CONVERGENCE OF THE ITERATIVE CONDITIONAL ESTIMATION AND APPLICATION TO MIXTURE PROPORTION IDENTIFICATION

Wojciech Pieczynski

INT/GET, Département CITI, CNRS UMR 5157
9, rue Charles Fourier, 91000 Evry, France
tél. 01 60 76 44 25, fax 01 60 76 44 33, e-mail <u>Wojciech.Pieczynski@int-evry.fr</u>

ABSTRACT

The iterative conditional estimation (ICE) is an iterative estimation method of the parameters in the case of incomplete data. Proposed since about fifteen years, ICE works under weak hypotheses and has been successfully applied in many unsupervised processing problems. In particular, it gave good results in unsupervised image segmentation based on complex models like hidden fuzzy Markov fields, hidden evidential Markov fields, or triplet Markov fields. However, there were no general theoretical results concerning its asymptotic behavior until now. The aim of this paper is to provide a general theorem, and to specify two applications: the mixture proportion estimation in a very general setting, and estimation of the components means in Gaussian mixture. The position of ICE with respect to the "Expectation-Maximization" (EM) method is also briefly discussed.

Index Terms—Iterative conditional estimation, incomplete data, mixture estimation

1. INTRODUCTION

The iterative conditional estimation (ICE) is an iterative estimation method of parameters in the case of incomplete data. Its use asks for very weak hypotheses and it can be performed in relatively complex situations. Proposed in [13], ICE has been successfully applied in many problems of statistical signal or image segmentation; let us mention [3, 5, 7, 8, 11, 15] among recent papers. Before, ICE also gave good results in different complex models, like hidden fuzzy (also said "mixed-states") Markov fields [16], hidden evidential Markov fields [2], multi-sensor Markov chains [10], or triplet Markov fields [1]. However, there were no theoretical results concerning the asymptotic behaviour of ICE until now. The aim of this paper is to provide a general theorem, and to show that it is applicable to the mixture proportion estimation in a very general setting.

2. CONVERGENCE OF ICE

Let $Z = (X, Y)$ be a couple of random variables whose distribution $p(z|\theta)$ depends on a parameter $\theta = (\theta_1, ..., \theta_m) \in R^m$. The problem of estimating θ from Y is a very important one for it can be requested as preliminary step in search of X from Y. Let $\hat{\theta}(z)$ be an estimator of θ defined from the complete data $z = (x, y)$. ICE is an iterative method producing a sequence (θ^q) in the following way:

(i) initialize θ^0;

(ii) compute $\tilde{\theta}_i^{q+1}(y) = E_{\theta^q}[\hat{\theta}_i(X, Y)|Y = y]$ for the components θ_i of the vector θ for which this computation is workable;

(iii) for other components θ_i, use the classical approximation simulating x_1^q, …, x_l^q according to $p(x|y, \theta^q)$ and setting $\tilde{\theta}_i^{q+1}(y) = [\hat{\theta}(x_1^q, y) + ... + \hat{\theta}(x_l^q, y)]/l$.

We see that ICE is applicable under very slight two hypotheses: existence of an estimator $\hat{\theta}(z)$ from the complete data $z = (x, y)$, and the ability of simulating X according to $p(x|y)$. The first hypothesis is not really a constraint because if we are not able to estimate θ from complete data $z = (x, y)$, there is no point in searching an estimator from incomplete ones y. The second hypothesis is often verified, because in many models $p(x|y)$ has to be simple enough to allow one searching $X = x$ from $Y = y$.

We will state an ICE convergence result in the following context. Let $(X \times Y, T, P_\theta)_{\theta \in \Theta}$ be a statistical model, with $X = X_1 \times X_2 \times ... \times X_n \times ...$, $Y = Y_1 \times Y_2 \times ... \times Y_n \times ...$, (the sets X_i are all equal and it is the same for the sets Y_i), and $\Theta \subset R^m$. We will denote by $X^n = (X_1, X_2, ..., X_n)$ and

$Y^n = (Y_1, Y_2, ..., Y_n)$ the corresponding random variables, and we will assume the random variables $Z_i = (X_i, Y_i)$ independent and identically distributed (i.i.d.). Otherwise, $B(\theta, \rho)$ will designate the open ball of centre θ and of radius ρ.

We can state the following result:

Theorem

Let $\hat{\theta}_1 = \hat{\theta}(X_1, Y_1)$ be an unbiased estimator with finite variance defined from (X_1, Y_1), and let $\hat{\theta}_n = \hat{\theta}_n(X^n, Y^n) = = [\hat{\theta}(X_1, Y_1) + ... + \hat{\theta}(X_n, Y_n)]/n$ be the corresponding unbiased estimator converging in L^2. For each θ', θ'' in Θ, we will set $\varphi(\theta', \theta'') = E_{\theta'}[E_{\theta''}[\hat{\theta}_1 | Y_1]]$.

Let $\theta^0 \in \Theta$. Under the following hypotheses:

(i) For every $\theta \in \Theta$, the deterministic sequence $(\theta^q)_{q \in N}$ recursively defined by θ^0 and $\theta^{q+1} = \varphi(\theta, \theta^q)$ converges to θ, and the convergence is uniform with respect to θ ;

(ii) for every $q \in N^* = \{1, 2, ..., n, ...\}$, the function $\theta' \to \varphi(\theta, \theta')$ is continue at θ^q, and its continuity is uniform with respect to θ ;

(iii) there exists a matrix M such that $Var_\theta[E_{\theta'}[\hat{\theta}_1 | Y_1]] \leq M$ for each θ, θ' in Θ,

there exists a sequence of natural numbers $(q(n))_{n \in N}$ tending to infinity such that the ICE sequence $[\tilde{\theta}_n^{q(n)}(Y^n)]_{n \in N}$ defined with θ^0 converges in probability.

Proof. Let $\theta, \theta^0 \in \Theta$ and let $(\theta^q)_{q \in N}$ be the deterministic sequence defined in (i). Let us show the following property (iv), which will allow us to find $(q(n))_{n \in N}$ such that $(\tilde{\theta}_n^{q(n)})_{n \in N}$ converges.

(iv) For each natural number $m \geq 1$ there exist two natural numbers $k(m)$ and $n(m)$ such that for $n \geq n(m)$, $P_\theta[\|\tilde{\theta}_n^{k(m)} - \theta\| \leq \frac{1}{m}] \geq 1 - \frac{1}{m}$ for each $\theta \in \Theta$.

Let $m \geq 1$. According to (i), there exists $k(m) \in N$, which will be taken minimal, such that $\|\theta^q - \theta\| < \frac{1}{2m}$ for $q \geq k(m)$ and each $\theta \in \Theta$. This implies that if

$\theta' \in B(\theta^{k(m)}, \frac{1}{2m})$, then $d(\theta, \theta') \leq \frac{1}{m}$. Let us consider the sequence of balls $B(\theta^1, \rho_1)$, $B(\theta^2, \rho_2)$, ..., $B(\theta^{k(m)}, \rho_{k(m)})$ (depending on θ) whose radius are defined - this is possible by virtue of (ii) – by backward recursions $\rho_{k(m)} = \frac{1}{2m}$, $\rho_{k(m)-1}$ such that $\theta' \in B(\theta^{k(m)-1}, \rho_{k(m)-1})$ implies $\varphi(\theta, \theta') \in B(\theta^{k(m)}, \frac{\rho_{k(m)}}{2})$ (for each $\theta \in \Theta$), ..., ρ_{q-1} such that $\theta' \in B(\theta^{q-1}, \rho_{q-1})$ implies $\varphi(\theta, \theta') \in B(\theta^q, \frac{\rho_q}{2})$ (for each $\theta \in \Theta$), ..., ρ_1 such that $\theta' \in B(\theta^1, \rho_1)$ implies $\varphi(\theta, \theta') \in B(\theta^2, \frac{\rho_2}{2})$ (for each $\theta \in \Theta$). It is then possible, by virtue of Bienaymé-Tchebychev inequality and (iii), to find n_1, n_2, ..., $n_{k(m)}$ such that for each $q = 1$, ..., $k(m)$, when $\theta' \in B(\theta^{q-1}, \rho_{q-1})$ and when $n_{q'} \geq n_q$, the probability that $E_{\theta'}[\hat{\theta}_{n_{q'}}(X^{n_{q'}}, Y^{n_{q'}}) | Y^{n_{q'}}]$ is in $B(\theta^q, \rho_q)$ is superior to $1 - \frac{1}{m2^{q+1}}$. Let us consider $n(m)$ the sup of n_1, n_2, ..., $n_{k(m)}$ and, for $q = 1$, ..., $k(m)$, let $U_1^{n(m)}$, ..., $U_{k(m)}^{n(m)}$ be the random variables defined by $U_q^{n(m)} = 1$ if $\tilde{\theta}_{n(m)}^q(Y^{n(m)}) \in B(\theta^q, \rho_q)$, and $U_q^{n(m)} = 0$ otherwise. Then $P[U_1^{n(m)} = 1] \geq 1 - \frac{1}{m2^2}$ and $P[U_q^{n(m)} = 1 | U_{q-1}^{n(m)} = 1] \geq 1 - \frac{1}{m2^{q+1}}$ for $q = 2$, ..., $k(m)$. The sequence $U_1^{n(m)}$, ..., $U_{k(m)}^{n(m)}$ being Markovian, this implies $P[U_1^{n(m)} = 1, U_2^{n(m)} = 1, ... U_{k(m)}^{n(m)} = 1] \geq (1 - \frac{1}{m2})(1 - \frac{1}{m2^2})...(1 - \frac{1}{m2^{k(m)+1}}) \geq 1 - \frac{1}{m}$. Finally, $P[U_{k(m)}^{n(m)} = 1] \geq 1 - \frac{1}{m}$, and it is still true replacing $n(m)$ by any $n \geq n(m)$, which gives (iv).

Let us consider the following sequence $(q(n))_{n \in N}$ of natural numbers. For $m = 1$, 2, ... let $n(1)$, $n(2)$, ... and $k(1)$, $k(2)$, ... be two sequences existing by virtue of (iv). Then we set $q(n) = k(1)$ for $1 \leq n \leq n(1)$, $q(n) = k(2)$ for $n(1) < n \leq n(2)$, ..., $q(n) = k(m)$ for $n(m) < n \leq n(m+1)$, ... and so on. Let us show that $\tilde{\theta}_n^{q(n)}$ converges in probability. Let $\varepsilon > 0$. By virtue of (iv), there exist $k(\varepsilon)$ and $n(\varepsilon)$ such that for $n \geq n(\varepsilon)$, $P[\|\tilde{\theta}_n^{k(\varepsilon)} - \theta\| \leq \varepsilon] \geq 1 - \varepsilon$ (we take the smallest k among all k verifying $\frac{1}{k} < \varepsilon$).

Knowing that $P[\|\tilde{\theta}_n^{k(m+m')} - \theta\| \leq \frac{1}{m+m'}] \geq 1 - \frac{1}{m+m'}$, for

each n verifying $n(m+m') < n \le n(m+m'+1)$, we can state that if $n \ge n(\varepsilon)$ and $n(m+m') < n \le n(m+m'+1)$, then $P[\|\tilde{\theta}_n^{q(n)} - \theta\| \le \frac{1}{m+m'}] \ge 1 - \frac{1}{m+m'}$. This implies that for each $n \ge n(\varepsilon)$ we have $P[\|\tilde{\theta}_n^{q(n)} - \theta\| \le \varepsilon] \ge 1 - \varepsilon$, which ends the proof.

Let us remark that (ii) and (iii) are rather subsidiary hypotheses and are generally easy to verify. In particular, the uniformity with respect to θ can be obtained assuming that $\Theta \subset R^m$ is compact, which can often be made in real situations. Therefore the main result is the relationship between the convergence of the deterministic sequence $(\theta^q)_{q \in N}$ defined in (i) and the convergence of the stochastic sequence $[\tilde{\theta}_n^{q(n)}(Y^n)]_{n \in N}$ defined by the ICE principle. We give in the two next sections two examples showing that the convergence of $(\theta^q)_{q \in N}$ can be established in two cases linked with the important problem of finite mixture identification.

3. MIXTURE PROPORTION ESTIMATION

Let $X_1 = \{\omega_1, \omega_2\}$, $Y_1 = R^m$, and $f(y_i) = p(y_i|x_i = \omega_1)$, $g(y_i) = p(y_i|x_i = \omega_2)$ two known probability densities with respect to some measure dy on R^m. We assume that $\theta = p(x_1 = \omega_1)$ in unknown and the problem is to estimate it from $Y^n = (Y_1, Y_2, ..., Y_n)$. In order to apply ICE, we use the classical efficient estimator from complete data $\hat{\theta}_n = \hat{\theta}(X_n, Y_n) = \hat{\theta}(X_n) = [1_{[X_1 = \omega_1]} + ... + 1_{[X_n = \omega_1]}]/n$. Then we obtain for ICE iterations $\tilde{\theta}_n^{q+1}(y^n) = [p_{\theta^q}(X_1 = \omega_1|y_1) + ... + p_{\theta^q}(X_n = \omega_1|y_n)]/n$. We can state

Proposition 1

The hypotheses of the theorem above are verified for each $\theta^0 \in \Theta$.

Proof. The deterministic sequence (θ^q) is given by $\theta^{q+1} = \int_{R^m} \frac{\theta^q f[\theta^q f + (1-\theta)g]}{[\theta^q f + (1-\theta^q)g]} dy$. Setting $\int_{R^m} \frac{\theta^q f(f-g)}{[\theta^q f + (1-\theta^q)g]} dy = h(\theta^q)$, this can be written $\theta^{q+1} = \theta^q + (\theta - \theta^q)h(\theta^q)$, which gives:

$$\theta^{q+1} - \theta = (\theta^q - \theta)(1 - h(\theta^q)) \qquad (1)$$

Let us show that for each $\varepsilon \in]0, 0.5[$, if $\theta^0 \in [\varepsilon, 1-\varepsilon]$, then there exists $\alpha \in]0, 1[$ such that $|\theta^{q+1} - \theta| \le (1-\alpha)^{q+1}$. Let $\varepsilon \in]0, 0.5[$. We verify that $h(0) = h(1) = 0$ and, otherwise, the second derivates of h is $h''(t) = \int_{R^m} \frac{-2fg(f-g)^2}{[tf + (1-t)g]^3} dy$, which is negative on $[0,1]$. This shows that h strictly positive on $]0,1[$ and thus, as h is continuous, there exists $\alpha^1 \in]0,1[$ such that $h(t) \ge \alpha^1$ for $t \in]\varepsilon, 1-\varepsilon[$. Otherwise, setting $p_{1,\theta^q}(\omega_1|y) = \frac{\theta^q f(y)}{\theta^q f(y) + (1-\theta^q)g(y)}$, $p_{\theta^q}^{(1)}(\omega_1) = \int_{R^m} p_{1,\theta^q}(\omega_1|y)f(y)dy$, and $p_{\theta^q}^{(2)}(\omega_1) = \int_{R^m} p_{1,\theta^q}(\omega_1|y)g(y)dy$, we see that $p_{\theta^q}^{(1)}(\omega_1)$ and $p_{\theta^q}^{(2)}(\omega_1)$ are probabilities, and thus they are both in $]0,1[$. As $h(\theta^q) = p_{\theta^q}^{(1)}(\omega_1) - p_{\theta^q}^{(2)}(\omega_1)$, there exists $\alpha^2 \in]0,1[$ such that $h(t) \le 1 - \alpha^2$ for $t \in]\varepsilon, 1-\varepsilon[$. Finally, we can state that there exists $\alpha = \sup(\alpha^1, \alpha^2) \in]0,1[$ such that for each $\theta^q \in [\varepsilon, 1-\varepsilon]$, we have $\alpha \le h(\theta^q) \le 1 - \alpha$, which also means that $\alpha \le 1 - h(\theta^q) \le 1 - \alpha$. Given (1), we can say that if $\theta^0 \in [\varepsilon, 1-\varepsilon]$, then $\theta^{q+1} - \theta = (\theta^0 - \theta)(1-\alpha)^{q+1}$. Finally, given that $|\theta^0 - \theta| < 1$, we have $|\theta^{q+1} - \theta| \le (1-\alpha)^{q+1}$, which implies the convergence, uniform with respect to θ, of the sequence (θ^{q+1}) to θ.

The hypothesis (ii) is verified coarsely, and hypothesis (iii) comes from the continuity of the function

$$(\theta, \theta') \to Var_{\theta'}[E_{\theta'}[\hat{\theta}_1|Y_1]] = Var_{\theta'}[\frac{\theta'f(Y_1)}{\theta'f(Y_1)+(1-\theta')g(Y_1)}] =$$

$$\int_{R^m}[\frac{\theta'f(Y_1)}{\theta'f(Y_1)+(1-\theta')g(Y_1)}]^2[\theta f(Y_1)+(1-\theta)g(Y_1)]dy -$$

$$[\int_{R^m}[\frac{\theta'f(Y_1)\theta f(Y_1)+(1-\theta)g(Y_1)}{\theta'f(Y_1)+(1-\theta')g(Y_1)}]dy]^2, \text{ defined on the compact}$$

set $[0,1]^2$.

4. COMPONENTS MEANS OF A MIXTURE ESTIMATION

Let us consider the problem of estimating the means of the components in Gaussian mixture

$$p(y) = \sum_{i=1}^{k} \alpha_i f_i(y) \qquad (2)$$

Thus we assume $\alpha_1, ..., \alpha_k$ known, and $f_1, ..., f_k$ Gaussian with unknown means $m_1, ..., m_k$ and known variances $\sigma_1^2, ..., \sigma_k^2$. The problem is to estimate

$m = (m_1,...,m_k)$ from an i.i.d. sample $Y^n = (Y_1,...,Y_n)$. Let us introduce the parameter $\theta = (\theta_1,...,\theta_k) = (\alpha_1 m_1,...,\alpha_k m_k)$ and let us consider the estimator from complete data

$$\hat{\theta}_i(x_1,...,x_n, y_1,...,y_n) = \frac{\delta_i(x_1)y_1 + ... + \delta_i(x_n)y_n}{n} \qquad (3)$$

where δ_j is the function defined on $\Omega = \{\omega_1,...,\omega_k\}$ by $\delta_j(x) = 1_{[x=\omega_j]}$. Then we can classically show that $\hat{\theta}(X^n, Y^n)$ is unbiased and converges in L^2. Applying the ICE principle to the estimator defined by (3), we obtain, for $i = 1,...,k$;

$$\theta_i^{q+1}(y^n) = \frac{y_1 p(x_1 = \omega_i | y, \theta^q) + ... + y_n p(x_n = \omega_i | y, \theta^q)}{n} \qquad (4)$$

Using the theorem above, we will show the following result :

Proposition 2

Let us consider the mixture model described above. Let us assume that $m_i^3 < 6\sigma_i^3 \sqrt{2\pi}$ for $i = 1$, …, $i = k$. Then for each $\theta^0 \in \Theta$, there exists a likelihood V_{θ^0} of θ^0 such that for θ in V_{θ^0} the deterministic sequence (θ^q) defined by $\theta^{q+1} = E_\theta[E[\hat{\theta}_1(X_1,Y_1)|Y_1,\theta^q]] = \varphi(\theta,\theta^q)$ converges to θ, and the convergence is uniform with respect to θ.

Proof. Let $\theta^0 \in \Theta$. It is sufficient to show that there exists a constant $c > 0$ and a likelihood V_{θ^0} of θ^0 such that $\|\theta^{q+1} - \theta\| \le c\|\theta^q - \theta\|$ for each θ in V_{θ^0}; in fact, the latter implies $\|\theta^{q+1} - \theta\| \le c^{q+1}\|\theta^0 - \theta\|$ which establishes the convergence uniform qith respect to θ. Using the classical continuity properties it is possible to show that the property above is be verified if $\varphi(\theta,\theta^q)$ admits continuous partial derivatives with respect to all components of θ^q, and if all these partial derivatives are in [0,1[at the point $\theta^q = \theta$. Thus let us consider the partial derivatives of $\varphi(\theta,\theta^q)$ with respect to θ_i^q, with $i \in \{1,...,k\}$. We can take, as an example, $i = 1$. Thus we consider θ and θ^q, in which all components are equal except the first one. Recalling that $f_1(y|\theta_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp[-(y - \frac{\theta_1}{\alpha_1})^2 / 2\sigma_1^2]$, we have

$$\frac{\partial f_1(y|\theta_1)}{\partial \theta_1} = \frac{1}{\sigma_1^2 \alpha_1}(\frac{\theta_1}{\alpha_1} - y)f_1(y|\theta_1) \qquad (5)$$

Otherwise, to simplify the notations let us put $f(y|\theta) = \sum_{1 \le i \le k} \alpha_i f_i(y|\theta_i)$ and $g(y|\theta) = \sum_{2 \le i \le k} \alpha_i f_i(y|\theta_i)$. Knowing that $p(x = \omega_1 | y, \theta^q) = \frac{\alpha_1 f_1(y|\theta_1^q)}{f(y|\theta^q)}$, $\theta_1^{q+1} = E_\theta[E[\hat{\theta}_1|Y_1,\theta = \theta^q]]$ is given by

$$\theta_1^{q+1} = \int_R y \frac{\alpha_1 f_1(y|\theta_1^q)}{f_1(y|\theta_1^q) + g(y|\theta^q)} f(y|\theta)dy \qquad (6)$$

Thus we have to derive (6) with respect to θ_1^q, and take the value of the derivative at θ_1. Recalling that $g(y|\theta)$ does not depend on θ_1 and taking (4) into account we have

$$\frac{\partial \varphi(\theta,\theta^q)}{\partial \theta_1^q}(\theta,\theta) = \int_R y \frac{\alpha_1 \frac{1}{\sigma_1^2 \alpha_1}(\frac{\theta_1}{\alpha_1} - y)f_1(y|\theta_1)g(y|\theta)}{[\alpha_1 f_1(y|\theta_1) + g(y|\theta)]^2} f(y|\theta)dy$$

$$= \frac{1}{\sigma_1^2} \int_{R^m} \frac{y(m_1 - y)f_1(y|\theta)g(y|\theta)}{\alpha_1 f_1(y|\theta_1) + g(y|\theta)} dy \le \frac{1}{\sigma_1^2} \int_0^{m_1} \frac{y(m_1 - y)f_1(y|\theta)g(y|\theta)}{\alpha_1 f_1(y|\theta_1) + g(y|\theta)} dy$$

the last inequality being due to the fact that $y(m_1 - y)$ is positive on $[0,m_1]$ and negative outside.

Otherwise, as $\frac{1}{\alpha_1 f_1(y|\theta_1) + g(y|\theta)} \le \frac{1}{g(y|\theta)}$, the last integral is inferior to $\frac{1}{\sigma_1^2} \int_0^{m_1} y(m_1 - y)f_1(y|\theta)dy$. Finally, as $f_1(y|\theta_i) \le \frac{1}{\sqrt{2\pi\sigma_1^2}}$,

we have $\frac{\partial \varphi(\theta,\theta^q)}{\partial \theta_1^q} \le \frac{1}{\sigma_1^3 \sqrt{2\pi}} \int_0^{m_1} y(m_1 - y)dy = \frac{m_1^3}{6\sigma_1^3 \sqrt{2\pi}}$, which ends the proof.

Let us notice that on the contrary to the example in the previous section, the theorem is used "locally" in the sense that the true parameter θ is assumed to be close enough to the θ^q. Of course, as θ is not known, such kind of hypotheses is difficult to verify in real situations.

5. ICE AND EM

The ICE method resembles to the well known iterative EM method [12, 17, 18], whose principle is

$$\tilde{\theta}^{q+1}(y) = \arg\max_\theta E_{\theta^q}[Log(p_\theta(X,Y)|Y = y] \qquad (7)$$

and whose wide use is justified by the good asymptotic behaviour of the Maximum Likelihood estimator. The main result is that in the EM sequence $(\tilde{\theta}^q(y))_{q \in N}$ the likelihood $p_{\theta^q}(y)$ increases; however, the convergence of $(\tilde{\theta}^q(y))_{q \in N}$

to $\arg\max_{\theta}[p_{\theta}(y)]$ is not guaranteed [4, 12], and the main theoretical result consists on the convergence of the sequence $(p_{\theta^q}(y))_{q\in N}$ [18]. The EM and ICE obtained sequences are different in general, and, as ICE does not use any "likelihood maximization", it is in general easier to implement. In fact, the "maximization" step in EM can pose calculation problems, especially in complex models. When its exact computation is impossible, one can consider different approximation techniques, often of stochastic nature. However, these different approximation methods are, in general, difficult to study from the theoretical viewpoint and it is difficult to show that the good asymptotic properties of the Maximum Likelihood estimator are saved [4, 12, 17, 18].

Let us mention that some relationships between EM and ICE have been studied in the frame of exponential models in [6], and it has been showed that in some situation ICE can give the same sequence as EM. In particular, when the principle of the EM leads to explicit calculations in the mixture proportion estimation above, the obtained formulas are strictly the same as the formulas obtained with ICE. Therefore, in this case ICE can be seen as a generalisation of EM. Otherwise, the ICE iterations in the example in section 4 are strictly the same that those obtained with EM; thus this example can also be seen as an original way of studying the EM convergence.

6. CONCLUSION

We considered the problem of the Iterative Conditional Estimation (ICE) convergence in the case of independent and identically distributed data. A general theorem has been established and its interest has been validated through two classical examples concerning mixture estimation. Some advantages of ICE over EM have also been briefly mentioned.

7. REFERENCES

[1] D. Benboudjema and W. Pieczynski, Unsupervised image segmentation using triplet Markov fields, *Computer Vision and Image Understanding*, Vol. 99, No. 3, pp. 476-498, 2005.

[2] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski, Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 39, No. 8, pp. 1789-1798, 2001.

[3] Y. F. Cao, H. Sun, and X. Xu, An unsupervised segmentation method based on MPM for SAR images, *IEEE Geoscience and Remote Sensing Letters*, Vol. 2, No. 1, pp. 55-58, 2005.

[4] O. Cappé, E. Moulines, and T. Ryden, *Inference in hidden Markov models*, Springer, Series in Statistics, 2005.

[5] C. Carincotte, S. Derrode, and S. Bourennane, Unsupervised change detection on SAR images using fuzzy hidden Markov chains, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 44, No. 2, pp. 432-441, 2006.

[6] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, *IEEE Trans. on Signal Processing*, Vol. 45, No. 10, pp. 2613-2615, 1997.

[7] S. Derrode and G. Mercier, Unsupervised multiscale oil slick segmentation from SAR images, *Pattern Recognition*, Vol. 40, No. 3, pp. 1135-1147, 2007.

[8] F. Destrempes and M. Mignotte, A statistical model for contours in images, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 5, pp. 626-638, 2004.

[9] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, Unsupervised segmentation of radar images using hidden Markov chains and hidden Markov random fields, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 41, No. 3, pp. 675-686, 2003.

[10] N. Giordana and W. Pieczynski, Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 465-475, 1997.

[11] J. Lapuyade-Lahorgue and W. Pieczynski, Partially Markov models and unsupervised segmentation of semi-Markov chains hidden with long dependence noise, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2007), Chania, Greece, May 2007.

[12] G. J. McLachlan and T. Krishnan, *The EM algorithm and extension*, Wiley, 1997.

[13] W. Pieczynski, Statistical image segmentation, *Machine Graphics and Vision*, Vol. 1, No. 1/2, pp. 261-268, 1992.

[14] D. Pierre-Loti-Viaud, Random perturbations of recursive sequences with an application to an epidemic model, *Journal of Applied Probability*, Vol. 32, pp. 559-578, 1995.

[15] J.-N. Provost, C. Collet, P. Rostaing, P. Pérez, and P. Bouthemy, Hierarchical Markovian segmentation of multispectral images for reconstruction of water depth maps, *Computer Vision and Image Understanding*, Vol. 93, No. 2, pp. 155-174, 2004.

[16] F. Salzenstein and W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, *Graphical Models and Image Processing*, Vol. 59, No. 4, pp. 205-220, 1997.

[17] F. Vaida, Parameter convergence for EM and MM algorithms, *Statistica Sinica*, Vol. 15, pp. 831-840, 2005.

[18] C. F. J. Wu, On the convergence properties of the EM algorithm, Annals of Statistics, Vol. 11, pp. 95-103, 1983.