

Calcul de vraisemblance pour l'identification d'un modèle biomathématique continu-discret d'EEG en épilepsie

Paul FROGERAIS^{1,2}, Jean-Jacques BELLANGER^{1,2}

¹INSERM, U642, F-35000 France

²Université de Rennes 1, LTSI, F-35000, France

LTSI, Campus de Beaulieu, Université de Rennes 1, 35042 Rennes Cedex,

Paul.Frogerais@univ-rennes1.fr, Jean-Jacques.Bellanger@univ-rennes1.fr

Résumé – Le problème abordé ici concerne l'identification d'un modèle continu-discret physiologiquement réaliste pour l'interprétation des signaux électroencéphalographiques de patients épileptiques. L'approche étant celle du maximum de vraisemblance, le calcul de celle-ci par la mise en oeuvre d'un filtre particulaire est présenté. Ce filtre approxime la densité d'importance optimale en discrétisant l'équation différentielle stochastique régissant l'état par la méthode de Runge-Kutta stochastique à l'ordre 2, et en sur-échantillonnant (par interpolation) de la suite observée. L'expérimentation sur deux signaux simulés montre l'apport de ce sur-échantillonnage en vue de maximiser la vraisemblance ainsi calculée.

Abstract – The identification of a continuous-discrete physiological model is presented to interpret EEG signals. The maximum likelihood approach is proposed and a particle filter is used to calculate the likelihood function. The stochastic differential equation is discretized using a 2 order Runge-Kutta method. The combination of this scheme with an interpolation of the observed data is used to approximate the optimal importance density. Experimentations with two simulated signals illustrate the enhancement that can be obtain when we search for the maximum likelihood.

1 L'interprétation des signaux EEG en épilepsie

Le problème posé est d'identifier un modèle basé sur la physiologie pour interpréter des signaux EEG enregistrés chez des patients épileptiques. Ce modèle, déjà proposé par ailleurs [7], présente des non linéarités et son identification délicate peut être envisagée de diverses manières [2]. Nous abordons ici le problème de l'approximation de la vraisemblance dans la perspective d'identifier au sens du maximum de vraisemblance. Le modèle peut être mis sous la forme d'un système continu discret :

$$dX = f(X, \theta)dt + G(\theta)d\beta \quad (1)$$

$$Y_k = HX_k + v_k \quad (2)$$

résumant le système suivant :

$$\begin{cases} dx_i = x_{i+5}dt & i = 0, \dots, 4 \\ dx_5 = (Aa\mathcal{S}(x_1 - x_2 - x_3) - 2ax_5 - a^2x_0)dt \\ dx_6 = (Aa(m_p + C_2\mathcal{S}(C_1x_0)) - 2ax_6 - a^2x_1)dt + Aad\beta \\ dx_7 = (BbC_4\mathcal{S}(C_3x_0) - 2bx_7 - b^2x_2)dt \\ dx_8 = (GgC_7\mathcal{S}(C_5x_0 - C_6x_4) - 2gx_8 - g^2x_3)dt \\ dx_9 = (Bj\mathcal{S}(C_3x_0) - 2jx_9 - j^2x_4)dt \\ dx_{10} = (G_{PH}(x_6 - x_7 - x_8) - \frac{1}{\tau}x_{10})dt \end{cases} \quad (3)$$

avec,

$$\mathcal{S}(v) = 2e_0/(1 + e^{\tau(v_0 - v)}) \quad (4)$$

$$Y_k = x_{10}(t_k) + v_k \quad (5)$$

Le vecteur de paramètre $\Phi = (\theta, \gamma)$ est constitué de $\theta = (A, B, G)$ non connu et que l'on cherchera à estimer et de γ supposé connu. L'interprétation de (A, B, G) est liée

à la notion de balance excitation/inhibition dans le tissu neuronal [7] [1]. L'équation différentielle stochastique (1) a pour trajectoire d'état $X(t) \in \mathbb{R}^{11}$, $t \in [0, T]$ et $\beta(t)$ est un processus brownien modélisant des activités neuronales afférentes au sous système considéré. Son coefficient de diffusion σ est supposé constant, le terme $G(\theta)$ étant ici indépendant de θ . L'observation (2) est une forme linéaire $Y_k = HX_k$ de l'état $X_k = X(t_k)$ en N instants $t_k = k\Delta$.

2 Approximation de la fonction de vraisemblance

Dans un contexte d'estimation de θ , à partir d'une suite observée $y_{1:N} = \{y_1, \dots, y_N\}$, le problème considéré ici est celui d'une évaluation numérique de la fonction vraisemblance $\theta \rightarrow p_{Y_{1:N}}^\theta(y_{1:N})$. En s'appuyant sur la suite aléatoire $X_{1:N}$, la fonction de vraisemblance peut s'écrire [5] :

$$p_{Y_{1:N}}^\theta(y_{1:N}) = E^\theta[p_{Y_1|X_1}(y_1, X_1)] \times \prod_{k=2}^N E^\theta[p_{Y_k|X_k}(y_k, X_k)|Y_{1:k-1}] \quad (6)$$

La connaissance de la densité $p_{X_k|Y_{1:k-1}}^\theta$ est nécessaire pour l'évaluation de l'espérance conditionnelle (6). Elle peut être théoriquement exprimée grâce aux équations récurrentes du filtre bayésien [3]. Ce filtre ne peut être appliqué en pratique dans le cas général. Ainsi a-t-on recours très souvent à une méthode de Monte Carlo, l'échantillonnage d'importance.

L'échantillonnage d'importance

Cet échantillonnage peut être réalisé séquentiellement au moyen de filtres particuliers qui permettent d'appro-

cher la loi $p_{X_k|Y_{1:k-1}}^\theta$ par un ensemble de N_s particules \tilde{x}_k^i , $i = 1, \dots, N_s$, pondérées, chacune, par un poids w_k^i . Le calcul de l'espérance dans (6) est alors approché par :

$$E^\theta [p_{Y_k|X_k}(y_k, X_k)|Y_{1:k-1}] \simeq \sum_{i=1}^{N_s} w_k^i p_{Y_k|X_k}(y_k, \tilde{x}_k^i) \quad (7)$$

Ainsi, une approximation de la log-vraisemblance est donnée par l'expression suivante :

$$\hat{L}(y_{1:N}, \theta) = \sum_{k=1}^N \log \sum_{i=1}^{N_s} w_k^i p_{Y_k|X_k}(y_k, \tilde{x}_k^i) \quad (8)$$

Un filtre particulaire met en jeu N_s particules \hat{x}_k^i et leur poids associé w_k^i . A l'étape initiale on génère ces particules dans la loi initiale p_{X_1} . Les étapes suivantes, $k > 1$, consistent à simuler N_s nouvelles particules à partir de N_s particules $\hat{x}_{1:k-1}^i$ obtenues à l'étape $k-1$ et de l'observation $y_{1:k}$, selon une loi d'importance de densité $q_{X_k|X_{1:k-1}, Y_{1:k}}$. A chaque particule ainsi générée est associé un poids défini récursivement [3] :

$$w_k^i = w_{k-1}^i \frac{p_{X_k|X_{k-1}}(\hat{x}_k^i, \hat{x}_{k-1}^i) p_{Y_k|X_k}(y_k, \hat{x}_k^i)}{q_{X_k|X_{1:k-1}, Y_{1:k}}(\hat{x}_k^i, \hat{x}_{1:k-1}^i, y_{1:k})} \quad (9)$$

Il est possible d'approcher le calcul de l'espérance dans (6) par $E^\theta [p_{Y_k|X_k}(y_k, X_k)|Y_{1:k-1}] \simeq \sum_{i=1}^{N_s} w_k^i [5]$

Considérons dans un premier temps la technique bootstrap [4] utilisant comme densité d'importance la densité de transition $p_{X_k|X_{k-1}}(x_k, x_{k-1})$ avec des poids

$w_k^i = w_{k-1}^i p_{Y_k|X_k}(y_k, \hat{x}_k^i)$. En pratique on constate que les étapes de simulation et de pondération seules ne suffisent pas car au fil des itérations la majeure partie du poids total du nuage se concentre sur une particule. Or la simulation et la pondération des particules de poids très faible représente un temps de calcul non négligeable alors qu'elles apportent peu en terme d'estimation. Ce phénomène appelé dégénérescence des poids est décrit dans [6]. Une des méthodes employée pour l'éviter est une étape supplémentaire de rééchantillonnage [4] qui consiste à tirer dans les N_s particules suivant une loi donnée par les poids. Ce rééchantillonnage peut-être systématique ou conditionné par le résultat d'un test sur la variance des poids. Ainsi est-il possible qu'à une itération k de l'algorithme plusieurs particules aient la même position. Si la dynamique de l'état est faiblement bruitée (ou presque déterministe), ces particules évolueront les unes très proches des autres, ce qui limite alors fortement la capacité d'exploration du nuage.

Lorsque la loi d'observation est très localisée (ce qui est notre cas), l'évolution du nuage se faisant indépendamment de l'observation, aucune particule ne visite la zone de forte vraisemblance. Après rééchantillonnage seules quelques particules (les plus vraisemblables) seront sélectionnées. Ce qui limite là aussi la capacité exploratrice du nuage. Ces deux phénomènes décrits par exemple dans [6] sont appelés dégénérescence des particules.

Une des solutions pour éviter la dégénérescence des particules est de prendre en compte l'information apportée par l'observation à l'instant k lors de la génération du nuage. Les particules évolueront non plus selon la loi de transition $p_{X_k|X_{k-1}}(x_k, x_{k-1})$ mais selon ce que l'on appelle la densité d'importance optimale [3] :

$p_{X_k|X_{k-1}, Y_k}(x_k, \hat{x}_{k-1}^i, y_k)$ et seront pondérées récursivement par $w_k^i = w_{k-1}^i p_{Y_k|X_{k-1}}(y_k, \hat{x}_{k-1}^i)$ pour suivre (9).

Densité d'importance optimale et simulation de l'EDS : proposition d'une méthode.

Pour appliquer la méthode bootstrap il est possible d'utiliser la méthode d'Euler pour simuler sur $[t_{k-1}, t_k]$ l'équation différentielle stochastique (1) et générer ainsi des particules selon une loi approchant $p_{X_k|X_{k-1}}$ en introduisant le schéma au pas $\Delta' = \Delta/r$ (r entier > 0) :

$$X'_{n,k} = X'_{n-1,k} + f(X'_{n-1,k}, \theta)\Delta' + G(\theta)W_{r,n}^r \quad (10)$$

avec $n = 0, \dots, r-1$ et $k = 1, \dots, N$. La suite des variables aléatoires $W_{r,n}^r$ représente les incréments d'un Brownien sur un intervalle Δ' (leur variance $(\sigma_w^r)^2$ étant proportionnelle à $1/r$). Pour diminuer r il est possible d'utiliser une méthode de type Runge-Kutta stochastique d'ordre 2 ou 4 [8]. Dans notre application, à l'ordre deux, on aboutit à un schéma de la forme :

$$X'_{n,k} = f_{rk2}(X'_{n-1,k}, \theta, \Delta') + G_{rk2}(\Delta', \theta)W_{r,n} \quad (11)$$

$$W_{r,n} \sim \mathcal{N}(0, \sigma_w^r), \quad \sigma_w^r = \sigma_w^1 \frac{1}{\sqrt{r}}$$

Mais pour appliquer la méthode d'échantillonnage d'importance optimal en des instants $k\Delta + n\Delta'$, $0 < n < r$, on se heurte à la difficulté de simuler dans la loi $p_{X'_n|X_{k-1}, Y_k}$. A notre connaissance, il n'a pas été proposé de méthode pour y répondre. Dans [3] le calcul de $p_{X_k|X_{k-1}, Y_k}$ et de la densité $p_{Y_k|X_{k-1}}$, nécessaire au calcul des poids, est donné dans le cas où (1) est remplacé par un système markovien discret de la forme $X_k = u(X_{k-1}) + W_k$ avec W_k gaussien et de matrice de covariance C_w inversible. Ce calcul ne peut être appliqué à (11) sans aménagement car C_w est singulière et de plus r pourra être plus grand que 1. Nous proposons ici une méthode qui contourne ces difficultés. Elle consiste à recalculer la moyenne et la covariance de la loi d'importance optimale du système discret (11) (ici gaussienne) en considérant des échantillons pris aux instants $k\Delta'$: $\tilde{y}_{n,k} = HX_{n,k} + v_k$. Ne disposant réellement que d'échantillons aux instants $k\Delta$, nous proposons d'introduire un suréchantillonnage (par interpolation) de l'observation pour disposer de pseudo observations $y'_{n,k}$, avec $y'_{0,k} = y_k$ et $y'_{r,k} = y_{k+1}$. Cette approche se justifie ici par un bruit d'observation faible et en supposant que $1/\Delta$ est supérieur à la fréquence d'échantillonnage sans perte pour HX_k .

Considérons les deux situations :

1. Le pas d'échantillonnage physique Δ est suffisant pour obtenir une bonne approximation par Runge-Kutta au pas Δ de la trajectoire d'état.
2. Au pas Δ l'erreur d'approximation par Runge-Kutta 2 stochastique est trop importante.

Dans la situation 1) nous appliquons notre technique d'échantillonnage d'importance pour la suite observée $y_{1:N}$ avec $\Delta' = \Delta$ dans (11) et nous en déduisons la log-vraisemblance $\hat{L}(y_{1:N}, \theta)$ (8). Dans la situation 2) nous introduisons les observations $y'_{1:rN}$ et appliquons la même technique au pas Δ' sur (11) pour cette fois obtenir une vraisemblance $\hat{L}'(y'_{1:rN}, \theta)$ calculée sur les rN échantillons.

3 Expérimentations

A titre d'illustration nous donnons des résultats en simulation en choisissant deux jeux de paramètres $\theta_1 =$

(5, 20, 50) et $\theta_2 = (20, 10, 50)$ correspondant à deux type de dynamique du signal EEG à l'approche et en début de crise (Ces valeurs de paramètres ont été obtenues par identification empirique sur signaux réels). On note ici $x10_{1:rN}^{\theta_j} = \{HX'_{n,k}; k = 1, \dots, N, n = 0, \dots, r - 1\}$ la suite obtenue en calculant les $X'_{n,k}$ par la méthode de Runge-Kutta à l'ordre deux avec un pas suffisamment fin $\Delta' = 1/(256 \times r)$ où $r = 8$ et $N = 2048$.

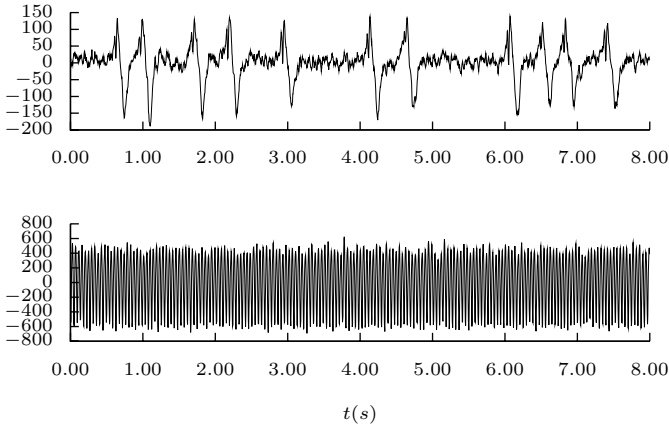


FIG. 1 – Une réalisation des deux types d'activité simulés (en haut) $y_{1:N}^{\theta_1}$ (en bas) $y_{1:N}^{\theta_2}$.

Pour chaque valeur θ_j correspondants, trois signaux sont construits à partir de $x10_{1:rN}^{\theta_j}$. Le premier signal (simulant le signal réel) $y_{1:N}^{\theta_j} = \{y_n^{\theta_j}, n = 1, \dots, N\}$ correspond au signal $x10_{1:rN}^{\theta_j}$ sous échantillonné à 256Hz (on prend un échantillon sur 8) auquel on ajoute un bruit de mesure blanc centré gaussien $v_n \sim \mathcal{N}(0, \sigma_v)$. Ensuite Le signal $y_{1:N}^{\theta_j}$ est interpolé d'un facteur 8 par deux méthodes : interpolation linéaire et interpolation par spline cubique pour créer respectivement les deux derniers signaux $yl_{1:rN}^{\theta_j} = \{yl_k^{\theta_j}, k = 1, \dots, rN\}$ et $y_{cs}^{\theta_j} = \{y_{cs_k}^{\theta_j}, k = 1, \dots, rN\}$. La figure 2 représente ces différents signaux sur quelques échantillons. L'écart entre les signaux interpolés et $x10_{1:rN}^{\theta_j}$ est dû au bruit v_n et à l'erreur d'interpolation. Pour différentes valeurs de σ_v la distance suivante est calculée

$$\epsilon_i^{\theta_j}(\sigma_v) = \frac{1}{rN} \sum_{k=1}^{rN} (x10_k - y_i^{\theta_j})^2 \quad i \in \{cs, l\} \quad (12)$$

et représentée figure (3). Pour des fortes valeurs de σ_v^2 cette distance est très proche de σ_v^2 ce qui suggère que l'interpolation apporte peu de bruit complémentaire. On remarque que pour des faibles valeurs de σ_v^2 , cette distance est bornée inférieurement par un seuil correspondant à l'erreur due à l'interpolation seule. Cette erreur pourra être prise en compte pour l'évaluation de la vraisemblance quand σ_v est du même ordre de grandeur.

Les deux signaux $y_{1:N}^{\theta_1}$ et $y_{1:N}^{\theta_2}$ correspondent respectivement aux situations 1) et 2) décrites à la fin de la section 2. Une réalisation de la vraisemblance $A \rightarrow \hat{L}(y_{1:N}^{\theta_j}, A, B_j, G_j)$ calculée pour différentes valeurs de A sur ces deux signaux (B_j et G_j étant fixés) est représentée figure (4) et (5). La variance du bruit additif pour les deux modèles est $\sigma_v^2 = 10$.

La courbe de la figure (4) présente un maximum global

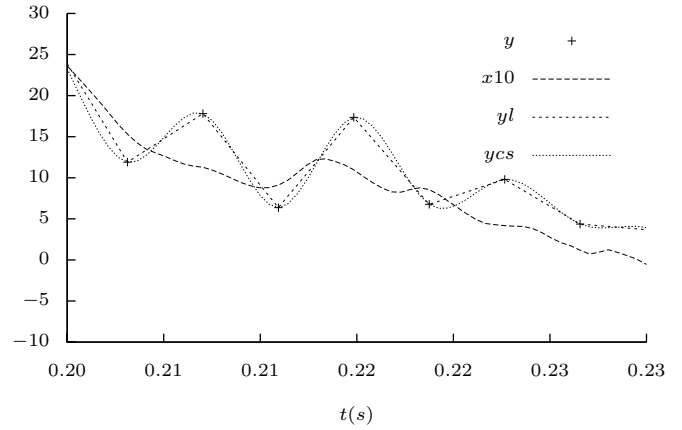


FIG. 2 – Les quatre signaux construit par la simulation du modèle avec $\theta_1 = (5, 20, 50)$ et $\sigma_v = 10$

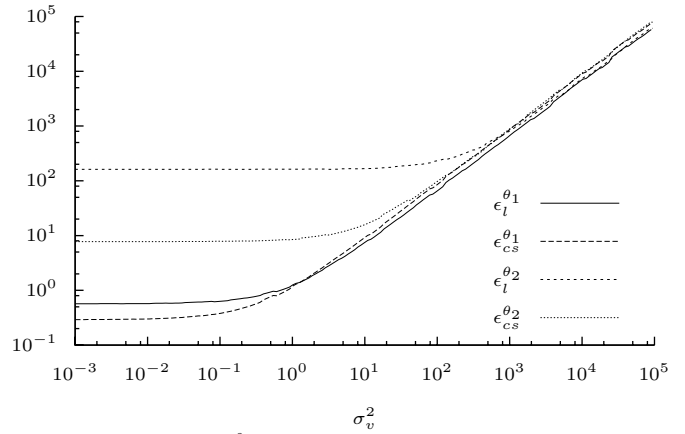


FIG. 3 – distances $\epsilon_i^{\theta_j}(\sigma_v)$ entre signaux interpolés et signal non bruité $x10_{1:rN}^{\theta_j}$.

aux alentours de la vraie valeur ($A = 5.0$). Par contre pour la situation 2) figure (5), le mode aux alentours de la vraie valeur ($A = 20.0$) n'est pas un maximum global. Dans cette situation, pour montrer l'amélioration apportée par le calcul de la vraisemblance sur les signaux interpolés, les vraisemblances $\hat{L}'(x10_{1:rN}^{\theta_2}, A, 10, 50)$, $\hat{L}'(yl_{1:rN}^{\theta_2}, A, 10, 50)$ et $\hat{L}'(y_{cs}^{\theta_2}, A, 10, 50)$ sont calculées (avec $\sigma_v^2 = 10$) pour différentes valeurs de A et représentées figure(6). La vraisemblance $\hat{L}'(x10_{1:rN}^{\theta_2}, A, 10, 50)$ correspond ainsi à la courbe de vraisemblance que l'on obtiendrait si on échantillonnait le système à une fréquence $1/\Delta'$ pour être dans la situation 1). Elle nous sert donc de référence. La vraisemblance calculée sur le signal interpolé par la méthode spline cubique $\hat{L}'(y_{cs}^{\theta_2}, A, 10, 50)$ suit assez bien cette référence par rapport à la vraisemblance sur le signal $yl_{1:rN}^{\theta_2}$. A noter que dans cette situation ($\theta = \theta_2, \sigma_v^2 = 10$) l'interpolation par spline cubique est bien meilleure (figure (3)). Pour les trois vraisemblances on retrouve le maximum global aux alentours de la vraie valeur ($A = 20$). Cette méthode s'avère donc nécessaire pour identifier le maximum de vraisemblance contrairement au cas de la figure (5)

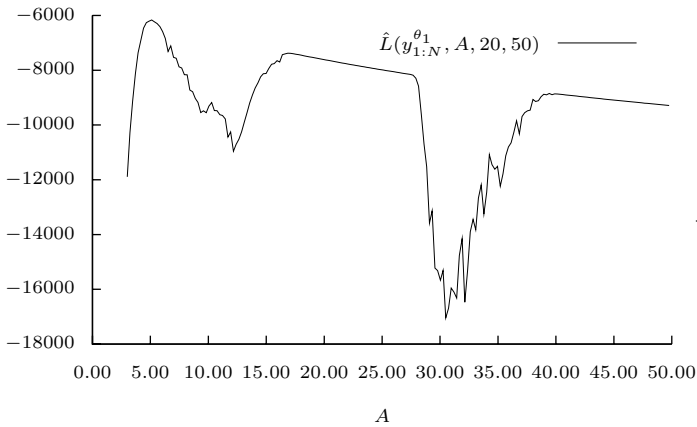


FIG. 4 – La vraisemblance $\hat{L}(y_{1:N}^{\theta_1}, A, 20, 50)$ pour différentes valeurs de A ($\sigma_v^2 = 10$).

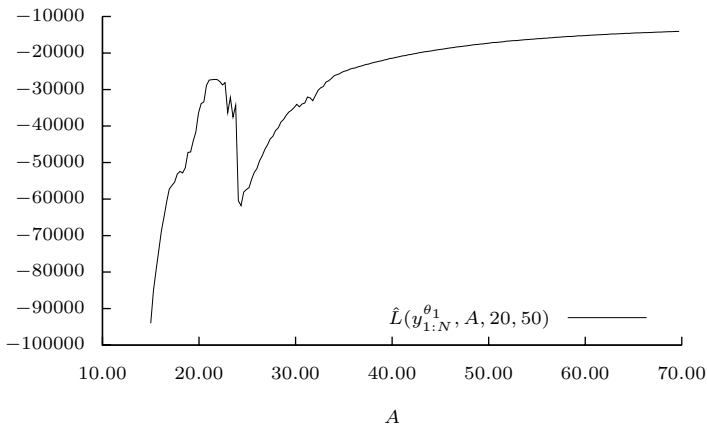


FIG. 5 – La vraisemblance $\hat{L}(y_{1:N}^{\theta_2}, A, 10, 50)$ pour différentes valeurs de A ($\sigma_v^2 = 10$).

4 Conclusion

Le problème d'identification abordé dans le contexte de l'interprétation physiologique de signaux EEG en épilepsie n'est pas aisée, et nous explorons ici l'approche du maximum de vraisemblance. Plus précisément, le problème auquel nous essayons d'apporter une solution est celui du calcul de cette vraisemblance au moyen d'algorithmes particuliers dans le cas où la fréquence d'échantillonnage de l'observation est insuffisante pour discrétiser sans trop d'erreur l'équation différentielle stochastique sous-jacente. La méthode proposée, simple à implémenter et pertinente dans le cas où le bruit d'observation est faible, consiste à fournir à un algorithme particulière des pseudo-observations obtenues par interpolation de la sortie. Ceci nous permet d'utiliser une méthode d'échantillonnage d'importance optimal grâce à laquelle le phénomène classique de dégénérescence du nuage particulaire est plus facilement contrôlé. L'amélioration apportée est démontrée dans un cadre de simulation sur deux instances du modèle. Ce travail concernant le calcul de la vraisemblance est actuellement poursuivi pour réaliser en particulier une maximisation globale de la vraisemblance par rapport à un ou plusieurs paramètres (algorithmes essais particuliers ou génétiques) s'appuyant sur des méthodes plus locales

(EM, gradient). L'application concrète de ces techniques dans l'analyse de beaucoup de données patient nécessitera évidemment des heuristiques pour limiter les temps de calcul en intégrant un maximum d'information a priori.

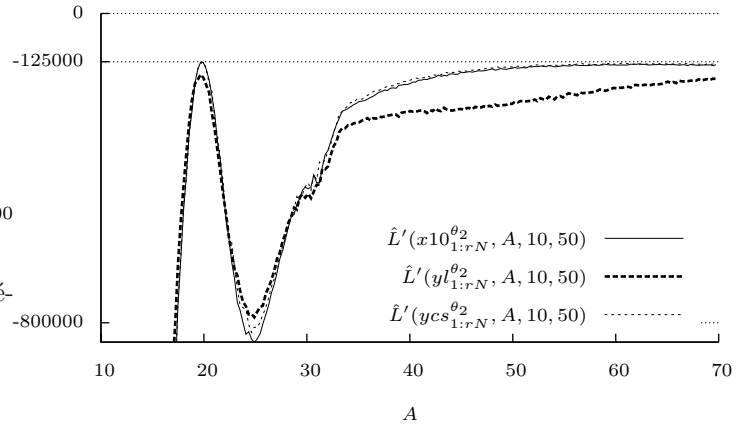


FIG. 6 – Les vraisemblances $\hat{L}'(y_{1:rN}^{\theta_2}, A, 10, 50)$ et $\hat{L}'(x_{10}^{\theta_2}, A, 10, 50)$ pour différentes valeurs de A .

Références

- [1] J. J. Bellanger, O. Djelloul, and F. Wendling. Modélisation et identification de modèles neuronaux pour les signaux eeg en épilepsie. In *gretsi*, 2003.
- [2] J. J. Bellanger, A. Hernandez, F. Wendling, and P. Frogerais. Mesure de l'excitation et de l'inhibition dans le tissu neuronal en épilepsie par identification d'un modèle dynamique non linéaire d'activité eeg. In *gretsi*, 2005.
- [3] Olivier Cappé, Eric Moulines, and Tobias Ryden. *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, springer edition, 2005.
- [4] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings*, 1993.
- [5] M. Hürzeler and H. R. Künsch. *Approximating and Maximizing the likelihood for a general state-space model*. 2001.
- [6] N. Oudjane. *Stabilité et approximations particulières en filtrage non linéaire. Application au pistage*. PhD thesis, Université de Rennes 1, 2000.
- [7] F. Wendling, F. Bartolomei, J. J. Bellanger, and P. Chauvel. Epileptic fast activity can be explained by a model of impaired gabaergic dendritic inhibition. *European Journal of Neuroscience*, 15(9) :1499–1508, 2002.
- [8] Joshua Wilkie. Numerical methods for stochastic differential equations. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 70(1) :017701, 2004.