

Classification de signaux non-stationnaires : Comparaison d'une approche paramétrique et d'une approche non-paramétrique

Isabelle VINCENT, Christian DONCARLI et Eric LE CARPENTIER

Laboratoire d'Automatique de Nantes, URA 823, Ecole Centrale de Nantes, 1 rue de la Noë, 44072 Nantes Cedex 03.

Membres du GdR TdSI 134

RÉSUMÉ

Nous étudions dans cet article la classification supervisée de signaux monocomposantes non-stationnaires. Nous présentons une méthode paramétrique fondée sur l'identification de la fréquence instantanée et du taux de modulation des signaux à classer par un filtre de Kalman linéaire, utilisant l'approximation de Tretter. La classification s'effectue dans ce cas à partir de la distance bayésienne mesurant les écarts entre les paramètres estimés et les paramètres caractéristiques obtenus lors de la phase d'apprentissage. La méthode non-paramétrique s'appuie sur l'utilisation d'une distance entre représentations temps-fréquence. Les performances de ces méthodes sont alors comparées sur deux exemples de signaux synthétiques : des chirps parfaits puis des signaux à bande étroite.

ABSTRACT

We study here the supervised classification of non-stationary monocomponent random signals. We first present a parametric method which realize the identification of instantaneous frequency and modulation rate of the signals to be classified using linear Kalman filtering based on the Tretter approximation. The classification step uses then the Bayesian distance between the estimated parameters and the characteristic ones which are computed during the learning phasis. The non-parametrical method is based on the use of a distance between time-frequency representations. We compare at last the methods performances. Two examples are studied : chirps embedded in noise and narrowband signals.

Introduction

La classification supervisée et la reconnaissance de signaux non-stationnaires noyés dans du bruit est une étape importante dans de nombreux problèmes de décision. Le choix de la mise en œuvre d'une méthode paramétrique ou non-paramétrique se détermine souvent en fonction de la présence d'hypothèses sur la structure des signaux traités, qui permettent alors éventuellement de conjecturer un modèle fiable de génération de ceux-ci.

Nous proposons dans un premier temps une méthode paramétrique adaptée au cas particulier de la classification des signaux monocomposantes noyés dans du bruit blanc gaussien, [4]. Cette méthode conduit à une règle de décision optimale (au sens de la théorie bayésienne de la décision) déduite d'un algorithme d'estimation paramétrique fondé sur un filtre de Kalman et du calcul analytique des probabilités *a posteriori* de chaque classe. Nous présentons ensuite une méthode de classification non-paramétrique basée sur l'utilisation de distances entre représentations temps-fréquence du groupe de Cohen [3].

Après avoir brièvement défini les notations concernant la classe de signaux étudiés (partie II), nous présentons donc la méthode d'estimation paramétrique, qui nous fournit

non seulement les estimations des paramètres du modèle mais aussi la matrice de l'erreur d'estimation (partie III). En rappelant ensuite la propriété de gaussianité asymptotique de la distribution des estimations (le filtre de Kalman utilisé pouvant être considéré comme une approximation d'un critère de Maximum de vraisemblance récursif) nous pouvons calculer analytiquement la règle de décision minimisant la probabilité d'erreur de classification (partie IV).

La cinquième partie fait l'objet de la classification non-paramétrique. Après avoir justifié le choix de la distribution temps-fréquence utilisée pour la représentation des signaux utilisés, nous présentons la distance utilisée pour quantifier les écarts entre ces distributions bidimensionnelles et en déduisons finalement la règle de décision appliquée.

Un exemple de classes de signaux à modulation linéaire de fréquence noyés dans du bruit blanc est présenté dans la partie VI, et les résultats concernant la classification par les deux approches sont comparés. Nous terminerons par une étude pratique de la robustesse de ces méthodes à l'aide de la comparaison des résultats de classification de signaux à bande étroite. Les résultats présentés sont tous issus de tests statistiques réalisés sur une large population de signaux aléatoires. Nous tirerons les conclusions



de cette étude dans la dernière partie.

1 Notations

Considérons le cas de la classification entre deux classes Ω_1 et Ω_2 (les résultats présentés ici s'étendent immédiatement au cas d'un nombre de classes supérieur à deux). On suppose que les signaux à classer sont à modulation linéaire de fréquence noyés dans du bruit blanc complexe circulaire w_k centré, de variance $2\sigma^2$:

$$\begin{cases} y_k &= a \exp(j\phi_k) + w_k \\ \phi_k &= \frac{\alpha}{2} k^2 + \lambda k + \phi_0 \end{cases} \quad (1)$$

où ϕ_0 est une variable aléatoire uniformément distribuée sur $[0, 2\pi]$ et $\Lambda = [\alpha \lambda]^T$ est un vecteur gaussien de moyenne et de matrice de variance inconnus. L'appartenance d'un signal à une classe est déterminée par les valeurs des coefficients (λ, α) , et ce sont donc ces paramètres que nous chercherons à estimer pour les utiliser ensuite au sein d'une procédure de classification.

2 Estimation paramétrique

Considérons le vecteur d'état suivant :

$$X_k = [\phi_k \ \lambda_k \ \alpha_k]^T \quad (2)$$

où λ_k et α_k sont la fréquence instantanée et le taux de modulation du signal y_k définis par :

$$\begin{cases} \lambda_k &= \frac{\phi_{k+1} - \phi_{k-1}}{2} \\ \alpha_k &= (\phi_{k+1} - 2\phi_k + \phi_{k-1}) \end{cases} \quad (3)$$

Nous déduisons alors l'équation d'évolution du vecteur d'état suivante :

$$X_{k+1} = F X_k + v_k \quad (4)$$

où

$$F = \begin{bmatrix} 1 & 1 & 1/2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Le bruit d'état v_k est une variable aléatoire centrée et de variance notée Q (nulle si le signal est à modulation linéaire de fréquence). Sous la condition que l'approximation de Treutter soit valide, c'est à dire si le rapport signal sur bruit est supérieur à 5 dB, [5], le bruit additif de mesure w_k peut être modélisé par un bruit de modulation de phase b_k et l'équation de mesure du filtre de Kalman devient :

$$\arg(y_k) = H X_k + b_k \quad (\text{mod } 2\pi) \quad (6)$$

où H est la matrice ligne $[1 \ 0 \ 0]$. Sous cette approximation la séquence $(b_k)_{k=1, \dots, N}$ est alors blanche centrée de variance σ^2/a^2 . Les équations du filtre sont alors :

$$\begin{cases} \hat{b}_k &= \arg(y_k) - H \hat{X}_{k-1/k-1} \\ \hat{r}_k &= \hat{r}_{k-1} + \mu (|\hat{b}_k|^2 - \hat{r}_{k-1}) \\ \hat{P}_{k/k}^{-1} &= \nu (F \hat{P}_{k-1/k-1} F^T)^{-1} + H^T \hat{r}_{k-1}^{-1} H \\ \hat{K}_k &= \hat{P}_{k/k} H^T \hat{r}_k^{-1} \\ \hat{X}_{k/k} &= F \hat{X}_{k-1/k-1} + \hat{K}_k \hat{b}_k \end{cases} \quad (7)$$

où $\hat{X}_{k/k}$ est l'estimation du vecteur d'état X_k connaissant les mesures jusqu'à l'instant $k-1$, $\hat{P}_{k/k}$ sa matrice de variance-covariance, \hat{K}_k le gain du filtre, \hat{r}_k l'estimation de la variance du bruit de mesure, μ et ν des facteurs d'oubli le premier permettant de régler la largeur de l'horizon de prise en compte de r_k et le second devant être fixé à 1 si le signal est un chirp et à une valeur inférieure à 1 sinon et où la détermination de \hat{b}_k est prise dans $]-\pi, \pi]$. Ce dernier coefficient permet finalement de réaliser une estimation de la matrice de variance de l'estimation lorsque l'on ne connaît pas la matrice de variance Q du bruit d'état.

Les premières valeurs obtenues par le filtre étant très sensibles aux valeurs initiales fixées par l'opérateur, nous avons choisi de mettre en œuvre un passage retour du filtre. Celui-ci se comporte comme le passage aller pour lequel on définit la matrice d'évolution d'état suivant :

$$F_r = \begin{bmatrix} 1 & -1 & 1/2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Il présente en outre le principal avantage d'être initialisé par les valeurs finales du filtre aller. Finalement, nous paramétrons les signaux à modulation linéaire de fréquence par les estimations $[\hat{\lambda}_1 \hat{\alpha}_1]$ et leur matrice de variance-covariance. Ces dernières correspondent aux dernières estimations obtenues par le filtre retour.

Si l'hypothèse de variation linéaire de fréquence du signal n'est pas vérifiée et en particulier lorsque le signal est à bande étroite, nous devons utiliser une procédure d'estimation adaptative. Nous avons donc choisi de paramétrer le signal par l'ensemble des estimations de la fréquence instantanée $(\hat{\lambda}_k)_{k=1, \dots, N}$ obtenues lors du passage retour du filtre asymptotique défini pour la même valeur de ν que celle utilisée pour le filtre aller. Ce filtre est défini suivant les équations précédentes (7) pour des matrices K_∞ et P_∞ constantes et vérifiant :

$$\begin{cases} K_\infty &= P_\infty H^T r^{-1} \\ P_\infty^{-1} &= \nu (F P_\infty F^T)^{-1} + H^T r^{-1} H \end{cases} \quad (9)$$

où r est la variance asymptotique du bruit de mesure. Dans ce cas, nous avons en outre l'expression analytique exacte de la covariance de la fréquence instantanée estimée [4].

3 Classification paramétrique

Pour un problème de classification supervisée paramétrique entre les deux classes Ω_1 et Ω_2 , les données sont :

- un ensemble d'apprentissage pour chaque classe Ω_c noté $\underline{\Lambda}_c = \{\underline{\Lambda}_{c,1}, \dots, \underline{\Lambda}_{c,n_c}\}$, où $\underline{\Lambda}_{c,i}$ représente l'estimation du vecteur de paramètres $\Lambda_{c,i}$ de matrice de covariance $V_{c,i}$
- une estimation \underline{V} de la matrice de variance V

Le problème est alors de décider si $\Lambda = \Lambda_1$ ou $\Lambda = \Lambda_2$. La règle optimale de la décision (au sens de la minimisation

de l'erreur de classification) s'écrit alors :

$$\Lambda \in \Omega_d \text{ si } \text{Prob}(\Omega_d | \underline{\Lambda}, \underline{\Lambda}_1, \underline{\Lambda}_2) = \arg \max_c \text{Prob}(\Omega_c | \underline{\Lambda}, \underline{\Lambda}_1, \underline{\Lambda}_2)$$

où $\text{Prob}(\Omega_c | \underline{\Lambda}, \underline{\Lambda}_1, \underline{\Lambda}_2)$ est la probabilité *a posteriori* de la classe Ω_c conditionnellement au vecteur à classer et aux ensembles d'apprentissage. Si nous supposons que les deux classes ont la même probabilité *a priori*, que la fonction de densité de probabilité (FDP) $f(\underline{\Lambda} | \Omega_c, \Lambda_c)$ de $\underline{\Lambda}$ conditionnellement à la classe et à la vraie valeur Λ_c peut être approchée par une loi gaussienne de moyenne Λ_c et de matrice de covariance V et que la FDP $f(\underline{\Lambda}_{c,i} | \Omega_c, \Lambda_c)$ de $\underline{\Lambda}_{c,i}$ conditionnellement à Λ_c est gaussienne de moyenne Λ_c et de variance $V_{c,i}$, alors en utilisant la règle bayésienne et de la même manière que ce qui a été présenté dans [2], la règle de décision devient :

$$\Lambda \in \Omega_d \text{ si } h_d(\underline{\Lambda}) = \arg \min_c h_c(\underline{\Lambda}) \quad (10)$$

avec

$$\begin{cases} h_c(\underline{\Lambda}) &= \text{Log det}(V + V_c) + \\ & (\underline{\Lambda} - \mu_c)^T (V + V_c)^{-1} (\underline{\Lambda} - \mu_c) \\ V_c^{-1} &= \sum_{i=1}^{n_c} V_{c,i}^{-1} \\ \mu_c &= V_c \sum_{i=1}^{n_c} V_{c,i}^{-1} \underline{\Lambda}_{c,i} \end{cases}$$

4 Classification non-paramétrique

Dans le cas non-paramétrique, on choisit naturellement de représenter les signaux à classer dans le plan temps-fréquence. Plus précisément nous travaillons avec la distribution de Wigner-Ville des signaux analytiques associés aux signaux à classer. Une fois donnés les ensembles d'apprentissage, $z_c = [z_{c,1}, \dots, z_{c,n_c}]$, et le signal z à classer, la règle de décision est la suivante :

$$z \in \Omega_d \text{ si } \Delta_d(z) = \arg \min_c \Delta_c(z) \quad (11)$$

avec

$$\begin{cases} \Delta_c(z) &= \iint (WV_z(t, f) - WV_c(t, f))^2 dt df \\ WV_z(t, f) &= \int z(t + \tau/2) z^*(t - \tau/2) e^{-2j\pi f\tau} d\tau \\ WV_c(t, f) &= \frac{1}{n_c} \sum_{i=1}^{n_c} WV_{z_{c,i}}(t, f) \end{cases}$$

La distance choisie est ici la norme L_1 entre la RTF de l'élément à classer et la RTF moyenne de chaque classe. Cette distance a été choisie en référence à d'autres travaux sur le sujet [1].

5 Application

Nous allons donc comparer les deux méthodes de classification présentées ci-dessus dans un premier temps sur des signaux monocomposantes à modulation linéaire de fréquence noyés dans du bruit blanc gaussien. Nous étudierons ensuite la robustesse de la première méthode au modèle du signal dans le cas particulier où le signal est à bande étroite. Nous comparerons finalement les résultats obtenus par la première méthode à ceux de la méthode non-paramétrique.

5.1 Signaux à modulation linéaire de fréquence

Les signaux de simulation ont été générés suivant la loi :

$$x_k^{ci} = e^{j(\alpha^c \frac{k^2}{2} + \lambda^c k + \phi^{ci})} + b_k^{ci} \quad (12)$$

où $c = 1, 2$, $i = 1, \dots, 100$ et $k = 1, \dots, N$ représentent respectivement la classe du signal, la réalisation et l'indice temporel, N étant le nombre d'échantillons total de chaque réalisation fixé à 128. La séquence $(b_k^{ci})_{k=1, \dots, N}$ est une séquence de bruit blanc gaussien circulaire centrée de variance $2\sigma^2$. Cette variance, égale à 0.3162, est identique pour les deux classes et a été choisie de manière à ce que le rapport signal sur bruit soit approximativement égal à 5 dB. La phase initiale de chaque réalisation ϕ^{ci} est aléatoire uniformément distribuée sur $[0, 2\pi]$. Les deux classes se distinguent donc uniquement par les paramètres α_c et λ_c qui ont été fixés aux valeurs suivantes :

$$\begin{cases} \alpha^1 &= 2\pi \times 7.87 \cdot 10^{-4} \\ \lambda^1 &= 2\pi \times 0.1 \end{cases} \quad \begin{cases} \alpha^2 &= 2\pi \times 8.82 \cdot 10^{-4} \\ \lambda^2 &= 2\pi \times 0.096 \end{cases}$$

Pour ces valeurs la fréquence instantanée réduite d'un signal de la première classe varie en fonction du temps de 0.63 à 1.26 et celle d'un signal de la seconde classe de 0.60 à 1.31.

Les paramètres ont été estimés par le filtre de Kalman aller-retour pour lequel les paramètres ν et μ sont respectivement fixés à 0 et à 0,1. La méthode non-paramétrique a été mise en œuvre avec la distribution de Wigner-Ville et la norme L_1 . La population d'apprentissage est tirée au hasard dans la population totale pour un cardinal variant de 2 à 20 par classe, et le pourcentage d'individus mal classés est évalué sur la population restante. La règle de classification bayésienne donne alors la courbe représentée sur la figure suivante :

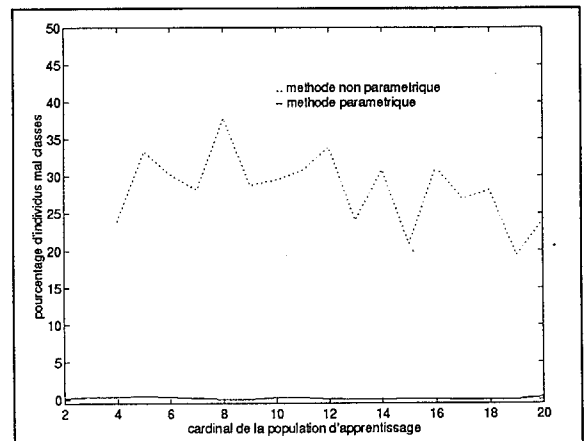


Figure 1: Classification de chirps

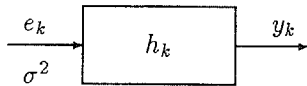
Nous constatons sur cette figure que la classification paramétrique présente d'excellentes performances. En revanche les résultats obtenus par la classification par représentations temps-fréquence sont beaucoup moins bons. Ceci s'explique par le fait que le modèle choisi dans la méthode d'identification par filtrage de Kalman correspond exactement au modèle du signal. Le filtre de



Kalman étant en outre linéaire, aucune approximation n'est faite lors de l'estimation des paramètres α^c , λ^c , qui est par conséquent non biaisée. La procédure de classification paramétrique est donc optimale.

5.2 Signaux à bande étroite

Nous étudions maintenant les résultats de classification de signaux à bande étroite par les mêmes méthodes. Les signaux sont issus de la sortie d'un filtre linéaire h_k passe-bande de fréquence centrale λ_k variable selon une loi temporelle linéaire ($\lambda_k = k \alpha_0 + \lambda_0$) :



Nous disposons de 2 classes de 50 signaux de 128 échantillons, dont les paramètres sont fixés à des valeurs identiques à celles des signaux à modulation linéaire de fréquence parfait. La fréquence centrale du filtre varie donc de 0.63 à 0.1.26 pour un signal de la première classe et de 0.60 à 1.31 pour un signal de la seconde. La variance du bruit d'entrée a été fixée à 0.01 pour tous les signaux et la largeur de bande est constante égale à $0.003F_e$. Cependant elle ne représente qu'un facteur d'échelle pour les signaux et n'influe donc pas sur la convergence de la méthode d'identification.

Pour la classification paramétrique, la méthode d'identification choisie est le filtre de Kalman aller-retour asymptotique, cependant ce type de signaux ne peuvent être paramétrés par les valeurs estimées à un instant fixé de la fréquence instantanée et du taux de modulation. Les tests systématiques de classification qui en découlent échouent alors totalement (50% d'individus mal classés). Nous avons donc choisi de les paramétrer par l'ensemble des estimations de la fréquence instantanée $[\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_N]$. Les estimations issues du filtre de Kalman asymptotique sont obtenues pour le paramètre ν fixé à 0.9. La matrice de variance covariance des estimations est alors constante et connue analytiquement [4], ce qui nous permet d'utiliser à nouveau la distance Bayésienne pour l'étape de classification.

La méthode de classification non-paramétrique est mise en œuvre avec la distribution de Wigner-Ville et la norme L_1 . Les distributions de Wigner-Ville d'un chirp parfait étudié précédemment et d'un signal à bande étroite :

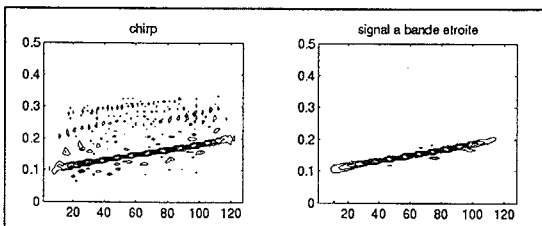


Figure 2: Distributions de Wigner-Ville d'un chirp et d'un signal à bande étroite

La figure suivante illustre les résultats de classification obtenus par la distance bayésienne :

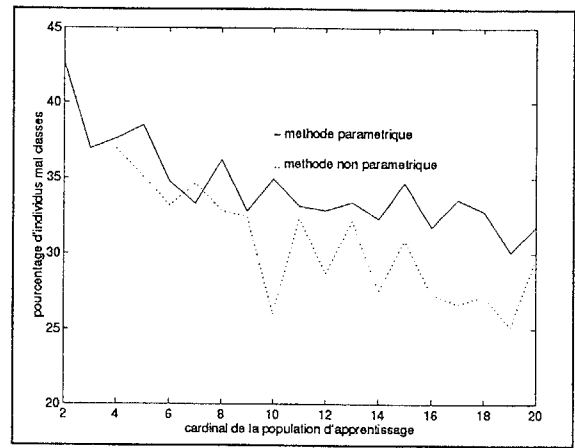


Figure 3: Classification de signaux à bande étroite

Nous constatons ici une détérioration des résultats de classification obtenus par les deux méthodes par rapport à ceux obtenus dans le premier exemple. Cependant, pour les signaux considérés la méthode non paramétrique est sensiblement meilleure que la méthode paramétrique, qui s'appuie sur une modélisation approchée des signaux.

6 Conclusion

Ainsi nous avons pu constater sur les résultats de classification de signaux monocomposantes non-stationnaires présentés sur la figure 1, que la méthode paramétrique est optimale lorsque le modèle conjecturé est identique au modèle de synthèse des signaux. Dès que l'hypothèse faite sur le modèle des signaux s'avère fautive, nous constatons une nette détérioration des résultats obtenus par la méthode paramétrique (Figure 3). La méthode non-paramétrique, loin d'être optimale dans le premier cas, conduit à de meilleurs résultats que ceux fournis par la méthode paramétrique dans le second cas.

References

- [1] B. Boashash and P. O'Shea. A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis. *IEEE Trans on acoustics, speech and signal proc.*, 38(11):1829–1841, Nov 1990.
- [2] C. Doncarli and E. Le Carpentier. An optimal approach for random signals classification. *IEEE PAMI*, 13(9), sept 1991.
- [3] P. Flandrin. *Temps-Fréquence*. Hermès, 1993.
- [4] E. Le Carpentier, I. Vincent, F. Auger, and C. Doncarli. Classification of non stationary sinusoidal signals. In *Eusipco*, pages 1429–1432, 1994.
- [5] S.A. Tretter. Estimating the frequency of a noisy sinusoid by a linear regression. *IEEE Trans on Information Theory*, pages 832–835, 1985.