Expectation Maximization pour les chaînes de Markov cachées discrètes

Soit $(X_k)_{0\leq k\leq n}$ une chaîne de Markov discrète à valeurs dans $\{1,\ldots,r\}$, de matrice de transition Q et de loi initiale ν. On considère que cette chaîne est uniquement observée au travers des variables $(Y_k)_{0\leq k\leq n}$, indépendantes conditionnellement à $(X_k)_{0\leq k\leq n}$ et telles que pour tout $0\leq \ell\leq n$, la loi de Y_ℓ sachant $(X_k)_{0\leq k\leq n}$ est une gaussienne de moyenne μ_{X_ℓ} et de variance v_{X_ℓ}. Le paramètre inconnu est ici $\theta=\{\mu_1,\ldots,\mu_r,v_1,\ldots,v_r\}$.

1. Écrire la logvraisemblance jointe de $(X_{0:n},Y_{0:n})$: $\theta\mapsto \log p_\theta(X_{0:n},Y_{0:n})$.
2. Écrire la quantité intermédiaire de l'EM $Q(\theta,\theta')$ pour tout θ, θ' :
$$Q(\theta,\theta')=\mathbb{E}_{\theta'}\left[\log p_\theta(X_{0:n},Y_{0:n})|Y_{0:n}\right]\,.$$
3. Écrire cette quantité en faisant apparaître les probabilités
$$\omega^\theta_{k-1,k}(i,j)=\mathbb{P}_\theta\left(X_{k-1}=i,X_k=j|Y_{0:n}\right)\,,$$
pour $0\leq k\leq n$.
4. À l'itération $p\geq 0$, on dispose de l'estimation $\hat\theta^{(p)}$. Écrire l'estimateur $\hat\theta^{(p+1)}$ en maximisant $\theta\mapsto Q(\theta,\hat\theta^p)$, donner la valeur de $\theta^{(p+1)}$ en fonction des $\omega^{\theta^{(p)}}_{k-1,k}(i,j)$, $1\leq k\leq n$, $1\leq i,j\leq r$.
5. Dans le cas où on souhaite également apprendre la loi initiale de la chaîne de Markov, i.e. si l'on note $\nu_i=\mathbb{P}(X_0=i)$, $1\leq i\leq r$ et que $\theta=\{\mu_1,\ldots,\mu_r,v_1,\ldots,v_r,\nu_1,\ldots,\nu_r\}$, donner les équations de mise à jour de ν.
6. En admettant que l'on peut intervetir les calculs de gradients et les espérances, calculer le gradient de la logvraisemblance des observations : $\theta\mapsto\nabla_\theta\log p_\theta(Y_{0:n})$.
7. En déduire un algorithme de mise à jour des paramètres de type "descente de gradient".
8. **Bonus:** Calcul des $\omega^\theta_{k-1,k}(i,j)$, $1\leq k\leq n$, $1\leq i,j\leq r$.
   (a) Montrer que l'on peut calculer récursivement $\mathbb{P}_\theta\left(X_k=i|Y_{0:k}\right)$, $0\leq k\leq n$, $1\leq i,j\leq r$.
   (b) Montrer que l'on peut calculer récursivement, de $k=n$ à $k=0$, $\mathbb{P}_\theta\left(X_k=i|Y_{0:n}\right)$, $0\leq k\leq n$, $1\leq i\leq r$.
   (c) Conclure.