



DÉTECTION ET ESTIMATION POUR LA TRANSMISSION SUR UN CANAL INCONNU

Ghassan KAWAS KALEH
Robert VALLET

Ecole Nationale Supérieure des Télécommunications,
46, rue Barrault 75634 Paris CEDEX 13.

RÉSUMÉ

Nous présentons un traitement itératif basé sur l'algorithme EM pour l'estimation à maximum de vraisemblance des paramètres d'un canal de transmission à mémoire, linéaire ou non linéaire. Les fonctions de vraisemblance calculées par l'algorithme permettent à la fin des itérations d'effectuer aisément des décisions optimales sur les symboles suivant le critère du minimum de la probabilité d'erreur. Le système proposé permet d'économiser la perte due à la transmission de symboles connus usuellement utilisés pour l'estimation du canal.

SUMMARY

We present an iterative method for joint channel parameters estimation and symbol detection via the EM algorithm. Channel parameters, including noise variance, are estimated using maximum likelihood criterion. Minimum error probability decisions on symbols are easily obtained at the end of iterations. The proposed receiver is valid for both linear and nonlinear channels. It allows an improvement in the system throughput by sparing the usual transmission of known symbols employed for channel identification.

1. INTRODUCTION

Nous présentons ici un nouveau récepteur pour la transmission d'un bloc de symboles d'information à travers un canal à mémoire (dispersif). La détection des symboles est donc effectuée en présence de bruit gaussien et d'interférences entre symboles. Le récepteur se distingue des récepteurs connus par les deux particularités suivantes: il n'a pas besoin d'une identification préalable du canal, et il est valable pour les canaux linéaires et non linéaires. Dans le cas où une estimation grossière du canal est disponible, le récepteur est capable d'améliorer l'estimation en utilisant l'observation véhiculant les symboles inconnus.

Les récepteurs connus dans la littérature (voir dans [1] une synthèse et une bibliographie de récepteurs pour canaux linéaires et non linéaires), supposent que la transmission d'une séquence de symboles d'information est accompagnée de celle d'une séquence de symboles connus (par le récepteur) qui sert à estimer la réponse impulsionnelle du canal ou à adapter les coefficients de l'égaliseur incorporé dans le récepteur. Ces symboles connus font réduire le taux effectif de transmission (throughput). Dans certaines applications comme la transmission sur canaux ionosphériques ou la transmission avec évocation de fréquence, ce taux peut atteindre la valeur 0,5 [2]. Notre récepteur vise à supprimer cette perte. Les paramètres nécessaires pour calculer la vraisemblance a

posteriori d'un symbole sont estimés à partir de la totalité de l'observation et suivant le critère du maximum de vraisemblance en utilisant l'algorithme itératif EM (Expectation-Maximization) présenté dans [4]. Nous utilisons surtout la version établie par Baum et al. [5] pour le cas particulier où l'observation manquante est une chaîne de Markov. Lorsque la convergence est atteinte, les fonctions de vraisemblance utilisées par l'algorithme fournissent des décisions optimales sur les symboles en minimisant la probabilité que le symbole décidé soit différent du symbole émis.

D'autres part, on remarque que dans des systèmes de transmission récents, un traitement sophistiqué est effectué; les fonctions estimation de canal et détection de symboles sont itérées autant que possible en vue d'une amélioration des performances [2, 3]. L'étude présentée ici est motivée par ce traitement itératif; elle élabore une procédure d'itération dictée par la théorie.

2. MODÉLISATION

Les symboles d'information a_1, a_2, \dots, a_K sont complexes, i.i.d. et prennent des valeurs équiprobables dans l'alphabet M -aire $\{\alpha_m; m = 1, 2, \dots, M\}$. L'observation (en bande de base) est complexe, à temps discret (obtenue par l'échantillonnage au rythme de Nyquist d'une observation continue) et s'écrit



$$y_t = h(x_t) + w_t, \quad t = 1, 2, \dots, K \quad (1)$$

où w_t est un bruit i.i.d. gaussien complexe centré et de variance inconnue σ^2 et

$$x_t \triangleq (a_t, a_{t-1}, \dots, a_{t-L}) \quad (2)$$

x_t prend des valeurs équiprobables dans l'ensemble $\{\xi_n; n = 1, 2, \dots, M^{L+1}\}$. Le système $h(\cdot)$ est à temps discret, inconnu, linéaire ou non linéaire et de mémoire finie L . Il modélise le canal équivalent en bande de base. Dans le cas général, l'observation y_t est un vecteur de dimension égale au rapport du rythme d'échantillonnage (de l'observation continue) au rythme des symboles. Cependant, afin de simplifier la présentation, y_t est supposée scalaire; l'extension au cas vectoriel ne posant pas de difficultés.

On note par s_t l'état du système à l'instant t , i.e.

$$s_t \triangleq (a_{t-1}, a_{t-2}, \dots, a_{t-L}) \quad (3)$$

Il y a M^L états. La séquence $\{s_t\}$ est une chaîne de Markov finie et du premier ordre. On suppose que l'état initial s_1 et l'état final s_{K+1} sont connus; l'extension des résultats au cas d'états initial et final inconnus est simple.

On note y_t^n la séquence y_t, y_{t+1}, \dots, y_m . Ayant reçu l'observation totale y_1^K , le récepteur fournit une décision optimale sur chacun des symboles a_t en minimisant la probabilité que le symbole issu de la décision soit différent du symbole émis. Ceci est équivalent à chercher

$$\max_m \max_{\theta} P[a_t = \alpha_m; \theta / y_1^K] \quad (4)$$

où la vraisemblance *a posteriori* $P[\cdot]$ est indexée par θ qui désigne l'ensemble des paramètres nécessaires à déterminer la probabilité $P[\cdot]$. Au lieu d'estimer $h(\cdot)$ par l'intermédiaire de ses coefficients de Volterra [10], nous proposons d'estimer les paramètres

$$\theta \triangleq \{\sigma^2, h(\xi_n); n = 1, 2, \dots, M^{L+1}\} \quad (5)$$

3. ESTIMATION DES PARAMÈTRES

Soit la séquence d'observation y_1^K de densité de probabilité $p(y_1^K; \theta)$ indexée par le vecteur des paramètres θ . L'estimé $\hat{\theta}$ suivant le critère du maximum de vraisemblance et basée sur l'observation y_1^K est la valeur de θ qui maximise la vraisemblance $p(y_1^K; \theta)$. Soit $\{\vec{\xi}_l; l=1, 2, \dots, M^K\}$ l'ensemble des M^K réalisations de la séquence x_1^K . La vraisemblance se calcule au moyen de l'expression

$$p(y_1^K; \theta) = \sum_{l=1}^{M^K} L(x_1^K = \vec{\xi}_l, y_1^K; \theta) \quad (6)$$

$$L(x_1^K = \vec{\xi}_l, y_1^K; \theta) \triangleq P[x_1^K = \vec{\xi}_l] p(y_1^K; \theta / x_1^K = \vec{\xi}_l) \quad (7)$$

avec $p(\cdot)$ désignant une densité de probabilité conditionnelle. Malgré l'indépendance entre les paramètres et

les symboles, la définition de $L(\cdot)$ par (7) contient les probabilités *a priori* des symboles car ceci permet l'utilisation de la notation (7) pour définir d'autres vraisemblances *a posteriori*.

Trouver la solution θ qui maximise (6) est un problème inextricable. On peut cependant le trouver en utilisant l'algorithme EM qui, par itérations successives, garantit un accroissement monotone de la vraisemblance. Au lieu de maximiser la vraisemblance (6) directement, l'algorithme maximise une fonction auxiliaire [5] définie par

$$Q(\theta, \theta') \triangleq \sum_{l=1}^{M^K} L(x_1^K = \vec{\xi}_l, y_1^K; \theta) \log L(x_1^K = \vec{\xi}_l, y_1^K; \theta') \quad (8)$$

où le rôle de données manquantes dans l'algorithme EM est assumé par x_1^K . On peut en effet montrer suivant [5] que

$$Q(\theta, \theta') \geq Q(\theta, \theta) \Rightarrow p(y_1^K; \theta') \geq p(y_1^K; \theta) \quad (9a)$$

et que l'égalité est obtenue si et seulement si : pour tout $\vec{\xi}_l$,

$$L(x_1^K = \vec{\xi}_l, y_1^K; \theta) = L(x_1^K = \vec{\xi}_l, y_1^K; \theta') \quad (9b)$$

Les relations (9) constituent le fondement de l'algorithme. L'algorithme commence donc avec une estimée initiale $\theta^{(0)}$. L'estimée à la i -ième itération, notée $\theta^{(i)}$, est utilisée pour calculer une meilleure estimée $\theta^{(i+1)}$ en effectuant les deux étapes suivantes :

- Etape E (Expectation): calculer $Q(\theta^{(i)}, \theta')$.
- Etape M (Maximization): chercher $\theta^{(i+1)}$ qui maximise $Q(\theta^{(i)}, \theta')$ par rapport à θ' .

Généralement, si $Q(\theta, \theta')$ est continue en θ et θ' , l'algorithme converge vers un maximum local de la vraisemblance.

Cherchons $\theta^{(i+1)}$ qui maximise $Q(\theta^{(i)}, \theta')$ par rapport à θ' . On obtient de (7) et des propriétés du bruit

$$\log L(x_1^K = \vec{\xi}_l, y_1^K; \theta) = C_1 + \sum_{t=1}^K \sum_{n=1}^{M^{L+1}} \log p(y_t; \theta / x_t = \xi_n) \delta(x_t, \xi_n) \quad (10)$$

où C_1 , comme C_1 et C_2 dans (11) et (12) qui suivent, regroupe des termes indépendants de θ , et

$$\begin{aligned} \delta(x_t, \xi_n) &= 1; x_t = \xi_n \\ &= 0; x_t \neq \xi_n \end{aligned}$$

$$\log p(y_t; \theta / x_t = \xi_n) = C_2 - \log(\sigma^2) - \frac{1}{\sigma^2} |y_t - h(\xi_n)|^2 \quad (11)$$

En rapportant les relations précédentes dans (8) on obtient

$$Q(\theta, \theta') = C_3 + \sum_{t=1}^K \sum_{n=1}^{M^{L+1}} [-\log(\sigma^2) - \frac{1}{\sigma'^2} |y_t - h(\xi_n)|^2]$$

$$-h'(\xi_n)^2 \left[\sum_{l=1}^{M^k} L(x_l^K = \xi_l, y_l^K; \theta) \delta(x_l, \xi_n) \right] \quad (12)$$

La somme interieure (entre accolades) donne

$$L(x_t = \xi_n, y_t^K; \theta) \triangleq P[x_t = \xi_n] p(y_t^K; \theta / x_t = \xi_n) \quad (13)$$

Il reste à exprimer (12) en fonction de (13), et à annuler sa dérivée par rapport aux paramètres σ' et $h'(\xi_n)$ correspondant à θ' . On obtient ainsi les paramètres $\sigma^{2(i+1)}$ et $\{h^{(i+1)}(\xi_n)\}$ correspondant à $\theta^{(i+1)}$ qui maximise $Q(\theta^{(i)}, \theta')$ par rapport à θ' :

$$h^{(i+1)}(\xi_n) = \frac{\sum_t L(x_t = \xi_n, y_t^K; \theta^{(i)}) y_t}{\sum_t L(x_t = \xi_n, y_t^K; \theta^{(i)})}; n=1, \dots, M^{L+1} \quad (14)$$

$$\sigma^{2(i+1)} = \frac{\sum_n \sum_t L(x_t = \xi_n, y_t^K; \theta^{(i)}) |y_t - h^{(i+1)}(\xi_n)|^2}{\sum_n \sum_t L(x_t = \xi_n, y_t^K; \theta^{(i)})} \quad (15)$$

Le calcul de (14) et (15) est effectué à chaque itération. Il nécessite le calcul des vraisemblances (13) pour tout t et tout ξ_n . Ceci est effectué par les récursions dites "forward-backward" que nous produisons ci-dessous. Afin de simplifier l'écriture nous y négligeons le paramètre θ .

On remarque d'abord que la réalisation de l'événement $x_t = \xi_n$ est équivalente à celle du couple d'événements $s_t = i$ et $s_{t+1} = j$. Alors, en exploitant le caractère markovien de s_t , la relation (13) s'écrit

$$L(x_t = \xi_n, y_t^K) = P[s_t = i, s_{t+1} = j] p(y_t^{t-1} / s_t = i) p(y_t / x_t = \xi_n) p(y_{t+1}^K / s_{t+1} = j) = \gamma_t(i) \beta_{t+1}(j) P[s_{t+1} = j / s_t = i] p(y_t / s_t = i, s_{t+1} = j) \quad (16)$$

où

$$\gamma_t(i) \triangleq L(s_t = i, y_t^{t-1}) \quad (17)$$

$$\beta_{t+1}(j) \triangleq p(y_{t+1}^K / s_{t+1} = j) \quad (18)$$

On peut déduire la récursion "forward"

$$\begin{aligned} \gamma_t(j) &= \sum_i P[s_t = j] P[s_{t-1} = i / s_t = j] \\ &\quad p(y_t^{t-2}, y_{t-1} / s_{t-1} = i, s_t = j) \\ &= \sum_i \gamma_{t-1}(i) P[s_t = j / s_{t-1} = i] p(y_{t-1} / s_{t-1} = i, s_t = j) \end{aligned} \quad (19)$$

et la récursion "backward"

$$\begin{aligned} \beta_t(i) &= \sum_j p(y_t^K / s_t = i, s_{t+1} = j) P[s_{t+1} = j / s_t = i] \\ &= \sum_j \beta_{t+1}(j) P[s_{t+1} = j / s_t = i] p(y_t / s_t = i, s_{t+1} = j) \end{aligned} \quad (20)$$

Signalons que ces deux récursions sont utilisées dans [6] pour le cas d'un canal linéaire de réponse impulsionnelle connue.

Précisons maintenant les étapes de l'algorithme. Des tableaux de mémoires sont réservés pour contenir les valeurs $\gamma_t(i)$ et $\beta_t(i)$ données par (17) et (18). On dispose de l'observation y_1^K .

Initialisation : $\theta^{(0)}$ est arbitraire, $\gamma_1(i) = \beta_{K+1}(j) = 1$ pour les états i et j connus et 0 ailleurs.

L'algorithme suit les étapes suivantes :

1. Calculer d'après (19) $\gamma_t(i)$ pour $t = 2, 3, \dots, K+1$ et pour $i = 1, 2, \dots, 2^L$.
2. Calculer d'après (20) $\beta_t(i)$ pour $t = K, K-1, \dots, 1$ et pour $i = 1, 2, \dots, 2^L$.
3. Calculer les vraisemblances (16) pour $t = 1, 2, \dots, K$ et pour $i = 1, 2, \dots, 2^{L+1}$.
4. Estimer les paramètres en utilisant (14) puis (15).
5. Si $\sum_i \gamma_{K+1}(i)$ est inférieur à un seuil fixé aller à 1, sinon fin.

La figure 1 montre l'évolution de l'estimation des paramètres pour un canal de mémoire $L = 2$, des symboles binaires 1 ou -1, et un rapport énergie par bit sur densité spectrale de puissance monolatérale de bruit de 7 dB.

4. DÉTECTION DES SYMBOLES

Soit $\hat{\theta}$ l'estimée obtenue par l'algorithme EM. La décision optimale qui minimise la probabilité d'erreur sur un symbole a_t est de décider en faveur du symbole α_m qui maximise

$$L(a_t = \alpha_m, y_t^K; \hat{\theta}) \triangleq P[a_t = \alpha_m] p(y_t^K; \hat{\theta} / a_t = \alpha_m) \quad (21)$$

Soit Ξ_m l'ensemble des M^L réalisations de x_t pour lesquelles $a_t = \alpha_m$:

$$\Xi_m \triangleq \{\xi_n = (\alpha_m, a_{t-1}, \dots, a_{t-L})\}$$

La vraisemblance a posteriori (21) se calcule par la formule

$$L(a_t = \alpha_m, y_t^K; \hat{\theta}) = \sum_{\xi_n \in \Xi_m} L(x_t = \xi_n, y_t^K; \hat{\theta}) \quad (22)$$

Les termes de la somme sont disponibles car ils ont été calculés par l'algorithme EM en utilisant (16).

5. COMPARAISON AVEC UN SYSTÈME UTILISANT UNE SÉQUENCE D'APPRENTISSAGE

Nous comparons ici le système proposé à un système, appelé système 2, utilisant un détecteur de séquence à maximum de vraisemblance (détecteur de Viterbi) et un préambule de symboles connus qui sert à identifier le canal suivant le critère du moindre carré. Comme dans [2], nous supposons que dans ce système le bloc de K symboles est composé de $K/2$ symboles binaires connus et de $K/2$ symboles d'information. Les deux systèmes disposent de la même puissance et emploient une modulation monodimensionnelle. Pour transmettre K bits d'information, le système proposé utilise K symboles binaires 1 ou -1, et le



système 2 utilise $K/2$ symboles quaternaires 1, 3, -1 ou -3. On ne choisit pas la modulation en quadrature car cet avantage doit être accordé équitablement aux deux systèmes. Ainsi le système proposé gagne 4 dB environ qui représente la dégradation en performance de la modulation monodimensionnelle quaternaire par rapport à la modulation binaire.

Soit E_b (resp. E'_b) l'énergie reçue par bit dans le système proposé (resp. le système 2). L'énergie par bloc est la même pour les deux systèmes. Elle est KE_b pour le premier. Dans le système 2 une énergie $\frac{K}{2} E_b$ est utilisée pour les symboles connus et $\frac{K}{2} 2E'_b$ pour les symboles d'information. On en déduit que $E_b = 2E'_b$. Ceci donne un deuxième avantage de 3 dB au système proposé. Le gain total du système proposé est donc 7 dB. La figure 2 montre les résultats de simulation des deux systèmes.

6. CONCLUSION

Nous avons présenté un estimateur de paramètres à maximum de vraisemblance, conjointement avec un détecteur optimal de symboles pour la transmission sur des canaux linéaires et non linéaires. Le récepteur obtenu permet d'économiser la transmission d'une séquence d'apprentissage. Les simulations ont montré une bonne convergence lorsque le rapport signal à bruit est faible. Comme le détecteur de Viterbi, la complexité s'accroît exponentiellement avec la mémoire du canal et linéairement avec la durée du message.

REFERENCES

[1] S. Benedetto, E. Biglieri and V. Castellani, "Digital Transmission Theory," Prentice-Hall, Inc., N.J., 1987.
 [2] Perl, Shpigel and Reichman, "Adaptive Receiver for Digital communication over HF Cannel," IEEE JSAC, Vol. SAC-5, pp. 304-308, Feb. 1987.
 [3] Perkins and McRae, "A High Performance HF Modem," Proc. Conf. Military Electronics Defense, pp. 137-152, Hanover, W. Germany, May 1982
 [4] Dempster, Laird and Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Roy. Stat. Soc., Ser. 39, pp. 1-38, 1977.
 [5] Baum, Petrie, Soules and Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains," Ann. Math. Stat., vol. 41, pp. 164-171, 1970.
 [6] Chang and Hancock, "On receiver structures for channels having memory," IEEE Trans. Inform. Theory, vol. IT-12, pp. 464-468, oct. 1966.

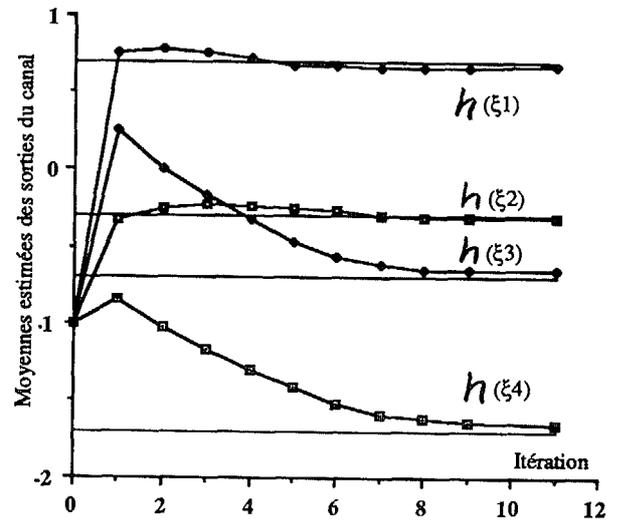


Figure: 1 Convergence des paramètres vers les valeurs nominales.

Le canal est linéaire, de réponse impulsionnelle 0,5; 0,7 et 0,5
 Le signal comporte $K=100$ symboles.
 Les quatre autres valeurs sont symétriques par rapport aux quatre premières.

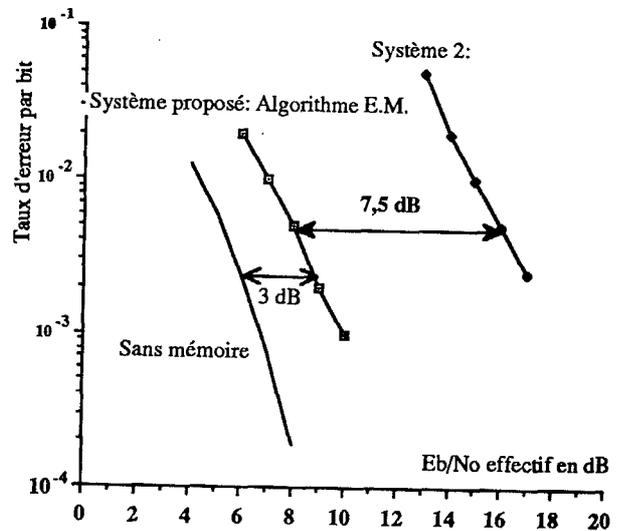


Figure 2: Comparaisons des performances entre l'agorithme E.M. et l'algorithme de Viterbi avec une séquence d'apprentissage de $K/2$ symboles.