

Identification des Coefficients du Model MA du Canal par Sous-estimation de la Densité de Probabilité du Signal Reçu

Jugurta MONTALVÃO^{1,2}, Bernadette DORIZZI¹, João Cesar MOTA³

¹EPH, INT
9 rue Charles Fourier, 91011, Évry Cedex, France

²CCFT, UNIT
264 rue Lagarto, 49010-390, Aracaju, Brésil

³DEE, UFC
BP 6001, 60455-760, Fortaleza, Brésil

jugurta.montalvao@int-evry.fr, Bernadette.Dorizzi@int-evry.fr
mota@dee.ufc.br

Résumé – Nous présentons une stratégie d’estimation autodidacte des coefficients d’un canal de communication avec une structure transversale par estimation paramétrique de la densité de probabilité du signal reçu. Pour diminuer le nombre de gaussiennes qui composent cette densité de probabilité, nous proposons un sous-modèle obtenu par regroupement de quelques gaussiennes en une seule avec dispersion non-radiale. Cela engendre une famille d’estimateurs du canal de communication permettant de varier le rapport entre qualité d’estimation et complexité d’estimateur. Quelques résultats de simulation sont aussi présentés en guise d’illustration.

Abstract – An unsupervised approach to linear channel parameters estimation is proposed. It is based on signal probability density function (pdf) estimation and its originality consists in sub-modeling the pdf by replacing its intrinsic groups of radial basis Gaussians by a few number of nonradial ones, obtained through reduction of the initial ones. The resulting family of channel estimators provides a wide tradeoff between estimation accuracy and computational burden. Some illustrative simulation results are also carried out at the end of this article.

1. Introduction

Ce sujet s’inscrit dans le cadre de l’estimation autodidacte de la réponse impulsionnelle d’un système de communication numérique échantillonné de façon synchrone.

Ce travail a été développé dans le contexte limité d’un système fondé sur un alphabet discret de modulation, un canal modelisé par une structure transversale (i. e. MA, du anglais : *moving average* [8]) et une source de bruit gaussienne blanche.

Dans ce contexte, la fonction densité de probabilité (fdp) conjointe d’une séquence de M échantillons consécutifs du signal reçu (i. e. une fenêtre temporelle d’observations) est une somme de gaussiennes dans l’espace de dimension M . En plus, cette fdp possède comme paramètres les coefficients de la structure transversale du canal et la variance du bruit.

Donc, théoriquement il est possible de réaliser une estimation de ces paramètres à partir du signal reçu, même si on néglige la dépendance statistique entre les fenêtres d’observations. Cette procédure correspond donc à une identification aveugle des coefficients du canal.

L’idée de faire une estimation du canal à partir de fenêtres d’observations « supposées » indépendantes apparaît aussi dans quelques articles récents comme, par exemple Adali et al. [2], qui font appel à une procédure d’estimation appelée *Partial Likelihood*. Moulines et al. [3] utilisent aussi la même idée mais ils justifient leur approche à partir du *Split Data Likelihood*, introduit par Rydén [4], qui à son tour, montre

que ce type d’estimateur est consistant et asymptotiquement normal.

Cependant, d’un point de vue pratique, le nombre de gaussiennes nécessaires pour construire une estimation adéquate de la fdp M -dimensionnelle peut devenir trop grand pour une mise en œuvre d’une stratégie d’identification du canal.

Dans ce travail, nous utilisons une sous-estimation de la fdp du signal reçu — à un nombre de gaussiennes réduit — ce qui rend possible l’implémentation d’une procédure d’estimation avec une quantité de calcul contrôlée. L’estimation des coefficients du canal résulte d’un processus d’optimisation basé sur la minimisation de la divergence de Kullback-Leibler [1] entre la fdp et sa sous-estimation.

Néanmoins, il est évident que l’utilisation d’une sous-estimation de la fdp affecte la qualité de l’estimation, ainsi que le nombre de minima locaux de la fonctionnelle.

Pour éviter les minima locaux, nous utilisons une formalisation de cette approche en termes de physique statistique pour lui appliquer une adaptation de la technique du recuit déterministe (*deterministic annealing*) [6].

Le modèle du canal et les notations utilisées sont présentés dans la section 2.

Dans la section 3, nous présentons notre approche d’estimation du canal ainsi que la fonctionnelle utilisée et les heuristiques qui soutiennent le choix des paramètres de l’algorithme de minimisation de cette fonctionnelle.

Quelques essais ont été réalisés avec le schéma de modulation 2-PAM. Les résultats correspondants sont présentés dans la section 4.

Finalement, les conclusions et perspectives se trouvent dans la section 5 de cet article.

2. Modélisation mathématique du canal

Le schéma de la figure 1 représente un modèle simple d'un système de communication numérique constitué en partie d'une source d'information qui émet un symbole à chaque intervalle de Δt secondes. Ces symboles émis appartiennent à un alphabet fini et forment une séquence de variables aléatoires indépendantes identiquement distribuées. Le modèle est aussi constitué d'une source de bruit $b(n)$ additif, blanc et gaussien, caractérisé par une moyenne nulle et une variance σ_b^2 . Finalement, la partie du modèle que nous appellerons le canal est modélisée par une structure MA d'ordre N , dont les coefficients correspondent à une réponse impulsionnelle échantillonnée, $f(n)$.

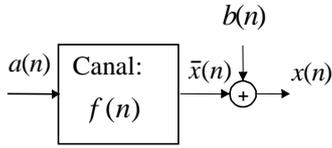


FIG. 1 : modèle d'un système de transmission de données numériques

Pour avoir une représentation mathématique simple, nous allons utiliser les vecteurs colonnes suivants :

- symboles : $\mathbf{a}(n) = [a(n) \ \cdots \ a(n - N - M + 2)]^T$;
- coefficients du canal : $\mathbf{f} = [f_0 \ \cdots \ f_{N-1}]^T$;
- observations synchrones : $\mathbf{x}(n) = [x(n) \ \cdots \ x(n - M + 1)]^T$.

En plus, dans cet article, les réalisations possibles du vecteur non bruité $\bar{\mathbf{x}}(n) = [\bar{x}(n) \ \cdots \ \bar{x}(n - M + 1)]^T$ seront appelées états du canal, ou simplement **états**.

2.1 La loi de probabilité du signal reçu

Dès que l'alphabet de la modulation est discret et que le bruit est gaussien et blanche, la fdp du signal reçu consiste en une mixture de gaussiennes donnée par :

$$p(\mathbf{x}; \mathbf{f}, \sigma_b^2) = \frac{1}{C \sqrt{(2\pi\sigma_b^2)^M}} \sum_{i=0}^{C-1} \exp\left(-\frac{(\mathbf{x} - \mathbf{c}_i)^T \mathbf{R}_{x-c}^{-1} (\mathbf{x} - \mathbf{c}_i)}{2\sigma_b^2}\right) \quad (1)$$

où $\mathbf{c}(n) = E_{\mathbf{x}}\{\mathbf{x}(n) | \mathbf{a}(n)\}$ donne les centres des gaussiennes qui, à leur tour, correspondent aussi aux états du canal, et $C = 2^{M+N-1}$ est le nombre total de ces états.

2.2 Une loi de probabilité approximative

Dans nos travaux précédents sur l'égalisation bayésienne [5], nous avons utilisé un sous-modèle d'une loi de probabilité conditionnelle de \mathbf{X} — où \mathbf{X} est la variable

aléatoire dont \mathbf{x} est une réalisation — obtenu à partir d'une somme de gaussiennes non-radiales, à savoir :

$$\tilde{p}(\mathbf{x}; \mathbf{f}, \sigma_b^2) = \frac{1}{\tilde{C} \sqrt{|\mathbf{R}_{x-c}|} (2\pi)^M} \sum_{i=0}^{\tilde{C}-1} \exp\left(-\frac{(\mathbf{x} - \tilde{\mathbf{c}}_i)^T \mathbf{R}_{x-c}^{-1} (\mathbf{x} - \tilde{\mathbf{c}}_i)}{2}\right) \quad (2)$$

où les $\tilde{C} = 2^{P+Q+1}$ nouveaux centres sont aussi des moyennes conditionnelles :

$$\tilde{\mathbf{c}} = E_{\mathbf{x}}\{\mathbf{x}(n) | a(n-d+P), \dots, a(n-d-Q)\} \quad (3)$$

En plus, \mathbf{R}_{x-c} est la matrice de covariance des observations autour de chaque centre, c'est-à-dire :

$$\mathbf{R}_{x-c} = E_{\mathbf{x}}\{\mathbf{x}(n)\mathbf{x}(n)^T | a(n-d+P), \dots, a(n-d-Q)\} \quad (4)$$

et les paramètres d , P et Q sont des nombres entiers qui contrôlent la justesse du sous-modèle \tilde{p} . Par exemple, si $P+Q = M+N-2$, le nombre de gaussiennes sera égal au nombre d'états, tandis que \mathbf{R}_{x-c} sera égal à la matrice identité multipliée par la variance du bruit. Cela veut dire que, dans ce cas, \tilde{p} ne sera pas un sous-modèle mais un modèle complet.

3. Approche proposée pour l'estimation du canal

Afin d'effectuer l'identification du canal, on a construit un sous-modèle de la loi de probabilité du signal reçu : $\tilde{p}(\mathbf{x}; \mathbf{f}, \sigma_b^2)$, où les paramètres à estimer sont le vecteur d'estimation des coefficients du canal et la variance du bruit. On notera respectivement $\hat{\mathbf{f}}$ et $\hat{\sigma}_b^2$ une estimation de \mathbf{f} et σ_b^2 . Ensuite, on a choisi une façon d'ajuster ce sous-modèle à la loi de probabilité de \mathbf{X} .

Comme il s'agit d'une comparaison entre lois de probabilité, notre choix s'est fait sur la divergence de Kullback-Leibler :

$$K(p, \tilde{p}) = E_{\mathbf{x}} \left[\ln \left(\frac{p(\mathbf{x}; \mathbf{f}, \sigma_b^2)}{\tilde{p}(\mathbf{x}; \hat{\mathbf{f}}, \hat{\sigma}_b^2)} \right) \right] \quad (5)$$

Il est clair que vis-à-vis de l'adaptation des paramètres, la minimisation de $K(p, \tilde{p})$ correspond simplement à la maximisation de l'espérance $E_{\mathbf{x}}[\tilde{p}(\mathbf{x}; \hat{\mathbf{f}}, \hat{\sigma}_b^2)]$.

3.1 La fonctionnelle utilisée

Il ressort du dernier paragraphe que notre tâche d'estimation du canal peut se résumer à la minimisation de l'espérance d'une fonction non linéaire de $\mathbf{x}(n)$. Cependant, vu la complexité de la loi de probabilité du signal reçu, et la difficulté d'une analyse théorique de l'effet de l'approximation de l'estimateur de cette loi par un sous-modèle, nous ne savons pas si les paramètres estimés $\hat{\mathbf{f}}$ et $\hat{\sigma}_b^2$ auront une solution unique pour n'importe quelle valeur de d , P et Q . Néanmoins, d'après nos résultats

expérimentaux, il est fort possible que la fonctionnelle construite de cette façon possède suivent un nombre important de minima locaux (évidemment, ce nombre varie en fonction de d , P , Q et M).

Pour contourner ce problème, nous avons adapté la technique du recuit déterministe [6] à notre cas. Nous ne détaillerons pas ici, faut de place, comment nous avons réalisé cette adaptation. Toutefois, le résultat que nous avons trouvé est la fonctionnelle

$$J(\hat{\mathbf{f}}, T) = -T \cdot E_{\mathbf{x}} \left[\ln(\tilde{p}(\mathbf{x}; \hat{\mathbf{f}}, T)) \right] \quad (6)$$

où T est une sorte de « température » associée à la variance estimée du bruit : $T = 2\hat{\sigma}_b^2$.

Finalement, le recuit est implémenté à partir d'une minimisation de cette fonctionnelle pendant que la température baisse lentement. Pour la minimisation de J nous avons appliqué le gradient stochastique, ce qui réduit le coût de calcul. Cependant, la mise au point de cette stratégie impose le choix d'un certain nombre de paramètres, ce qui sera étudié dans la suite.

3.2 L'algorithme et ces paramètres

Afin de minimiser la fonctionnelle de façon stochastique, nous devons d'abord calculer le gradient : $\nabla_{\hat{\mathbf{f}}} \ln(\tilde{p}(\mathbf{x}; \hat{\mathbf{f}}, T))$

où $\nabla_{\hat{\mathbf{f}}} \stackrel{\text{def}}{=} [\partial/\partial \hat{f}_0 \quad \partial/\partial \hat{f}_1 \quad \dots \quad \partial/\partial \hat{f}_{N-1}]^T$.

Cependant, l'expression analytique de ce gradient n'est pas évidente car la matrice $\hat{\mathbf{R}}_{x-c}$ (voir éq. 2) dépend du vecteur $\hat{\mathbf{f}}$. Donc, à la place de formules analytiques trop compliquées, nous avons développé un algorithme stochastique qui peut être vu comme une variation de l'algorithme EM [9] (pas 1 et 2), associé à la technique du recuit (pas 3). Cela peut se résumer par :

Initialisation : $\hat{\mathbf{f}}^{(0)} = [0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0]^T$, $T^{(0)}$

$k = 1$ jusqu'à NS (NS : nombre d'observations à chaque cycle)

1) calculer $\hat{\mathbf{R}}_{x-c}^{(k)}$ (analytiquement, à partir de $\hat{\mathbf{f}}^{(k)}$ [5])

2) mettre à jour $\hat{\mathbf{f}}^{(k+1)}$ ($\hat{\mathbf{f}}^{(k+1)} = \hat{\mathbf{f}}^{(k)} + \alpha_f \nabla_{\hat{\mathbf{f}}}^{\#} \ln(\tilde{p})$), où

l'indice # indique que la matrice $\hat{\mathbf{R}}_{x-c}^{(k)}(\hat{\mathbf{f}})$ est vu par l'opérateur gradient comme une matrice constante.

3) baisser la température : $T^{(k+1)} < T^{(k)}$

fin

Nous avons donc à choisir le pas d'adaptation de $\hat{\mathbf{f}}$, α_f , et la façon dont nous allons baisser la température. Nous avons choisi un taux de décroissance exponentiel. Néanmoins, il faut encore déterminer la vitesse de cette décroissance.

Pour cela nous nous sommes basés sur trois heuristiques, à savoir :

Heuristique 1 : La variance du signal reçu est supérieure ou égale à la plus grande valeur propre de \mathbf{R}_{x-c} .

Heuristique 2 : Le nombre de gaussiennes utilisées pour engendrer l'estimateur de la fdp est toujours choisi inférieur ou égal au nombre d'états du canal, donc la dispersion

moyenne de chaque gaussienne qui compose \tilde{p} est supérieure ou égale à l'énergie du bruit : $\text{tr}(\mathbf{R}_{x-c})/M \geq \sigma_b^2$.

Heuristique 3 : Pour avoir une probabilité non nulle de recevoir des observations venues de tous les états, le nombre d'observations (NS) utilisées pour chaque cycle du recuit doit être supérieure ou égal au nombre supposé d'états ($2^{M+\hat{N}-1}$).

D'après les heuristiques 1 et 2, nous choisissons les valeurs initiales et finales de la température pour chaque cycle de recuit (i. e. $T^{(0)} = 2\hat{\sigma}_x^2$, $T^{(NS)} = 2\hat{\sigma}_b^2$), aussi bien que le taux de

décroissance exponentiel : $\eta = 1 - 10^{\frac{2}{NS}(\log(\hat{\sigma}_x) - \log(\hat{\sigma}_b))}$. Enfin, la température est mise à jour à chaque nouvelle observation par : $T^{(k)} = 2\hat{\sigma}_x^2(1 - \eta)^k$.

D'autre part, l'heuristique 3 nous donne une référence floue pour les choix du nombre d'observations de chaque cycle et du pas d'adaptation du gradient. Pour les simulations présentées dans cet article, nous avons utilisé $NS = 10 \times 2^{M+\hat{N}-1}$ et $\alpha_f = 1/(2 \times 2^{M+\hat{N}-1})$.

Enfin, pour faire démarrer l'algorithme proposé on a besoin d'une estimation de l'énergie du signal reçu ($\hat{\sigma}_x^2$), aussi bien que de celle du bruit ($\hat{\sigma}_b^2$). Cependant, nous avons observé que l'algorithme n'était pas très sensible à la qualité de ces estimations.

Concernant l'ordre du canal, on suppose que $\hat{N} \geq N$, ce qui nous amène à fixer $M = \hat{N}$ et $d = \hat{N} - 1$.

4. Les résultats des simulations

Pour les simulations montrées dans cet article nous avons utilisé le canal $\mathbf{f} = [-0.5 \quad 0.7071 \quad -0.5]^T$, et un rapport signal/bruit (SNR, donné par $10 \log(\sigma_x^2/\sigma_b^2)$) égal à 20 dB. Il faut noter que ce canal possède une caractéristique particulière : sa représentation polynomiale en z présente deux zéros sur la circonférence unité¹. L'alphabet de modulation utilisé est bipolaire (2-PAM) et le nombre de gaussiennes qui composent la loi de probabilité conjointe de la variable aléatoire M -dimensionnel \mathbf{X} est 2^{M+3-1} .

D'après notre stratégie, nous pourrions réduire ce nombre et obtenir une estimation sous-optimale acceptable. Par exemple, soit $M = 3$, la loi de probabilité de \mathbf{X} se compose donc de 32 gaussiennes. Sur la figure 2, est montré un résultat typique d'un essai d'estimation avec \tilde{p} composé de seulement 8 gaussiennes. Sur cette figure, chaque courbe montre l'évolution de l'estimation d'un coefficient du canal, tandis que la courbe de décroissance exponentielle représente l'évolution de la température.

¹ Cela pose des problèmes vis-à-vis de l'identification autodidacte du canal. Les algorithmes basés sur les statistiques d'ordre supérieure (HOS), par exemple, partent de la prémisse que le canal ne possède pas des zéros sur la circonférence unité [8].

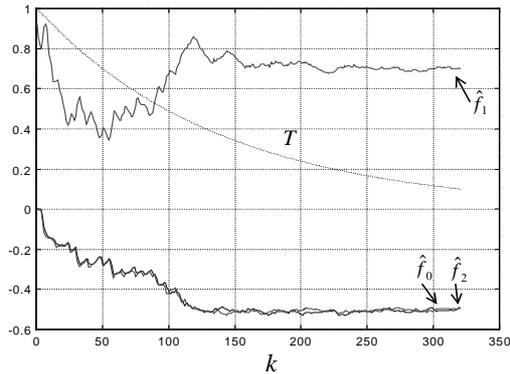


FIG. 2 : Essai d'estimation des coefficients du canal avec 8 gaussiennes SNR= 20 dB and $\hat{N} = 3$

Donc, la tâche d'estimation autodidacte de ce canal sévère est accomplie avec seulement 320 échantillons, ce qui correspond à un cycle du recuit.

Pour illustrer l'effet de la justesse du sous-modèle sur la qualité de l'estimation, le tableau 1 montre la moyenne et l'écart-type des estimations réalisées par trois sous-modèles différents : 2, 8 ou 32 gaussiennes.

TAB. 1 : Résultats après 3000 simulations Monte-Carlo avec SNR=20 dB et $\hat{N} = 3$ ($\Rightarrow NS = 320$).

Nombre de gaussiennes ↓		Coefficients du canal		
		-0.5	0.7071	-0.5
2	Moyenne	-0.3793	0.8162	-0.2170
	Écart-type	0.0286	0.0522	0.0180
8	Moyenne	-0.4885	0.6928	-0.4936
	Écart-type	0.0476	0.0619	0.0477
32 (pas de sous-modélisation)	Moyenne	-0.4997	0.7065	-0.5000
	Écart-type	0.0131	0.0131	0.0132

Malheureusement, la borne de Cramer-Rao pour l'estimation autodidacte de ce canal est trop difficile à calculer analytiquement. Cependant, juste pour avoir une référence d'évaluation des performances montrées sur le tableau, nous avons calculé la borne de Cramer-Rao pour une estimation **supervisée** à partir d'une fenêtre de 320 observations, ce qui donne un écart-type égal à 0.0056.

5. Conclusions et perspectives

Une nouvelle approche d'estimation de canaux linéaires a été proposée. Nous nous plaçons dans le cadre d'un modèle de canal MA, d'un alphabet de modulation numérique symétrique et d'un bruit additif blanc gaussien. Dans ce contexte, nous utilisons un sous-modèle de la loi de probabilité d'un bloc de M échantillons synchrones du signal reçu, ce qui nous permet de construire un estimateur de cette loi de probabilité même quand elle est trop complexe. En plus, puisque les coefficients estimés du canal sont aussi paramètres du sous-modèle engendré, l'estimation de la loi de probabilité du vecteur d'observation produit aussi une estimation du canal.

Bien que cette procédure puisse produire des estimateurs biaisés — voir, par exemple, les résultats du Tab. 1 avec 2 gaussiennes — elle rend possible une mise en œuvre de l'estimation autodidacte du canal à plusieurs rapports possibles entre le coût computationnel et la performance. Autrement dit, il nous est possible de choisir le nombre de gaussiennes du sous-modèle de la fdp, ce qui, en revanche, va affecter la qualité de l'estimateur.

Finalement, cette approche a été développée à partir d'un modèle simple de système de communication numérique. Toutefois, on peut envisager une adaptation de ce travail à d'autres situations, comme par exemple pour le cas où le bruit n'est plus gaussien ou pour des systèmes multi-utilisateurs.

Cependant, la suite plus immédiate à ce travail est l'application de cette stratégie d'estimation conjointement à l'égaliseur Bayesian que nous avons proposé dans [5] et qui est aussi basé sur la même idée d'un sous-modèle pour les lois de probabilité. Cela va nous donner, semble-t-il, des structures d'égalisation et d'estimation conjointe par maximisation de la vraisemblance qui seront facilement applicables aux cas où les structures basées sur l'algorithme de Viterbi [10] sont inapplicables.

Références

- [1] L. KULLBACK and R. A. LEIBLER. "On Information and Sufficiency", *Annals of Mathematical Statistics* 22, pp. 79-86, 1951.
- [2] T. ADALY, X. LIU and M. K. SÖNMEZ. "Conditional Distribution Learning with Neural Networks and Its Application to Channel Equalization", *IEEE Transaction on Signal Processing*, VOL. 45, NO. 4, pp. 1051-1064, 1997.
- [3] E. MOULINES, J. CARDOSO, E. GASSIAT. "Maximum Likelihood for Blind Separation and Deconvolution of Noise Signals using Mixture Models", *ICASSP-97, April 21-24, Munich, Germany*, 1997.
- [4] T. RYDEN. "Consistent and asymptotically normal parameter estimation for hidden Markov models." *Annals of Stat.* 22, 1884-1895, 1994.
- [5] J. R. MONTALVÃO, J. C. M. MOTA, B. DORIZZI, C. C. CAVALCANTE. "Bayesian Equalization Strategy: Performance versus Number of Clusters", *ITS'98/IEEE, São Paulo, Brazil*, 1998.
- [6] R. KENNETH. "Deterministic Annealing for Clustering Compression", *Classification, Regression, and Related Optimization Problems*, *Proceedings of IEEE*, vol. 86, NO. 11, pp. 2210-2239, 1998.
- [7] A. PAPOULIS. *Probability and Statistics*, Prentice Hall, 1990.
- [8] S. HAYKIN. *Adaptive Filter Theory*, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1996.
- [9] D. M. TITTERINGTON, A. F. M. SMITH, U. E. MAKOV, *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, 1985.
- [10] S. HAYKIN, *Blind Deconvolution*, PTR Prentice Hall, 1994.