, soit seulement $4N - 1$ degrés de liberté.

3.2. Apprentissage par maximum de vraisemblance

Notre étude est consacrée exclusivement au problème de l'apprentissage non supervisé, c'est-à-dire à l'estimation des paramètres $\boldsymbol{\theta}$ à partir de $\boldsymbol{y}$. Comme pour de nombreux problèmes d'estimation de paramètres de lois, c'est la recherche du maximum de la vraisemblance $f_{\mathbf{Y}}(\boldsymbol{y} \,; \boldsymbol{\theta})$ qui constitue l'outil privilégié, en raison de ses propriétés d'efficacité asymptotique. Son utilisation ne va pas pour autant sans difficultés, d'ordre à la fois théorique et pratique. NÁDAS [7] montre l'existence systématique de points singuliers de vraisemblance infinie. Dans le cas mieux connu des mélanges de populations gaussiennes, le même problème est étudié dans [10].

3.3. Algorithme EM pour les CMC

L'algorithme EM constitue une approche peu coûteuse permettant le calcul d'une estimée $\hat{\boldsymbol{\theta}}$ vraisemblable. Il s'agit d'une méthode à point fixe qui, partant d'une valeur initiale quelconque $\hat{\boldsymbol{\theta}}^0$, permet de générer une suite d'estimateurs $\{\hat{\boldsymbol{\theta}}^k\}$ de vraisemblance croissante. Chaque itération de l'algorithme est présenté classiquement en deux étapes :

Expectation (E)	Exprimer $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^k \,; \boldsymbol{y})$,
Maximization (M)	$\hat{\boldsymbol{\theta}}^{k+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^k \,; \boldsymbol{y})$,

où la fonction négative Q est définie par

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0 \,; \boldsymbol{y}) \triangleq \mathrm{E}\left(J(\boldsymbol{X}, \boldsymbol{Y} \,; \boldsymbol{\theta}) \mid \boldsymbol{Y} = \boldsymbol{y} \,; \boldsymbol{\theta}^0\right),$$

avec

$$J(\boldsymbol{x}, \boldsymbol{y} \,; \boldsymbol{\theta}) \triangleq \ln f_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y} \mid \boldsymbol{x} \,; \boldsymbol{\theta}_{Y|X}) P(\boldsymbol{X} = \boldsymbol{x} \,; \boldsymbol{\theta}_X)$$

la log-vraisemblance *a posteriori* pour $\boldsymbol{x}$. L'intérêt de l'algorithme EM provient de la relation aisément démontrable

$$\begin{aligned} \ln f_{\mathbf{Y}}(\boldsymbol{y} \,; \boldsymbol{\theta}) - \ln f_{\mathbf{Y}}(\boldsymbol{y} \,; \boldsymbol{\theta}^0) = \\ Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0 \,; \boldsymbol{y}) - Q(\boldsymbol{\theta}^0, \boldsymbol{\theta}^0 \,; \boldsymbol{y}) + D(\boldsymbol{\theta} \parallel \boldsymbol{\theta}^0), \end{aligned} \tag{6}$$

où $D(\boldsymbol{\theta} \parallel \boldsymbol{\theta}^0)$ est une pseudo-distance de Kullback-Leibler. Comme cette quantité est non négative, la non décroissance de la vraisemblance est assurée par maximisation de $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0 \,; \boldsymbol{y})$.

Il est facile de voir que chaque étape M de l'algorithme EM est constituée par la maximisation séparée de

$$\begin{aligned} Q_{Y|X}(\boldsymbol{\theta}_{Y|X}, \boldsymbol{\theta}^0 \,; \boldsymbol{y}) \triangleq \\ \mathrm{E}\left(\ln f_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{Y} \mid \boldsymbol{X} \,; \boldsymbol{\theta}_{Y|X}) \mid \boldsymbol{Y} = \boldsymbol{y} \,; \boldsymbol{\theta}^0\right) \end{aligned}$$

et de

$$Q_X(\boldsymbol{\theta}_X, \boldsymbol{\theta}^0\;; \boldsymbol{y}) \triangleq \mathrm{E}\left(\ln P(\boldsymbol{X}\;; \boldsymbol{\theta}_X) \mid \boldsymbol{Y} = \boldsymbol{y}\;; \boldsymbol{\theta}^0\right), \tag{7}$$

respectivement en $\boldsymbol{\theta}_{Y|X}$ et en $\boldsymbol{\theta}_X$, afin de constituer conjointement le vecteur $\boldsymbol{\theta}$ optimal. Dans le cas de la paramétrisation standard, ces équations de réestimation ont été proposées en 1970 [1] avant même que le principe général des méthodes EM ne soit reconnu. Le lecteur trouvera dans [9] et [3] une présentation illustrée (respectivement en reconnaissance de parole et en segmentation d'image) de ces équations désormais classiques, dites *formules de Baum-Welch.*

4. ALGORITHME EM TÉLÉGRAPHIQUE (EMT)

Le cas d'une CMC télégraphique diffère seulement du cas standard par le contenu de $\boldsymbol{\theta}_X$. En conséquence, la maximisation de $Q_{Y|X}$ est commune aux cas EM standard et EMT, c'est pourquoi seul le problème de la maximisation de Q_X est considéré dans la suite.

4.1. Étape E

En associant (7) et (5), on obtient

$$\begin{aligned} &Q_X(\boldsymbol{\theta}_X, \boldsymbol{\theta}^0\;; \boldsymbol{y}) \\ &\quad= \sum_{\boldsymbol{x}} P(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{Y} = \boldsymbol{y}\;; \boldsymbol{\theta}^0) \ln P(\boldsymbol{X} = \boldsymbol{x}\;; \boldsymbol{\theta}_X) \\ &\quad= \sum_n p_{1,n}^0 \ln p_n \sum_{m,n} \sum_{t=2}^{T} p_{t,mn}^0 \ln p_{mn}, \end{aligned}$$

en posant

$$\begin{aligned} p_{t,n} &\triangleq P(X_t = n \mid \boldsymbol{Y} = \boldsymbol{y}\;; \boldsymbol{\theta}), \\ p_{t,mn} &\triangleq P(X_{t-1} = m, X_t = n \mid \boldsymbol{Y} = \boldsymbol{y}\;; \boldsymbol{\theta}), \end{aligned}$$

et en repérant systématiquement par l'exposant ^{0} les quantités calculées pour $\boldsymbol{\theta}^0$ à l'aide de l'algorithme *forward-backward* classique [9]. Les expressions (1) et (2) permettent ensuite d'expliciter la dépendance de Q_X vis-à-vis des paramètres $\boldsymbol{\lambda}$ et $\boldsymbol{\mu}$:

$$\begin{aligned} Q_X(\boldsymbol{\theta}_X, \boldsymbol{\theta}^0\;; \boldsymbol{y}) &= \sum_n \alpha_n^0 \ln \mu_n + \beta_n^0 \ln(1-\lambda_n) \\ &+ s_n^0 \ln\left[1 + \frac{\lambda_n}{\mu_n(1-\lambda_n)}\right] - \ln \sum_n \frac{\mu_n}{1-\lambda_n}, \end{aligned} \tag{8}$$

avec

$$\alpha_n^0 \triangleq \sum_{t=1}^{T} p_{t,n}^0,\; \beta_n^0 \triangleq \sum_{t=2}^{T-1} p_{t,n}^0,\; s_n^0 \triangleq \sum_{t=2}^{T} p_{t,nn}^0. \tag{9}$$

4.2. Étape M

Le problème est maintenant de maximiser (8) en $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ sous les contraintes (3) et (4). Malheureusement, le dernier terme présent dans (8) rend cette maximisation inextricable. Pour résoudre ce problème, nous remplaçons Q_X par une nouvelle fonction peu différente :

$$\begin{aligned} R_X(\boldsymbol{\theta}_X, \boldsymbol{\theta}^0\;; \boldsymbol{y}) &\triangleq Q_X(\boldsymbol{\theta}_X, \boldsymbol{\theta}^0\;; \boldsymbol{y}) \\ &- \mathrm{E}\left(\ln P(X_1\;; \boldsymbol{\theta}_X) P(X_T\;; \boldsymbol{\theta}_X) \mid \boldsymbol{Y} = \boldsymbol{y}\;; \boldsymbol{\theta}^0\right)/2, \end{aligned}$$

à laquelle nous appliquons le principe d'optimisation itératif de l'EM. Ainsi, nous ne disposons plus de la condition suffisante (6) de croissance de la vraisemblance, mais il est possible de constater en pratique que l'écart entre Q_X et R_X est suffisamment faible pour que la vraisemblance demeure strictement croissante. En contrepartie, nous allons voir que la maximisation itérative de R_X se présente sous une forme particulièrement attrayante. Tout d'abord, explicitons R_X en fonction de $\boldsymbol{\lambda}$ et $\boldsymbol{\mu}$: $R_X(\boldsymbol{\theta}_X, \boldsymbol{\theta}^0\;; \boldsymbol{y}) = \sum_n R_n$, avec

$$R_n \triangleq \eta_n^0 \ln \mu_n(1-\lambda_n) + s_n^0 \ln\left[1 + \frac{\lambda_n}{\mu_n(1-\lambda_n)}\right], \tag{10}$$

et

$$\eta_n^0 \triangleq (\alpha_n^0 + \beta_n^0)/2. \tag{11}$$

Dans un premier temps, il est facile de maximiser R_X en $\boldsymbol{\lambda}$ en fonction de $\boldsymbol{\mu}$ en maximisant séparément chaque R_n :

$$\hat{\lambda}_n = \frac{s_n^0/\eta_n^0 - \mu_n}{1-\mu_n}. \tag{12}$$

Les contraintes d'inégalités (4) sont alors automatiquement satisfaites par $\hat{\boldsymbol{\lambda}}$, d'après (9) et (11). En reportant ensuite (12) dans (10), on obtient, à une constante près :

$$R_n = \gamma_n^0 \ln \frac{\mu_n}{1-\mu_n}, \text{ avec } \gamma_n^0 \triangleq \eta_n^0 - s_n^0 \geq 0. \tag{13}$$

La maximisation en $\boldsymbol{\mu}$ sous les contraintes (3) fait ensuite intervenir un paramètre de Lagrange ν, en fonction duquel on peut exprimer la condition d'annulation de la dérivée de $R_X + (\mathbf{1}^{\mathrm{t}}\boldsymbol{\mu})\nu$ pour chaque μ_n :

$$\nu \hat{\mu}_n^2 - \nu \hat{\mu}_n + \gamma_n^0 = 0. \tag{14}$$

Si $\nu > 4\gamma_n^0$, cette équation possède deux racines de part et d'autre de $1/2$, respectivement $\mu_n^-(\nu)$ et $\mu_n^+(\nu)$. Par combinaison, ce résultat autorise 2^N formes possibles pour $\hat{\boldsymbol{\mu}}(\nu)$, mais la prise en compte des contraintes (3) permet de réduire cette indétermination. Premièrement, elles rendent impossible le choix d'une racine de type « μ^+ » pour plus d'un seul état, ce qui laisse encore $N+1$ formes possibles. Deuxièmement, parmi les N formes incluant un « μ^+ », une analyse détaillée montre que la plus compétitive (au sens de la maximisation de R_X) est celle qui contient $\mu_{\bar{n}}^+(\nu)$, où $\bar{n}$ est l'état qui maximise γ_n^0, c'est-à-dire,

$$\bar{n} \triangleq \arg\max_n \gamma_n^0.$$

Il ne subsiste plus que deux formes possibles :

$$\forall n \neq \bar{n}, \mu_n = \mu_n^-\;; \mu_{\bar{n}} = \begin{cases} \mu_{\bar{n}}^- & : \text{cas « } \boldsymbol{\mu}^- \text{ »}, \\ \mu_{\bar{n}}^+ & : \text{cas « } \boldsymbol{\mu}^+ \text{ »}. \end{cases} \tag{15}$$

Parmi ces deux possibilités, un raisonnement analytique minutieux permet de montrer qu'une seule est compatible avec (3), suivant le résultat d'existence et d'unicité suivant :

Le maximum de R_X en $\boldsymbol{\mu}$ sous les contraintes (3) est atteint pour une unique valeur $\hat{\nu}$ de ν. Son argument prend la forme :

$$\hat{\mu} = \begin{cases} \mu^- & \text{si } \sum_n \sqrt{\gamma_{\bar{n}}^0 - \gamma_n^0} \leq (N-2)\gamma_{\bar{n}}^0, \\ \mu^+ & \text{sinon.} \end{cases} \qquad (16)$$

L'expression de $\hat{\nu}$ n'est pas explicite, mais on peut en obtenir un encadrement serré initial, puis raffiner le calcul par une méthode standard d'interpolation pour atteindre le respect de la contrainte $\mathbf{1}^t\hat{\boldsymbol{\mu}} = 1$ avec une précision arbitraire.

4.3. Algorithme

Tandis que l'obtention des équations de l'EMT s'avère plus délicate que dans le cas de l'EM standard, la mise en œuvre algorithmique n'est pas beaucoup plus compliquée. Dans les deux cas, l'étape la plus coûteuse reste le calcul des probabilités *a posteriori* $p_{t,n}^0$ et $p_{t,mn}^0$ par l'algorithme forward-backward.

Étant donné $\boldsymbol{\theta}^0 = \hat{\boldsymbol{\theta}}^k$, le calcul de $\hat{\boldsymbol{\theta}}_X^{k+1}$ se décompose sous la forme suivante :

- *Forward-backward* [9], [3] $\longrightarrow \{p_{t,n}^0\}, \{p_{t,nn}^0\}$;
- (9), (11), (13) $\longrightarrow \{\gamma_n^0\}$;
- (15) et interpolation $\longrightarrow \hat{\nu}$;
- (16) $\longrightarrow \hat{\boldsymbol{\mu}}$;
- (12) $\longrightarrow \hat{\boldsymbol{\lambda}}$,

qui constitue l'étape originale d'un algorithme EMT par rapport au cas standard des formules de Baum-Welch.

4.4. Discussion

En injectant (12) et (14) dans (1) et (2), on peut remarquer que l'algorithme EMT proposé fournit des formules de réestimation dont l'essentiel est très intuitif, à la façon des formules de Baum-Welch. En effet, les remises à jour de $\boldsymbol{p}$ et de la diagonale de $\boldsymbol{P}$ s'écrivent respectivement $\hat{p}_n = \eta_n^0/(T-1)$ et $\hat{P}_{nn} = s_n^0/\eta_n^0$, obtenues par moyennages empiriques des probabilités *a posteriori* $p_{t,n}^0$ et $p_{t,nn}^0$. En revanche les coefficients non diagonaux de $\hat{\boldsymbol{P}}$ ont des expressions plus complexes faisant intervenir le calcul du paramètre de Lagrange $\hat{\nu}$, conséquences des contraintes qui portent sur le modèle.

Du calcul des coefficients non diagonaux résulte une des principales qualités de l'algorithme EMT. En effet, ce calcul apporte une solution originale au problème de l'apprentissage par un faible nombre de données, souligné par Rabiner [9]. Typiquement, si $T < N^2 + 1$, les formules de réestimation de Baum-Welch fournissent une matrice de transition creuse, c'est-à-dire que les transitions absentes dans la séquence d'apprentissage seront interdites lors des utilisations ultérieures des paramètres estimés. Au contraire, l'identification d'un modèle télégraphique minimise ce risque dès lors que $T > N$.

5. CONCLUSION

PERROT *et coll.* [8] présentent un exemple réaliste d'utilisation du nouvel algorithme EMT pour la segmentation non-supervisée d'images médicales. Cet exemple montre que la paramétrisation télégraphique permet d'accroître la robustesse de la phase d'apprentissage EM, au sens où les cas de divergence de l'EMT sont plus rares que les divergences de l'EM. Cependant, dans la mesure où la réestimation des paramètres $\boldsymbol{\theta}_{Y|X}$, source de dégénérescence de la vraisemblance [7], n'est pas modifiée par la variante EMT, aucune garantie formelle de convergence ne peut être apportée. La recherche de solutions bayésiennes à ce problème, plus satisfaisantes que les solutions contraintes, figure parmi les perspectives de ce travail.

6. RÉFÉRENCES

[1] L. BAUM, T. PETRIE, G. SOULES & N. WEISS, « A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains », *Annals of Mathematical Statistics*, **41**, N°1, pp. 164–171, 1970.

[2] A. BLAKE & A. ZISSERMAN, *Visual reconstruction*. The MIT Press, Cambridge, 1987.

[3] P. DEVIJVER & M. DEKESSEL, « Champs aléatoires de Pickard et modélisation d'images digitales », *Traitement du Signal*, **5**, N°5, pp. 131–150, 1988.

[4] S. GEMAN & D. GEMAN, « Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**, N°6, pp. 721–741, novembre 1984.

[5] R. GODFREY, F. MUIR & F. ROCCA, « Modeling seismic impedance with Markov chains », *Geophysics*, **45**, N°9, pp. 1351–1372, septembre 1980.

[6] F. JENG & J. WOODS, « Compound Gauss-Markov random fields for image estimation », *IEEE Transactions on Signal Processing*, **SP-39**, N°3, pp. 683–697, mars 1991.

[7] A. NÁDAS, « Hidden Markov chains, the forward-backward algorithm, and initial statistics », *IEEE Transactions on Acoustics Speech and Signal Processing*, **ASSP-31**, N°2, pp. 504–506, avril 1983.

[8] G. PERROT, Y. GOUSSARD & J. IDIER, « Paramétrisation « télégraphique » de champs de Pickard pour la segmentation non-supervisée d'images », *Actes du GRETSI*, septembre 1995.

[9] R. RABINER, « A tutorial on hidden Markov models and selected applications in speech recognition », *Proceedings of the IEEE*, **77**, N°2, pp. 257–286, février 1989.

[10] R. REDNER & H. WALKER, « Mixture densities, maximum likelihood and the EM algorithm », *SIAM Rev.*, **26**, N°2, pp. 195–239, avril 1984.